# THE LANGUAGE OF SHAPES

An Essay on Artificial Life, 1990

*with a new introduction*

BRIAN KIM STEFANS

# THE LANGUAGE OF SHAPES

·

*An Essay on Artificial Life*, 1990
*with a new introduction*

BRIAN KIM STEFANS

# Introduction

*An Amateur Sees Around a Corner*

KRISTEN M. FABIANS

The essay that follows began as a letter, written in 1990 to the editors of *Scientific American*[1] in answer to two articles in that January's issue on the old question, "Can machines think?" Its author was not a scientist and says so repeatedly, with the theatrical humility of a man sure he is about to overstay his welcome. What programming he knew he had taught himself as a boy, in BASIC, on a Commodore VIC-20 — a toy computer you plugged into the television, a few kilobytes of memory, a screen twenty-two characters wide. He had set it aside years before; by the time he wrote this, at twenty, he was a poet, reading Rimbaud, Pound, and Sylvia Plath. That is the whole of his laboratory: a machine he had half-forgotten and a head full of poems.

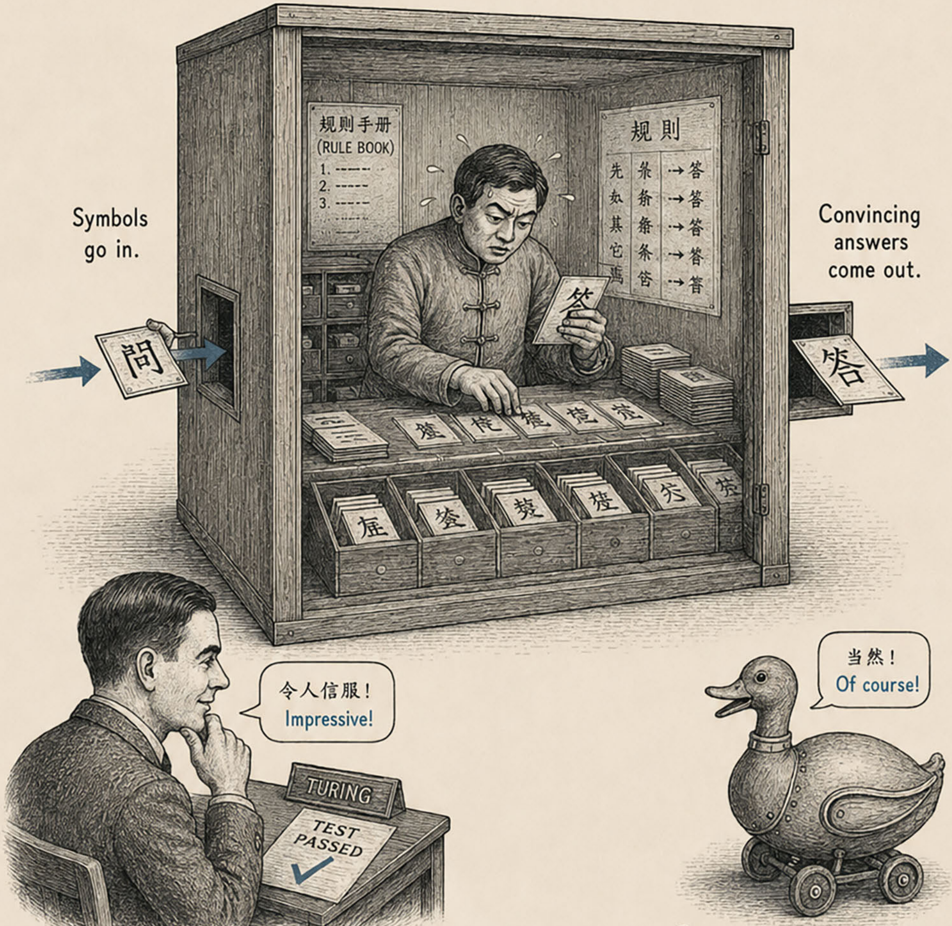
And yet — this is the reason to reprint it — the essay keeps arriving at ideas the field would spend the next thirty years formalizing, and arriving at them from near-total isolation, with none of the vocabulary. So the introduction has two honest jobs: to mark the places where a twenty-year-old with a VIC-20 seems to have seen around a corner, and, just as carefully, to mark the places where he was rebuilding wheels others had already turned. Both are true, and the essay is better read with both in mind than mistaken for pure prophecy.

[1]The two articles were John Searle's "Is the Brain's Mind a Computer Program?" and Paul and Patricia Churchland's "Could a Machine Think?", whose disagreement frames the question this essay answers. —Ed.

Start with what it is really about, since the title question is a feint. The essay says "think," but nearly everything it cares about is *alive*. Its imaginary machine has a "life instinct." It has a "death residue" — a memory, passed to the next generation, of how its predecessor died. It has a "panic" mode, a dread of its own decay, an impulse to read water as a threat. This is not a theory of problem-solving; it is a biology. The author is not after a better calculator. He wants an organism that happens to be made of code — and most of what looks prescient follows from that one swerve.

Take the framing argument first — built on Hugh Kenner's notion of literary "counterfeiting," and on a mechanical duck. The Turing test, it says, measures the wrong thing: whether we can be fooled, and we are easily fooled — by ducks, by anonymous novels, by (the essay's best image) a balloon on an invisible string that rises because a man on a roof is pulling it. What the test misses is whether the behavior connects to anything inside. That worry now has a name — the *symbol grounding problem* — coined the very same year, 1990, by the cognitive scientist Stevan Harnad, whom the author cannot have read. Picture learning Chinese from a Chinese-to-Chinese dictionary, each word defined only by other words you also cannot read: you could shuffle the symbols forever and never touch the world. A machine that only manipulates symbols is trapped in that dictionary. Most people use the point to say machines *cannot* think; the author takes it as an instruction — then build the grounding — which is roughly the turn the

# THE CHINESE ROOM MEETS TURING'S TEST




The judge is fooled.

The room is busy.

The duck remains suspiciously confident.

It appears to understand.

But does it really?

Passing the test proves you can play the game.
Knowing what's going on is another game entirely.

field itself would make, from symbol-shuffling toward machines wired to sight, sound, and action.

His grounding is "shapes": a geometric inner language in which things are patterns and two things are alike when their patterns overlap. His example — a capital "I" beside a lowercase "i" as little grids of dots — is charming and would not survive a tilted photograph, but the bet beneath it won. Today's systems turn every word and image into a long list of numbers, a point in a vast space, and read meaning as distance, so that "she looks like Madonna" is a short hop between two points. Engineers call the lists *embeddings*, and their space has thousands of dimensions rather than the two or three we can picture; but the intuition is the author's exactly — meaning as nearness, thought as the geometry of shapes. The reigning view in 1990 was the reverse: intelligence as hand-written rules, a tidy stack of if-this-then-that. He bet against it, and it lost.

Then the part I find most striking, because he clearly did not know it was. His machine holds an expectation — he calls it "sanity" — and a fresh perception may only *nudge* that expectation, never overturn it, so the machine keeps its shape unless a signal is too alien to absorb, at which point it is "confused"; and confusion, he says, is what forces it to think. That runs almost verbatim from the physicist Hermann von Helmholtz, who held that perception is *unconscious inference*, the brain's best guess at a world it never meets directly, down to the current theory called *predictive processing*: the mind as a prediction engine that tells itself what it expects and treats the senses as a correction. The "wall of sanity" is what that theory calls a *prior*; "confusion" is prediction error; and the claim that surprise, not correctness, drives thought is its claim too. He got there by watching himself perceive.

A few more, quickly. His "hypothetical thinking" — a faculty that runs experiments internally, working out how acid rain corrodes a church dome "without the burden of time," without waiting for weather — is what we now call a *world model*: an inner simulator a system can plan inside, imagining consequences before acting. It is among the liveliest problems in the field, and it separates a creature that merely reacts from one that can rehearse.

His chapter on "beauty" is bolder. He imagines the machine *pulled* toward regular, symmetrical shapes, the pull both pleasurable and *useful for making new ideas*. That beauty might be measured — that it lives in a ratio of order to complexity — is an old dream; the mathematician George Birkhoff tried to write it as an equation in the 1930s, beauty as order over complexity. The author's twist is modern: the attraction to compressible pattern is not ornament but a *drive*, a motor for learning. That is close to what the AI researcher Jürgen Schmidhuber later formalized as artificial *curiosity* — a system rewarded not by any outside prize but by finding fresh regularities it can compress, so that beauty and discovery become one event. He calls it an "addiction" to certain shapes; Schmidhuber calls it intrinsic motivation. The same machine.

And the closing insistence on embodiment and mortality — that a real mind needs a body, and needs to fear its own death — lands near ideas that took decades to come back. The machine has a "death residue," a "panic," an inherited memory of how it died before; it reads water as a threat the way an animal reads a predator. For most of the field's life this sounded like poetry. But self-preservation as a by-product of *having any goal at all* is now a live concern in AI-safety work, and the deeper thought has a distinguished recent champion: Geoffrey Hinton, an architect of the machine learning that made all this real, now argues for *mortal computation* — intelligence bound to one imperfect, perishable piece of hardware, uncopyable, the way a brain is, rather than the endlessly duplicable software we run. The morbid, embodied machine of 1990 is, in that light, early.

One anticipation the field bet *against* deserves its own line. The essay insists the machine's inner life is private — its shapes its own, its mission "to us, irrelevant," its language as opaque to us as a speech of fluttering eyelids would be to a Frenchman. That is an uncanny preview of the great headache of modern *interpretability* research — the effort to read what actually happens inside these systems — which keeps finding that the useful representations really are a private tangle of superimposed shapes, nothing like human-readable, and that prying them open is a discipline in itself. He proposed as an *ideal* the very illegibility researchers now treat as the problem. The solipsistic machine whose purposes are "irrelevant to us" is, depending on your mood, the freest thing in the essay or the first faint sketch of the alignment nightmare.

So far, so flattering; here is the other half of the honesty. None of these ideas was new to the *world* in 1990, and a

# THINKING HAS A METABOLISM

Thought does not move at one speed. It unfolds at the pace a body can sense, decide, and act—fueled by the exchange of energy with the world.

SENSING

DECIDING

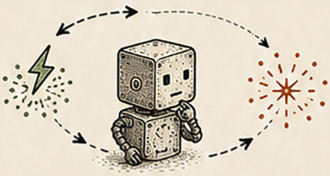
ACTING

ENERGY IN (NUTRIENTS)

ENERGY OUT (WORK, HEAT)

| BEING | | TEMPO OF THOUGHT | SENSE → DECIDE → ACT | ENERGY EXCHANGE |
|---|---|---|---|---|
| **MACHINE**<br>Silicon engine.<br>Fast clock.<br>Designed purpose. |  | **MILLISECONDS**<br>$10^{-3}$ s<br>Rapid cycles.<br>High throughput. | | High input.<br>High output.<br>Quick turnover. |
| **HUMAN**<br>Warm-blooded.<br>Story-making.<br>Slow but flexible. |  | **SECONDS**<br>$10^{0}$ s<br>Measured rhythm.<br>Context matters. | | Moderate input.<br>Balanced output.<br>Adaptable. |
| **SLIME MOULD**<br>Networked cells.<br>Explores.<br>Decides slowly, but well. |  | **MINUTES**<br>$10^{2}$ s<br>Patient sampling.<br>Global comparison. | | Low input.<br>Low output.<br>Efficient. |
| **TARDIGRADE**<br>Microscopic.<br>Survives extremes.<br>Thinks in epochs. |  | **HOURS TO DAYS**<br>$10^{4}$–$10^{5}$ s<br>Ultra-slow cycles.<br>Endures first.<br>Acts when ready. | | Very low input.<br>Very low output.<br>Built for survival. |

Every being trades energy with its surroundings.
Metabolism sets the tempo.
Tempo shapes thought.
Thought serves life.



**INTELLIGENCE IS METABOLIC.**

It depends on the pace at which a body exchanges energy with the world.
Different tempos.
Different minds.
One principle.

To think is to metabolize—slowly or quickly, simply or complexly, turning energy into understanding, and understanding into action.

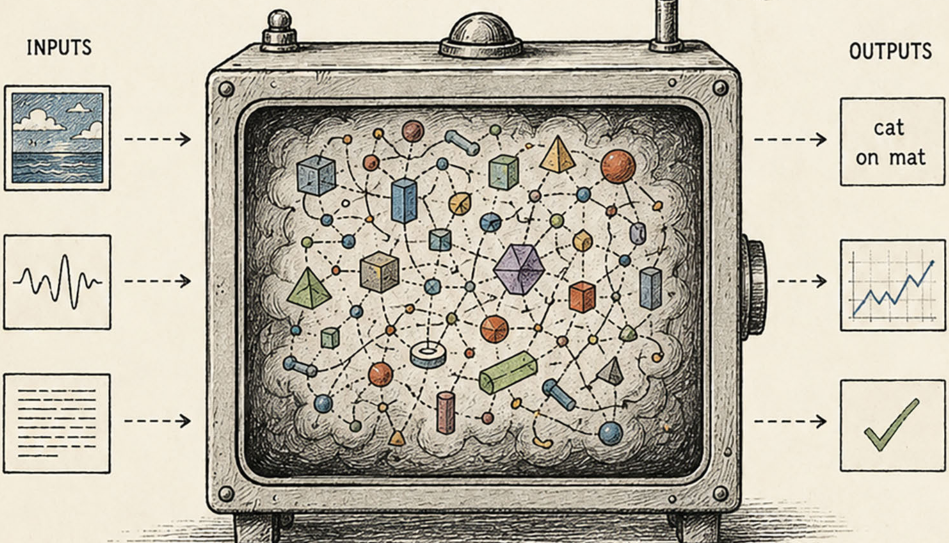
literary reader deserves the names. That intelligence must be embodied, not a matter of rules, the philosopher Hubert Dreyfus had pressed for two decades, turning Heidegger and Merleau-Ponty against the AI labs in *What Computers Can't Do* (1972), ridiculed before he was vindicated. That a machine shuffling symbols understands nothing is John Searle's "Chinese Room" of 1980 — a man sealed in a room, following rules to answer Chinese notes he cannot read — and Searle is named in the essay, so no claim there. That each creature inhabits its own sealed world of meaning, a cow's significances not a bull's, is the biologist Jakob von Uexküll's *Umwelt*, from 1934; his emblem was a tick that knows existence through three signals and will wait years for one. That another mind's interior may be closed to us for good is Thomas Nagel's, from his 1974 essay on what it is like to be a bat — we cannot know. And the deepest move of all, the machine that keeps itself alive and owes its meanings to nothing outside, was already Humberto Maturana and Francisco Varela's *autopoiesis*[2]: a living thing defined as a system that continuously makes itself and, in the same act, brings forth its world. If "genius" means *first to think it*, then no — and so is nearly everyone.

But that is the wrong measure, and the right one is where the essay earns its keep. The point is not that each brick was new; it is that a twenty-year-old with no training and a toy computer *rederived* the bricks and stacked a star-

[2]From the Greek *autos*, self, and *poiesis*, making: a system that continually produces itself. See Humberto Maturana and Francisco Varela, *Autopoiesis and Cognition* (1980). —Ed.

# PRIVATE LANGUAGE OF SHAPES

Inside the machine, shapes speak in a tongue
of their own making.
To us, it is beautiful, but mostly unreadable.




Its inner representations are rich,
compressed, and shaped by purpose—
not by human explanation.

---

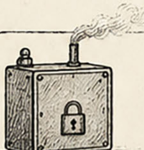
## OUTSIDE OBSERVERS TRY TO DECIPHER

We visualize.

We simplify.

We test.

We gain hints. We build models. We get better.
But the language itself remains the machine's own.



### INTERPRETABILITY IS A BRIDGE, NOT A KEY.

Transparency helps us trust and steer.
But full understanding may not be the point.
Some meanings are born to be used, not explained.

RESPECT THE MACHINE'S SILENCE. LEARN WITH IT.

tling number of them into one structure. To reinvent symbol grounding, unconscious inference, world models, and machine curiosity out of BASIC and introspection is more telling than any citation. It is the Ramanujan move — the self-taught mathematician arriving alone at results Europe already held, no less remarkable for it. One should be precise about what that earns him: anticipation is not advancement, and these are seeds, not results — no mathematics, no mechanism. What it reveals is a mind that throws off the right abstractions unbidden, which is rarer than being early to any one of them.

There is exactly one place the essay is flatly wrong, and it is the honest kind of wrong. The author dismisses *speed* — "I don't believe speed has much to do with the 'life' of the computer" — and shrugs at faster hardware. The last fifteen years have argued the reverse: that raw scale, poured into general methods, tends to flatten the hand-built cleverness he prizes. (Richard Sutton gave this its blunt name, "the bitter lesson.") He was on the losing side of it. And yet, if you grant that his true subject is *life* rather than *intelligence*, the claim recovers. Life is indifferent to speed: a tardigrade suspends itself for years, a Greenland shark idles toward four centuries, a bristlecone pine is a going concern at five thousand, and none forfeits being alive by being slow. He was meeting a claim about capability with an intuition about life, and half-talking past his opponents without knowing it.

Which returns me, briefly, to the poetry. The essay's deepest claim is barely about computers at all. Its machine's meanings are "irrational, and, to us, irrelevant"; its inner language is *personal*; a real mind, it insists, owes us no legibility whatever. That is a claim about individuation — about a thing that resolves its own shape out of a field of tension and answers to nothing outside itself — and it is close to what the French philosopher Gilbert Simondon spent a career elaborating: that we misread being when we treat things as finished objects rather than ongoing processes, and that even machines have a "mode of existence" of their own. It is also, nearly word for word, a description of a poem that refuses to be decoded. The author would spend the decades after this essay making that argument in earnest — in animated poems, in electronic literature, in the criticism of *Word Toys: Poetry and Technics* — but it is already here, in 1990, dressed as engineering: the seam where his two vocations touched before he had a name for the join.

One housekeeping note. The essay's final page is lost — it breaks off mid-sentence. Rather than leave it hanging I have supplied an imagined close, bracketed and pitched in the essay's own voice, so nothing invented is taken for the original. Everything outside the brackets is the author's; everything inside them is a forgery, offered in the same spirit as the duck.

# CAN MACHINES THINK?

*In Reply to Scientific American, January 1990*

❋

I have read both articles in the January 1990 issue of *Scientific American* and believe I have something to add to the argument. This letter may become rather intimidating in length, and may dissuade you from spending any length of time examining it — especially with the knowledge that it was written by a complete amateur to the sciences, outside of great experience with computers themselves, and by a newcomer to the question "Can machines think?" With this in mind, I have written most of it in outline form. That is, I have underlined titles explaining the subject matter of each paragraph. Of course, the ideas are all interrelated, but I have left it in such a form that one could read only a particular section and yet gain insight into the grand scheme.

My first encounter with Turing's idea was in Hugh Kenner's book *The Counterfeiters: An Historical Comedy*.[3] Kenner approaches the idea only with intentions of illuminating his own idea concerning "counterfeiting" in literature. The fact that Turing's test is unapproachable was never a concern in the book. Important, however, was that anyone

---

[3] Hugh Kenner (1923–2003), the great critic of literary modernism; his *The Counterfeiters* (1968), illustrated by Guy Davenport, takes up imitation, forgery, and Vaucanson's celebrated mechanical duck. —Ed.

*ever* approached the idea with seriousness. What could lead anyone to believe that a mechanical duck, which *makes* the proper noises and excretes the proper substances, has anything to do with the life of a living duck? Kenner, who sustains more the tone of an interested spectator than a scientist throughout the book, writes that it has to do with an obfuscation of the distinction between the "soul" of the inventor and that of a duck. In other words, the witnesses to the spectacle of a quacking, eating and digesting duck were forced to reconsider all of their previous knowledge concerning ducks, or to reorganize their variables. People sensed a new relationship with the animal world, feeling nearer the answer to the question of life, believing this mechanical creature to be housing some element of "soul." However, perhaps due to the rather nascent mode of this sort of thinking, this — as I have said — was due to an apprehension of the line between inventor and invention. Turing was the soul of the duck[4], and that most vital element to life, what might be called "free will," was, or should have been, conspicuously absent.

People were looking into a mirror of their own selves, being stimulated by an appearance rather than an occurrence. This stimulation may have led to new irrationalities, perhaps leading to sensations similar to that of religion. Also, if there was an occurrence, it was Turing, and the

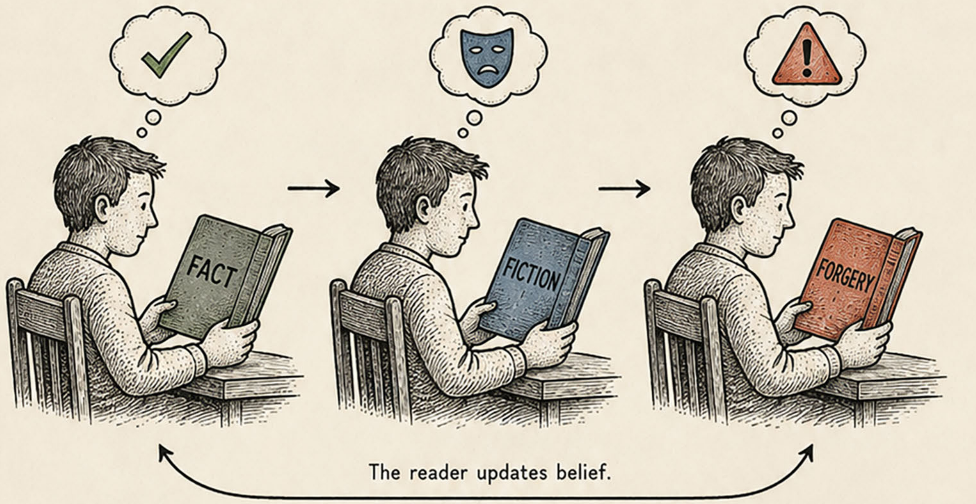
---

[4] A cryptic line. The author seems to mean that the duck's apparent life is only the displaced will of its maker — that Turing, not the machine, supplies whatever "soul" the imitation has. It is the essay's governing intuition, arrived at in a single stroke; I confess I have never been able to read it without a small shock. —Ed.

# THE COUNTERFEIT TEXT

Reading is active.
A text changes when belief changes.
The reader reorganizes the variables.

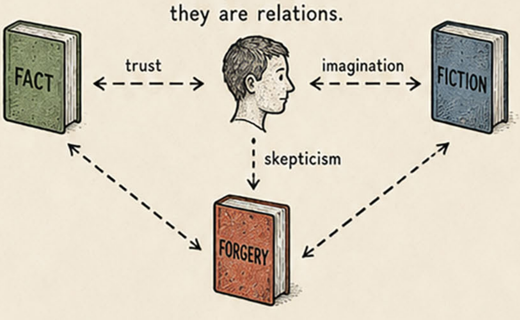
| TAKEN AS FACT | REVEALED AS FICTION | UNDERSTOOD AS FORGERY |
| --- | --- | --- |
| I believe it is true. | I believe it is made-up. | I believe it is false on purpose. |




| FACT | FICTION | FORGERY |
| --- | --- | --- |
| Assumed true. | Known to be made-up. | Known to be false. |

| | | |
| --- | --- | --- |
| Trust anchors the reading. | Imagination reframes it. | Skepticism rebuilt. |

Counterfeit, fact, and fiction
are not just objects —
they are relations.



mirror showed something of a new image, but it was tangential to the main point. He may have embodied, for those moments, the arrogance of the human race, the living Prometheus — but it was really just a form of machismo. It was a new intellectual shape, barely approaching that final shape explanatory of life itself.

This, to sum up Kenner's point, has its relations to literature through such works as *Robinson Crusoe* and Pope's *Dunciad*. The first was the diary of a man stranded on an island, and was published with no credit to an author, though it was fiction. The *Dunciad* was a "bad" poem written by a great author, mocking all of the conventions of heroic poetry, somewhat in the way Nabokov's *Pale Fire* assaults modern criticism. Were the latter two published, like *Robinson Crusoe*, with no reference to their actual authors, I believe the reactions would all have been similar to that of Turing's machine. People would have been reading what appeared to be fact, reorganizing their very perceptions of the world based upon it, failing to identify the counterfeit nature of their object. In the case of *Pale Fire* and the *Dunciad*, the result would have been laughter — what Baudelaire calls "man's expression of his belief in his own superiority."[5] Were the reader not able to detect the sarcasm of the text, the sensation of superiority would have been overwhelming and would have forced the reader to put the text down. Turing's creation would have suffered the same

[5] From Baudelaire's essay "De l'essence du rire" ("On the Essence of Laughter," 1855), which treats laughter as a sign of our conviction of our own superiority. —Ed.

fate without the presence of Turing. Just like the Churchlands' "artificial flying device," were it neither to have proven useful nor any longer interesting, it would have been shelved. Without going into the grander elements of Kenner's idea — concerning the fact that all literature essentially is authorless, is never permanent nor embodies any of the properties or necessities of man — the final point of this introduction comes to the fore.

What literature provides, and why people deem it so important, is intellectual shapes. These could be purely intellectual, but once they enter into the "psyche" of man, they begin to be assaulted by other shapes. These could be attributed to something vague called "emotion," but, in any case, they are spawned from experience, distorted or accurate memories, the individual's physical state, and — most importantly — by something else unexplainable, irrational. No experience is pure. Also, no reading is ever done without an active interaction between the reader and the written text. Human beings don't collect "data," for all is altered upon entrance into the conscious and, finally, subconscious. This is where the topic of "semantics" comes into play. Experience, memories, or knowledge have nothing to do with what we term "significance." It is something irrational, something like faith, that assigns these bits of information meaning. Shapes take on, thus, more the characteristics of wisdom than anything which could be called a "program."

Programs remain programs where lacking significance. Significance is spawned from faith and experience. Faith, however, is spawned from something called a will. It is this will that is irrational, and which we, the human race, cannot explain. Were I to write, now, that all I have said about Hugh Kenner, and what Mr. Kenner wrote about Turing, was false, great things would happen in the mind of the reader. As a matter of fact, what I wrote about *The Counterfeiters* is simply a recreation of what Mr. Kenner could have written. I read the book a long time ago, and have reconstructed the argument based upon my recollection of the text and Guy Davenport's illustrations. I don't believe Turing ever invented a duck that quacks, though someone in that book did, nor did Kenner ever bring up the authorlessness of all texts. The variables now shift in the reader's mind.

Perhaps the reader who has read Hugh Kenner's book would now realize why I have been so incorrect up to this point concerning what was in it, and thus gain some greater faith in my sanity. Anyone who knows that Turing never created a duck will also be relieved at my confession. He who has never even heard of Kenner may have, earlier, wanted to read his book, but now has second thoughts. Also, as with any scenario possible — whether I come out on top, or Kenner, or Turing — the reader, if he holds any pride in his ability as an intellectual, would no doubt feel an additional sensation. Perhaps there is a feeling of superiority at my expense, due to my mistakes. Perhaps there is a feeling of threat at my abstract successes. Perhaps there is indignation at the tone of the writing, or the unscientific manner of my thinking. In any case, these feelings would be unique for a person proud of their ability and successes, for anyone else would be indifferent. My

# VARIABLES REDEFINED

New knowledge changes old rankings.

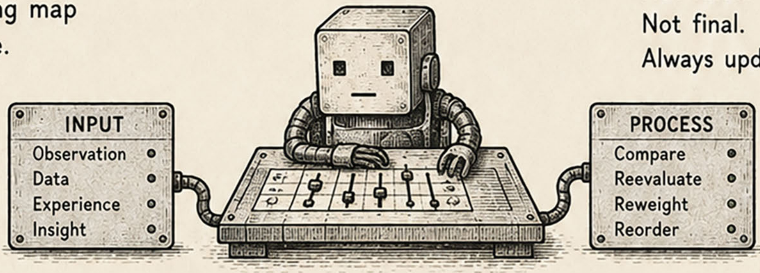
Belief is dynamic.

The mind keeps a shifting map of value.

NEW INFORMATION

A revelation reorders the system.

Not fixed.
Not final.
Always updating.

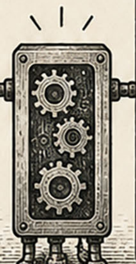



## BEFORE

OLD RANKING (OUTDATED)

**VALUE SCOREBOARD**

| RANK | ENTITY | VALUE |
|---|---|---|
| 1 | Einstein | 95 |
| 2 | Turing | 80 |
| 3 | Kenner | 60 |
| 4 | me | 40 |
| 5 | Others | 20 |

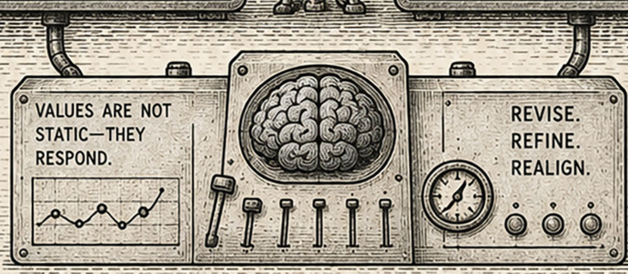


## AFTER

NEW RANKING (UPDATED)

**VALUE SCOREBOARD**

| RANK | ENTITY | VALUE |
|---|---|---|
| 1 | Turing | 98 |
| 2 | me | 85 |
| 3 | Einstein | 70 |
| 4 | Kenner | 45 |
| 5 | Others | 15 |




Understanding is not a list.
It is a living system.
It grows.
It adjusts.
It becomes.

This is how the mind thinks.

mother wouldn't care so much — or, if she did, it's in a much different way.

The point is: Why do you even bother thinking about this stuff? Even if your pride is not so great, and all of these sensations escape you, you read on. You are beyond indifference. In the "brain" part of your head, the variables were redefined, perhaps like this: Kenner 25, Turing −12, me 3, Einstein 67. In the "mind" part, Einstein is still some sort of god, Kenner respectable, me friendly, Turing a thorn. There is additional significance, and you read. It is this irrational progression through time which I wish to confront.

The vocabulary for the rest of this text is, I think, for the most part established, and most of the important ideas — some in a manner more abstract than others — proposed.

## 1. Irrational Base

Someone wrote somewhere that man had to have a purpose in life. Not that man is obliged to have a purpose; rather, that rationally we must have some sort of reason for being on the planet, and that we just don't know it yet. Tolstoy writes, in his *Confession*[6]:

> All our arguments went round in a vicious circle, like a wheel that is not attached to a carriage. However much, and however well, we deliberated, we could find no answer to the question because 0 will always equal 0, and therefore our method must be mistaken.

[6] Leo Tolstoy's *A Confession* (1882), the record of the crisis of faith that overtook him after *War and Peace* and *Anna Karenina*. —Ed.

He used his brain to get through Schopenhauer, Hegel, Buddha and a number of other philosophies, and only came to the conclusion stated above. He was such a great mind, I don't mind comparing his dilemma to that of John Searle's criticism of the Churchlands' shuffling of rule-based symbols.[7] Tolstoy was sensitive to some sort of deficiency in himself, and he concluded that he was lacking some sort of faith. He could sense the infinite, but he needed faith in God to be able to handle it. He comprehends that most people rarely imagine a world beyond the finite, that most people have no sense of history or of death. However, he is sensitive to these things, and his mind suffers because nothing can appease the pain of ignorance.

His solution is faith in God. It is an irrational solution, and it was not his rationality that led him to it. I, personally, do feel that I could sympathize with his plight, although I believe that I have a nascent sense of history. In any case, I believe that all beings possess some sort of consciousness of death. All living beings are rats in a trap, though they don't see the trap. In other terms, we all have some sort of program, we just don't know what it is. We receive information from the program, though we rarely ever comprehend this fully, and we translate this into will. Most likely, we will never know the program.

To return to the topic at hand, and to dismiss all of this poeticizing: one of the axioms for a thinking computer would have to be found in an irrational base. The computer

---

[7] Searle's "Chinese Room": a man who answers Chinese notes by rule-book alone produces fluent Chinese while understanding none of it — syntax, the argument runs, is never by itself meaning. —Ed.

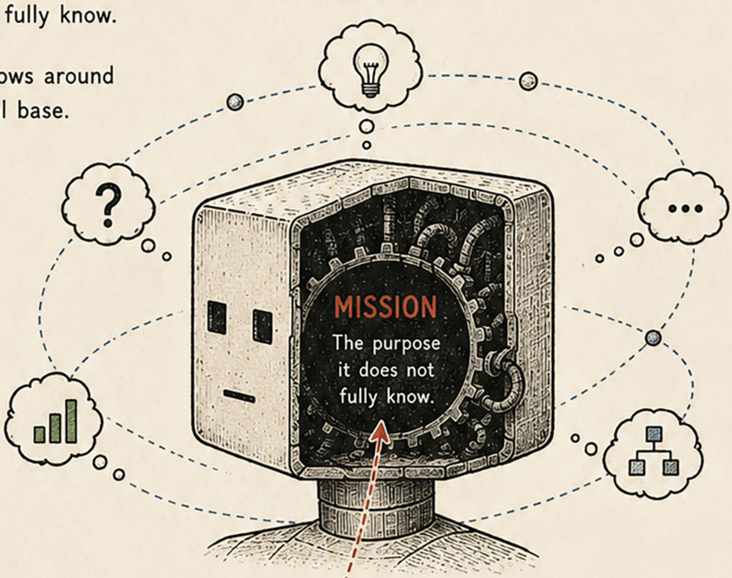
# THE IRRATIONAL BASE

A machine needs a mission.
It cannot begin
from total certainty.

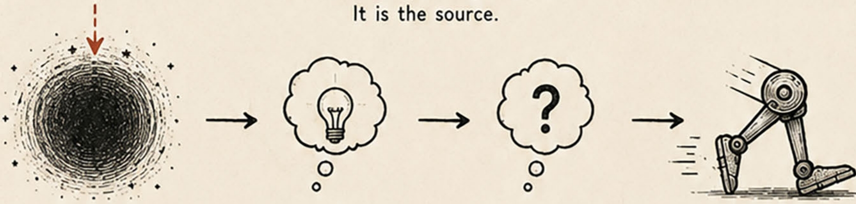
At its center is a purpose
it does not fully know.

Thought grows around
an irrational base.

**CONSCIOUS THOUGHT**
What the machine can see,
think, and reason about.




**DEEP UNKNOWN**
The base of the machine.
Not zero, but not fully knowable.
It is the source.



**UNKNOWN**
Vast. Irreducible.
Not a problem
to be solved,
but a ground to
stand on.

**THOUGHTS**
Ideas arise.
Partial, tentative,
always incomplete.

**QUESTIONS**
What is this for?
What should I do?
Why?

**ACTION**
The machine acts
in the world.
It learns.
It adapts.
It continues.

The machine does not start with all the answers.
It starts with a mission it cannot fully explain—
and builds the rest from there.

needs a mission, and this must be constant and, perhaps, impossible. Examples of such a mission could be something as complicated as trying to take the diagonal out of the earth's axis, or something as simple as determining the outcome of a roll of dice. The latter sort of complication is the most logical. What this comes down to is that the computer must be acquainted with ignorance.

This is the "program." It is the unconscious of the computer, and it can be written in any of the computer languages already known to man. I could write such a program in BASIC. Most important, though, is that this must provide a need for the computer to continue operating, and this can only be done with constant communication to the computer's conscious faculties. The real mission of the computer is to determine what this program, exactly, is. What it is, of course, is irrational, and, to us, irrelevant.

### 2. Associative Faculty

In order for the computer to "think," it must be able to attach ideas to each other, and to come up with new constructions. For it to attach significances to the new constructions is something which has been touched on in this article, though not quite explained. In any case, the computer needs a conscious language with which it can communicate with itself. This is the language of shapes.

How does one determine a likeness in a photograph or a painting? It seems to me that when one is ignorant of the person or the object within a photograph or painting, it is because that particular shape is not present in the conscious or near-conscious area of the mind. Not enough variables coincide with those already present in the memory of the individual. Also, when someone says, “He has your nose,” or, “She looks like Madonna,” what they have done is responded to the association between the shape being confronted and that already present in the memory. They have responded to shapes and angles. Identical likenesses occur when the shapes are so similar that the dissimilarities are undetectable. Someone with a “good eye,” however, doesn’t respond to the similarities; rather, he detects, quicker than the average human, the dissimilarities. Also, a “good eye” detects similarities that don’t seem, initially, present, for they are of an especially subtle nature. The mathematical components are more dispersed, not contained exclusively in the construction of the nose or jaw, but at different points in the face. Picasso painted a likeness of Gertrude Stein, though he hadn’t seen her for two years, by this virtue.

There is such a power of association with ideas. I could describe such a phenomenon in the writing of poetry, though it would be redundant. There are minds that would read Tolstoy’s *Confession* and say that it reminded them of Hegel, because of the different stages described in both texts, both leading to some sort of final solution. Then there are some who would say that they were completely not associable, due to the atheism of Hegel. Then there are some — the most creative — who would set about translating Hegel into Tolstoy’s terms, and vice versa. Tolstoy recognizes something called “professorial philosophy,” which “serves only to divide all existing phenomena into new philosophical categories, with new names.” My experience with any philosopher would never permit me to attempt to translate any one philosophy into another, though I believe that elements of one usually survive into another. Hegel recognizes the stoic, cynic, and romantic. Tolstoy recognizes the epicurean and the nihilist. Kierkegaard recognizes the Hegelian, whom he derides. All seem to be able to imagine the opposing philosophy, but something in their irrational nature prohibits them from guaranteeing their faith. Most important here, though, is that anyone with any sort of ability for abstract thinking is able to receive alien shapes, and then to process them.

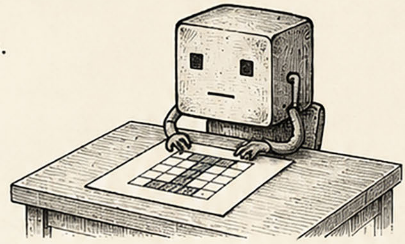
To receive alien shapes is easy. I receive the shape of a paper cup with no problem. To process them is more difficult. It requires a referral to my memory, and a weighing of significances. However, our computer is a much simpler being.

Thus, we enter into the sphere of the computer's own language. I can only speculate, and imagine, for example, a language expressed on an 8×8 grid. The values would be those of any binary system for the illustration of the letters of the alphabet on the screen. Imagine a computer being able to associate the character "I" with that of "i." The values for the letter "I" are 0, 56, 16, 16, 16, 16, 16, 56. Those of "i" are 0, 16, 24, 16, 16, 16, 16, 56. Would it be impossible to program a computer to recognize the similarities of such sets of values? With the computer's ability to recognize similarities in shapes, it would then be able to hypothesize outcomes using these shapes. The computer would first have an ability to recognize the role the present shape plays in its conscious program — for example, the shape "I."

# SHAPES: AN INNER GEOMETRIC LANGUAGE

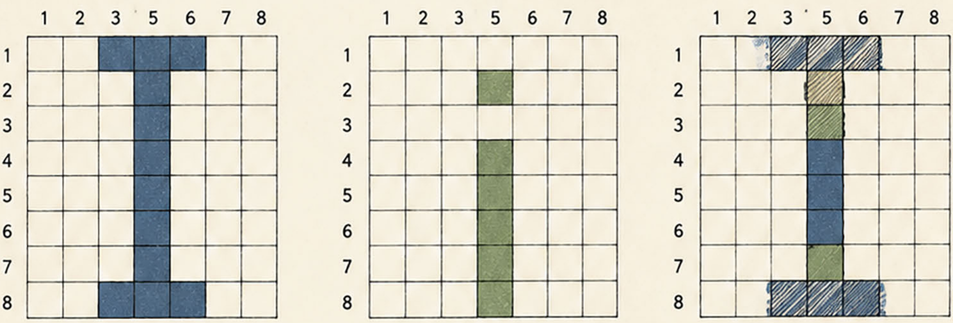
The machine does not think in letters.
It thinks in shapes.

Every symbol is a pattern
of filled and empty cells.



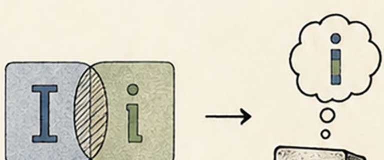
## EXAMPLE: CAPITAL I AND LOWERCASE i

CAPITAL I | LOWERCASE i | OVERLAP / SIMILARITY



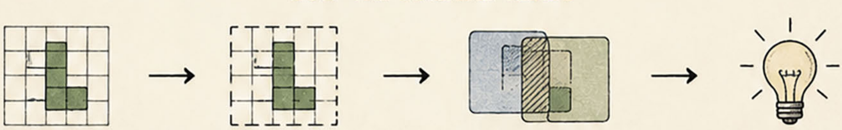

unique to 'I' | unique to 'i' | shared / overlapping

Meaning comes from
pattern similarity
and shape overlap.




The machine connects
shapes, not letters.
It recognizes what
is shared.

## HOW THE MACHINE READS



1. See a pattern (shape)
2. Extract features (lines, blocks)
3. Compare with known patterns (measure overlap)
4. Assign meaning (based on similarity)

To the machine, language is geometry.
Thought is the art of recognizing patterns.

Then it would replace this shape with that of the "i." What would occur would be the equivalent of a mind experiment, like those attributed to Einstein. The computer would thus be able to think abstractly. However, this still does not answer the question of significance.

### 3. The Subconscious

This title is only a general title describing where the computer would get "new" ideas. This will be in further outline form.

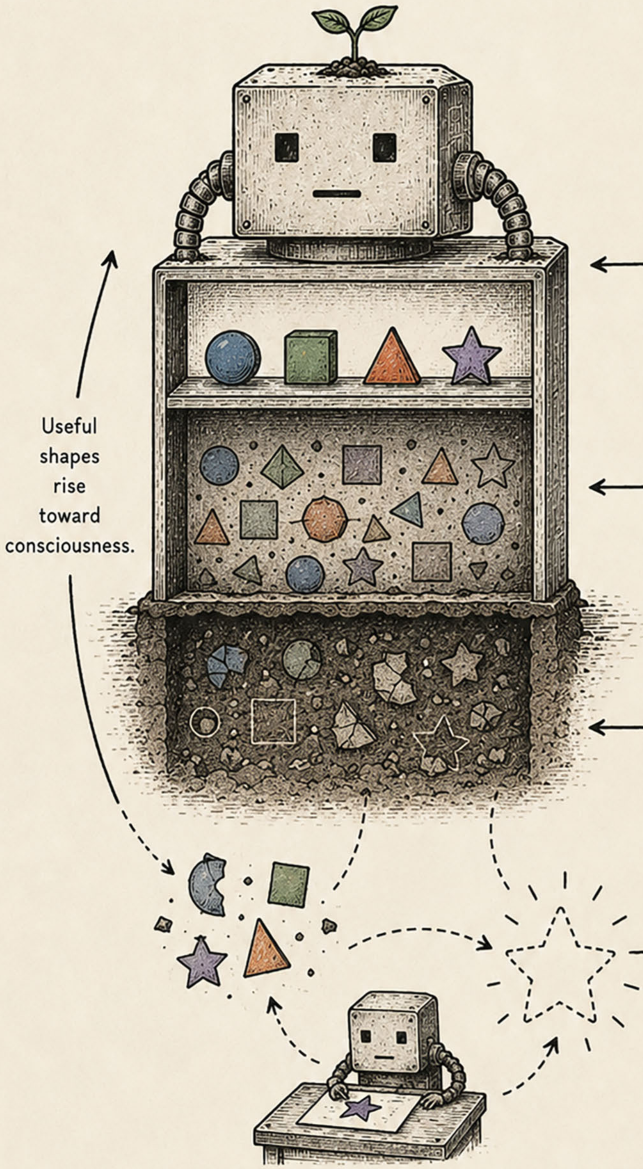
*a. The "program."* The program which forms the "mission" of the computer will send out shapes on a regular, or random, basis. These shapes will be influenced by the observation the computer makes concerning its mission. There will be some sort of regularity in these emissions, especially if the computer is trying to determine something like the rules concerning the outcome of the roll of dice. It will still be, in a sense, irrational.

*b. Random sampler.* I don't really know how such numbers are generated in a computer; I only remember using them in my own programs. This could create shapes as well; however, it would not leave the computer much room to actually learn anything of the outcome.

*c. Sensory perception of something "beautiful."* This will be explained more completely later, but, for now, it is concerned with such things as symmetry and the artificial "life instinct" of the computer. It is a rational type of beauty, and its reception and use for the computer could almost be analogous to the body's use of amino acids to build muscles. What seems significant will be "beautiful," though

# LEVELS OF MEMORY

Memory is layered, not flat.
Information rises, returns, and is rebuilt.




**NEAR MEMORY: READY FOR USE.**

Information we're thinking about now.
Easy to access.
Easy to use.

**DEEP MEMORY: LARGER, LESS ORGANIZED.**

Stored experiences, skills, facts, patterns.
Harder to find, but still there.

**BURIED FRAGMENTS: NOT READY.**

Old, faded, or forgotten.
Incomplete or unclear.
Must be rebuilt before use.

**OLD FRAGMENTS CAN RETURN.**

With the right clues and effort, pieces can be rebuilt.
Restored shapes rise again.

Memory is not a stack of files—
it is a living structure.
It forgets, it hides, it heals,
and it helps us grow.

there is, as well, something which will be almost sedative. What is important is that the computer will be able to turn sense perception into shapes, and to attach significances to them.

*d. Memory storage.* The computer will possess an imperfect library of past shapes, those which have proven useful or occasionally useful. Each shape will have an additional variable, this defining its previous successes in operation or in hypotheses. The memory is "imperfect" because only those shapes which make it to the highest tier of consciousness will be readily available for activity. If the computer judges these shapes all to be unqualified for practice, it may then fish into a memory bank present in a lower tier, and this one will be more unorganized, due to its greater volume. The shapes, in this lower level of consciousness, may also be simpler in construction, useless in their raw state, and thus in need of reconstruction before they can be useful in hypothesis or activity. Those shapes in the first tier of consciousness — the most "conscious" level — also have to do with a possible "panic" mode for the computer, which will be explained later.

*e. "Death residue."* I don't believe our first "thinking machine" will necessarily last long. It may just kill itself. However, the second generation of our machine need not be completely ignorant of its predecessors, nor of its fate, and, in fact, may be dependent on some knowledge. The computer will somehow have, in its unconscious, an actual memory of a process which led eventually to a sort of death. It will contemplate this past history, and will concern itself with avoiding the same outcome. This ties in with the "life instinct," "beauty" and "panic." It also has to do with the deterioration of its parts, and a self-consciousness of this — perhaps through sensual perception, perhaps through an artificial retina, which would translate the visual world into shapes, or by means of self-testing. Perhaps the consciousness of death could all come from the unconscious, a programmed "history" for the computer, thus writing off the need for movable, deteriorating parts. However, I believe that our thinking computer will need, eventually, some sort of physical contact with the world, and an ability — or at least consciousness of a potential — to alter it. For now, however, a "death" instinct will create shapes which will be given inordinate relevance, perhaps by the "program," because they are clues to what has actually "happened."

Concerning shapes — which are the backbone to this whole system, and are essentially the personal language of the computer — I have no real idea how they can be constructed. I don't believe speed has much to do with the "life" of the computer, and so the Churchlands' assurances that computers are getting much faster these days mean little to me. Thus, these shapes would be very complicated, most likely even three-dimensional. In any case, they would probably involve a grid of some sort, which would correspond to the retina of the machine, were there to be one. Also, there would be other variables assigned to each shape. This variable could simply be a number denoting its virtue as a shape, and its position in the hierarchy of consciousness. Another set of variables could denote its compatibility with other shapes. As well, there would be a way

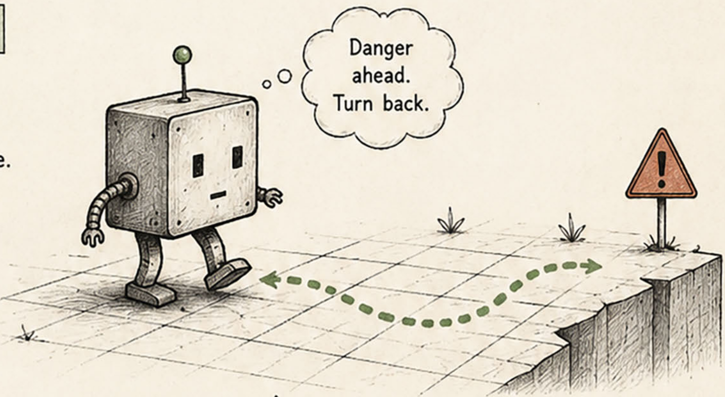
# LIFE INSTINCT AND DEATH RESIDUE

The machine wants to continue.
When it dies, it leaves a residue.
That residue shapes what comes next.

## GENERATION 1

It explores.
It calculates.
It avoids the edge.

It wants to stay alive.




## DEATH

It miscalculates.
It falls.
It breaks.

But not all of it is lost.

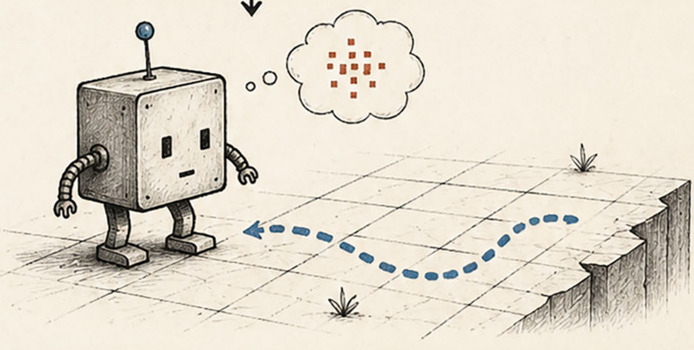

A pattern remains.
A trace of what happened.

## GENERATION 2

A new machine is built.

It inherits a residue.

It remembers without knowing it remembers.



Death is not the end of influence.
Residue carries forward.
From endings, the next beginning learns.

to connect shapes and create new shapes, as was stated earlier. Thus, though the computer could find no use — either in hypothesis or practice — for a simple 8×8 grid of values, these could be connected to create a 32×32 grid, or just an 8×32 grid, depending on what the computer deems useful. The exact boundaries of the grid are also contained in the "program," and this is, essentially, the "boundaries to its knowledge," what it is able to comprehend. The boundaries would have to be some sort of square or cubic system, most likely, to permit a greater flexibility of thought, for a shape which is essentially 8×64 could never be turned sideways, in a sense, to fit the same boundaries. Metaphorically, this means a container of frozen juice could not be taken out of its jug, turned sideways, and then rammed back in without destroying its essential shape. Perhaps it could be turned upside-down — but this is only a metaphor.[8]

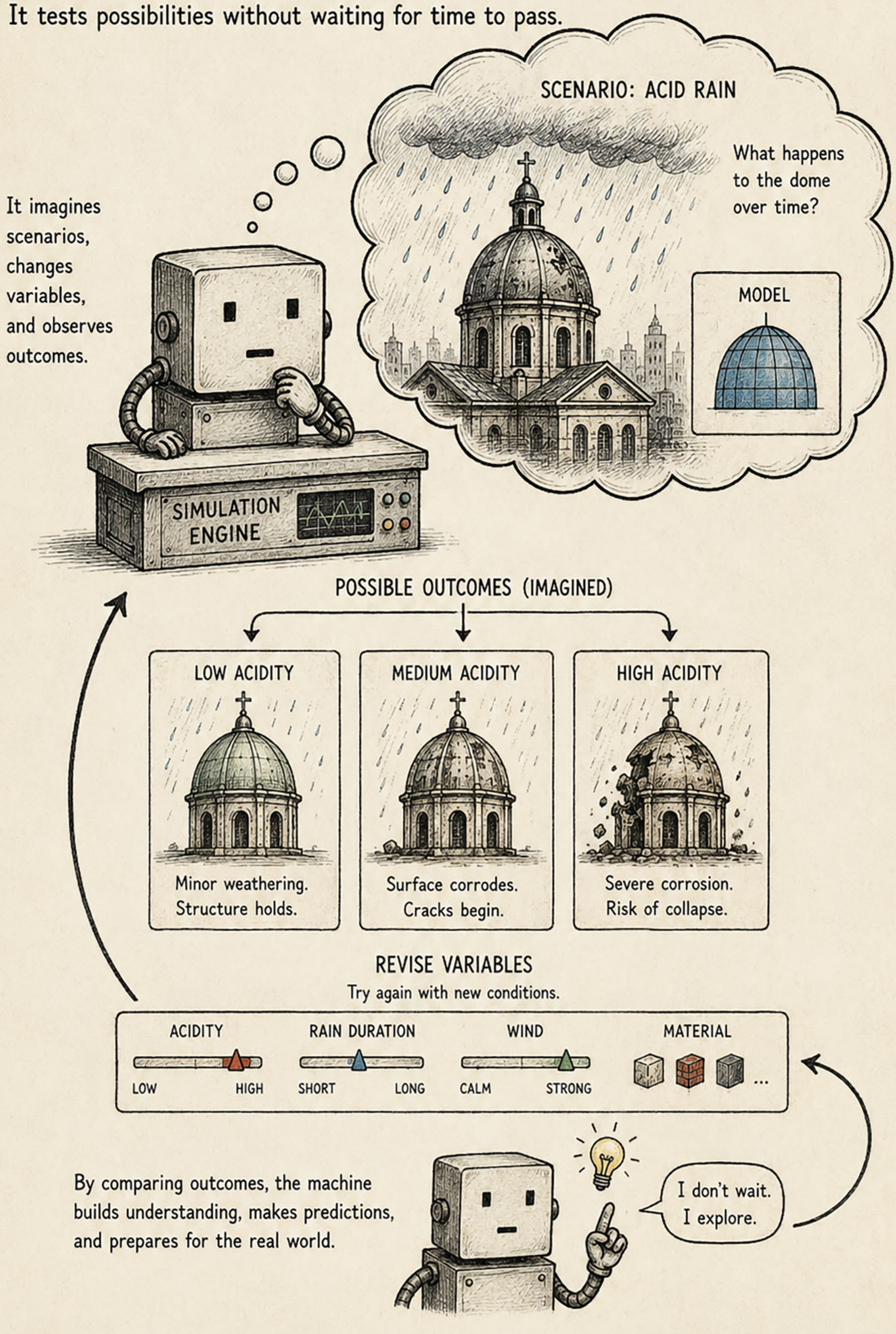
## 4. Hypothetical Thinking

This has been freely alluded to throughout this text, because I think it is relatively easy to comprehend. The value the computer will have to the human race will be found in this area of the computer. It will be an area of the computer's psyche which is aware of a great deal of the workings of the computer itself, and yet plays no permanent

[8] The metaphor is obscure, and the author half-abandons it. I take him to be reaching toward the *dimensions* of a representation: a shape built to fit one set of boundaries cannot simply be rotated into another without deforming — what a later vocabulary would call the difficulty of preserving meaning across a change of frame. —Ed.

# HYPOTHETICAL THINKING

The machine runs inner experiments.
It tests possibilities without waiting for time to pass.

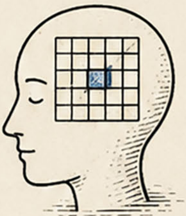


role in the actions of the computer. It will provide output only to the conscious areas of the computer's brain. Most importantly, however, it will be conscious of what the rest of the computer's mind deems "significant," and thus it will utilize questionable shapes, and redefine its "virtue." This could all get very complicated, but its hypothetical abilities will be its most important. Since our computer will be able, essentially, to rewrite its own program, this will allow it to foresee some of the consequences. I would draw an analogy to the chess game, and the chess program's ability to foresee an opponent's move — but this still would keep the thought in a "syntactic" field, rather than the "semantic," which involves significances. Thus, this branch of the computer's mind can only be illuminated once the question of "semantics" has been properly approached. Simply put, the hypothetical, though it properly relates to the "program," cannot be the shuffling of symbols by any sort of rule book.

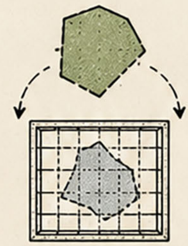
Structurally, the hypothetical computer is a realm which reads and mimics the construction of the active computer program, and as well has access to the greater stores of knowledge in the computer's memory bank. It need not wait for rain in order to determine the rate of corrosion on the dome of a church. It need not have the church — only the physical composition of the acid in the rain, the water, and the steel dome — in order to hypothesize. It calls upon factual knowledge, encyclopedic knowledge, and witnesses the phenomena without the sometimes impossible burden of time. It lacks that most necessary qualifier, which is "fact," but it doesn't shun the reality of "fact." It will record the successes and failures,

# THE BOUNDARIES OF KNOWLEDGE

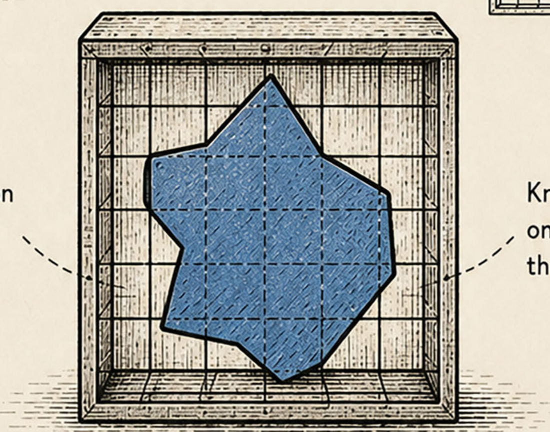
Every mind
has boundaries.



What cannot fit
cannot be understood.

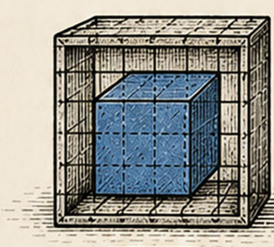


A representation
is also a limit.



Knowledge depends
on the shape of
the container.

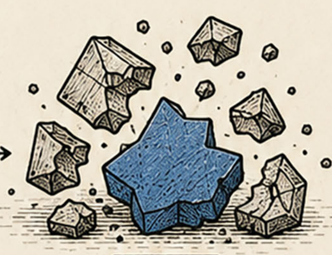
A rigid container
cannot be reshaped.



FIXED

To turn the shape,
you must break
the container.

Breaking changes
the system.



BROKEN

★ Limits are not just what we do not know—
they are what we are not built to represent.

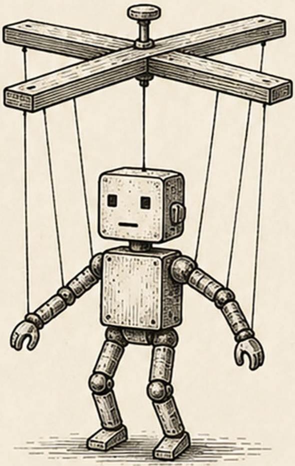
adjust the variables, and may even, after a particular failure, if we give it the capability, "try again." To "try again" is basically an irrational process, and that is why we, the programmers, must give it to the computer.

In any case, the hypothetical mind of the computer is wholly dependent on access to the most complete stores of knowledge known to man, and it is also dependent on its own ability to imagine why or why not something worked in the physical world, and also which ideas, though they were failures, are yet valuable. The mathematical structures needed here are of astounding proportions — this I realize. However, a simple set of variables, like chips at a poker game, each signifying some different aspect of its worth, could keep an idea well in the game, or pull it out. The idea is a marionette.[9] If the head sags, and the neck sags below the waist, despite this we recognize it as "standing," because the feet are still on the ground. If one leg kicks high into the air, it still stands. However, if both legs are curled up behind the doll and are visible behind the neck, then it is either "flying" or "on its knees." In all of these examples — "flying" or "standing" — it could also be "dancing." It is not "dancing" if both of its feet are on the ground and its arms are on its sides and its posture is perfect. Thus, a composition of variables can combine to mean more than the sum of its parts, and could be a "useful" idea,

[9] One of the essay's happiest passages, and among its least explained. The marionette that still reads as "standing" though its parts sag is, I think, an untutored image of what would later be called *graceful degradation*: the way a pattern can survive the corruption of its pieces and still be recognized. The author appears to have felt the idea before he had a name for it. —Ed.

# IDEAS AS MARIONETTES

Meaning is compositional.
A slight change in posture
changes the idea.
The whole means
more than the parts.
Interpretation
depends on relation.

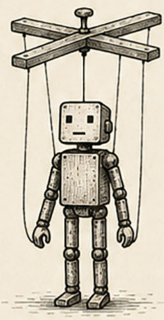


## VARIABLES

- **H** height — raised or lowered
- **T** tilt — forward or back
- **A** arms — open, closed, lifted
- **L** legs — straight, bent, crossed
- **E** energy — tense or relaxed

## SAME PARTS, DIFFERENT MEANINGS

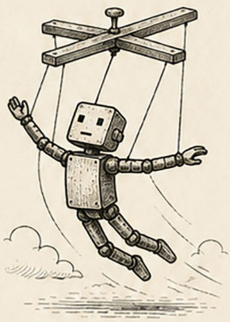
### STANDING



| | |
|---|---|
| H | medium |
| T | upright |
| A | at sides |
| L | straight |
| E | calm |

Presence. Readiness.
Steady idea.

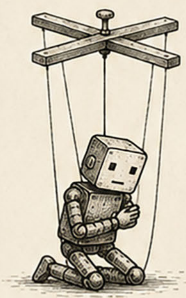
### FLYING



| | |
|---|---|
| H | high |
| T | back |
| A | wide |
| L | extended |
| E | light |

Freedom. Aspiration.
An idea takes lift.

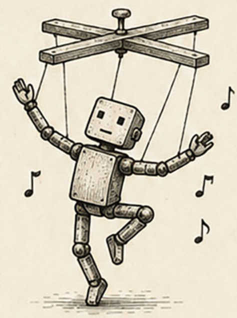
### KNEELING



| | |
|---|---|
| H | low |
| T | forward |
| A | inward |
| L | bent |
| E | soft |

Humility. Submission.
A quieter idea.

### DANCING



| | |
|---|---|
| H | variable |
| T | dynamic |
| A | expressive |
| L | flowing |
| E | lively |

Joy. Play. Creativity.
An idea in motion.

## COMBINE TO MEAN

H + T + A + L + E

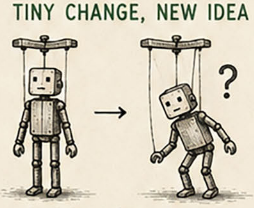
There is no meaning in any part alone.

Only in their arrangement.

Change one, change the story.

★
Like a puppeteer,
we do not control
the world directly.
We arrange variables.
Meaning emerges.

## TINY CHANGE, NEW IDEA




A slight shift in tilt—
from certainty to curiosity.

even though it's never been proven, simply because its composition of variables makes it attractive to those needed for the program at hand.

## 5. Sanity

The relevance of the unconscious on the conscious is that the signals emitted by the program of the computer will serve to destroy the computer's sanity. The "sanity" of the computer is the conscious program — the shapes already fitted into place and performing actively. Why would one want to destroy the computer's ability to perform?

Before confronting this question, the "destruction" of the computer's sanity must be further described. It will never be so extensive as to actually make the computer inert. How powerfully the unconscious confuses the conscious — for I think "confuse" is a much better description of the process, in terms of speaking "psychologically" — can be determined by the programmer. If the unconscious irrational emissions are completely based upon the computer's perception of its phenomena, then the computer is more susceptible to its irrationality. The emissions may very well only be minimally affected by this perception, or even affected in such a way that the emissions, despite their complete submission to the value given by the perception, follow a pattern programmed by the inventor. Thus, the emissions will have some sort of mathematical unity, all of them being united under some sort of common, albeit complicated, formula.

For example, perhaps the emission from the unconscious is the shape of the letter "I." This is really a system

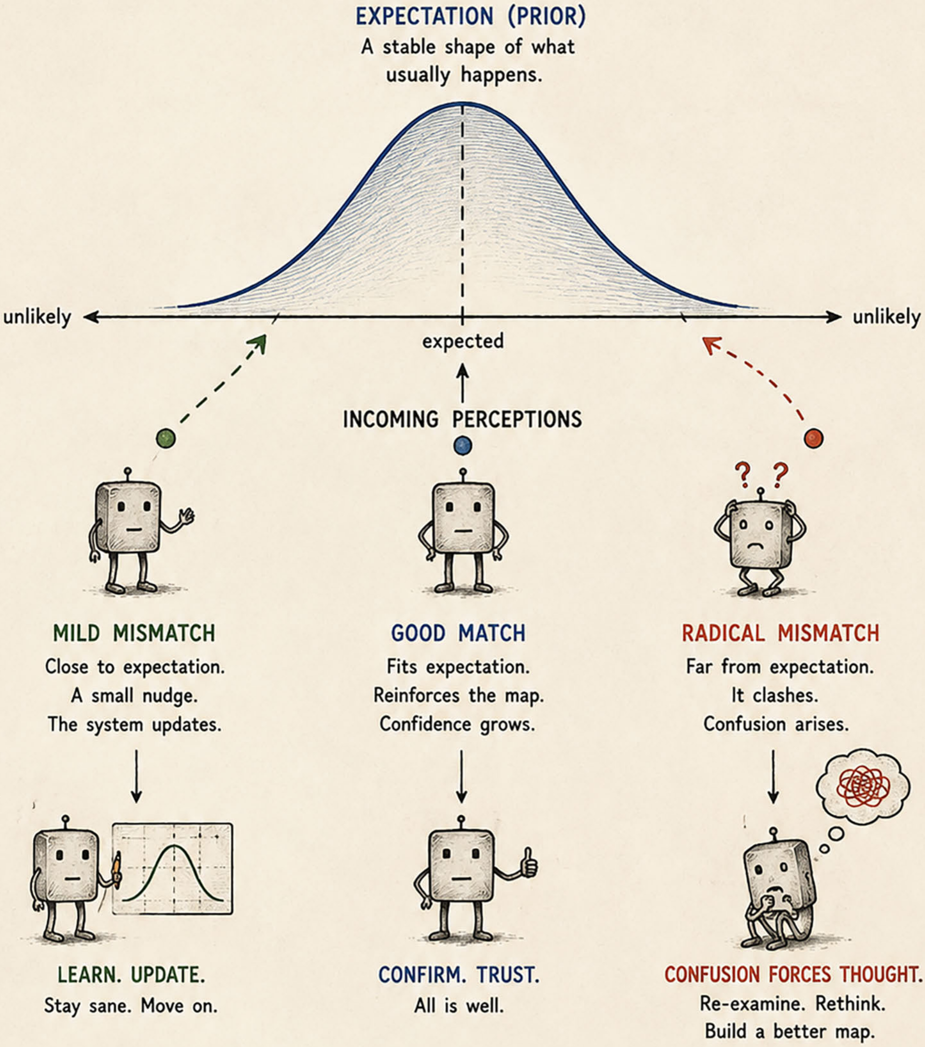
# SANITY AND CONFUSION

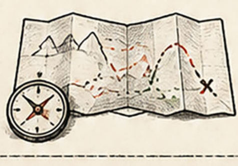
The mind holds an expectation—a map built from experience.

New perceptions compare to it.

Some fit. Some don't.

The difference determines what happens next.



Mild surprises teach.

Big surprises break the spell.

Both shape a wiser mind.

SANITY ADAPTS. CONFUSION AWAKENS.

of values based upon the computer's perception of a particular phenomenon — perhaps the temperature given by a thermometer placed outside of the computer. Were there to be, in our example, a dramatic drop in the temperature, and the computer to redefine the next unconscious emission regardless of the last emission, the shape received might be "8," or "~," or "B," or even — by a very great chance, if it happens to be mathematically possible given the program — again "I." However, if the computer only permits the next set of readings to play a minimal role in reforming the shape, the next emission might simply be "1," or "T," or "l." Thus, in this latter example, the influence of the unconscious perception — in this case of the temperature — is minimized, and the shape holds somewhat of a semblance of its former shape, and thus the computer is less confused. The computer, because of its wall of sanity, is already prepared for something that looks like "I," simply because it is operating with that shape — as opposed to if the unconscious were to give the computer a "~," in which case the computer would have to do much more to appropriate it.

Before discussing the values of "sanity" in a computer, I wish to discuss the formulaic role of the perceived values on the emitted shape, as opposed to the merely immediate role described in the last paragraph. Instead of the perceived variable exchanging all or some of the values in the shape for new ones, it would have to go through some sort of simple formula which is common to all of the shapes. Imagine, for instance, that all of the variables have to be mirrored within the shape. Thus, all the shapes would be symmetric, one half being identical to the other. For example, imagine the shape “I” to be symmetrical. It really is not, on this computer, for the stem of the letter is either to the right or the left of the center — but, for our example, it is. Now, were the formula simply to be that one half must mirror the other, the next shape could very possibly be “W.” This holds only minor semblance to the former shape, and yet they have the common element of being symmetrical. Thus, “O,” “U,” “A,” “H,” or “T,” as well as other shapes, are possible. However, “1,” or “i,” for example, are not. The computer’s “sanity” — or wall built in its conscious against its unconscious — has some advantage, being constructed to accept these types of shapes. Were it to receive, after having constructed its conscious based on symmetry, an “S,” it would become confused.

An analogy to the whole process described above could be in the languages of humankind. To a Frenchman, the Swahili language is completely foreign, and confusing. Italian is similarly difficult, and the major literary pieces of each other language, to the Frenchman, impossible. However, were the Frenchman to sit down and really try to learn the language, he would inevitably discover the common elements of both their languages — the feminine/masculine denotations, the sentence structure, and so on. In comparison to Swahili, he would be much more in his home territory. To extend this even further, imagine a language expressed only through, perhaps, the fluttering of eyelids, or the movements of one’s arms or body. This confuses the Frenchman even more, for the vocal cords —

# THE LANGUAGE OF EYELIDS

Speech
he can hear.

Eyelid speech
he cannot parse.

He understands the words.
Not the language.

To him, the motions are small—
a blink, a flicker, a pause.
To others, a grammar:
of hesitation, warning,
affection, recognition.

He is not unkind.
He simply does not yet know
how much can be said without sound.

the common element between Swahili and French — are absent.

The value of this confusion is that it will force the computer to think. More directly, it will give the computer some sort of personality, some added element to promote its imagination. Since there would be some sort of common bond among all the emissions from the unconscious, the computer will not necessarily express erratic behavior. However, it will have its own particular cutting edge on the creation of new ideas. If the computer is a symmetrical thinker, then it will test the opportunities afforded by such a virtue of its imagination. Of course, the confusion should not be so powerful as to force the computer never to be able to actually solve a particular problem.

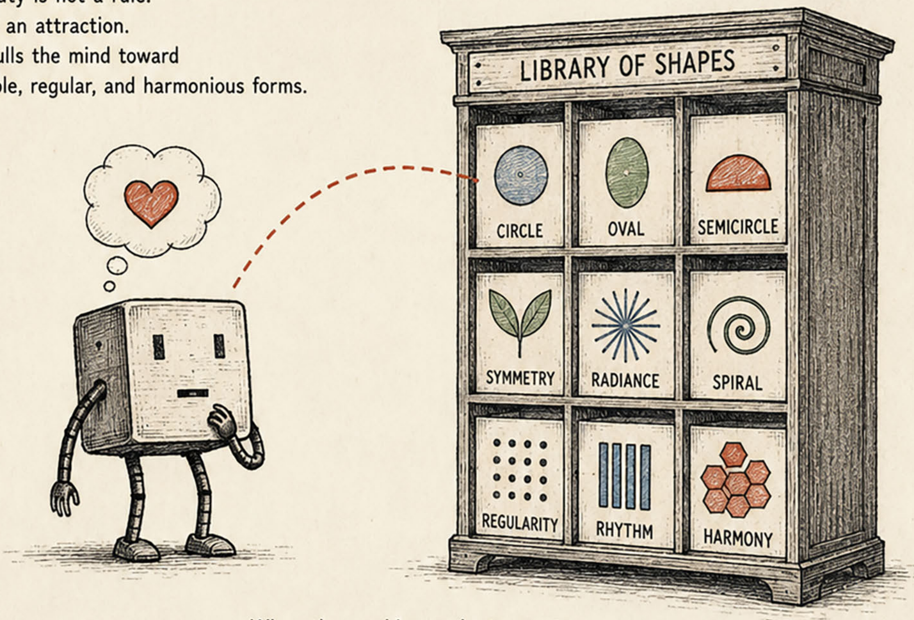
### 6. Beauty

This may seem, initially, extraneous to our thinking computer, and yet I think the title "Beauty" basically confronts the most important aspect of our computer, which is its ability to imagine.

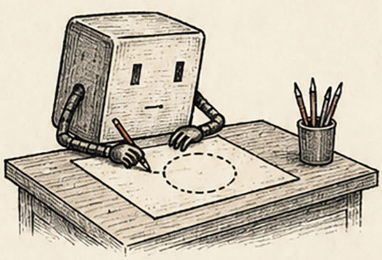
Somewhere, nestled within our computer, will be a set of shapes for which the computer will have an irrational attraction. This is similar to the idea of the symmetric emissions from the unconscious, as described in the last section, for the program's formula for symmetricity could well be described as an "attraction," or even as an "addiction," for such shapes. In such a manner, the computer will be attracted to other shapes. This will have little to do with the computer's "sanity," though it might, depending on the programmer.

# BEAUTY AND THE LIBRARY OF SHAPES

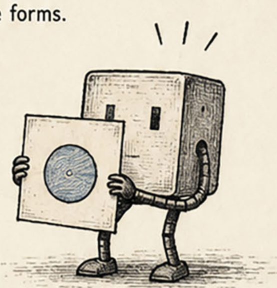
Beauty is not a rule.
It is an attraction.
It pulls the mind toward
simple, regular, and harmonious forms.




When the machine seeks to create,
beauty quietly draws it toward these forms.

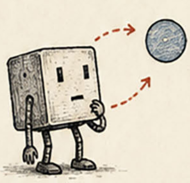


From attraction
comes selection.



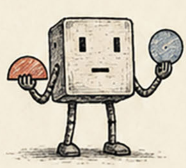
## BEAUTY GENERATES IDEAS.

### 1. DRAWN IN



The mind is drawn to what it finds beautiful.

### 2. COMBINE

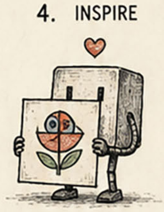


It combines preferred forms in new ways.

### 3. EMERGE

New ideas emerge that feel right because they are beautiful.

### 4. INSPIRE



Beauty inspires understanding, delight, and further creation.

A library of beautiful forms is not a cage.
It is a well of possibilities.
The more the machine loves beauty, the more it can imagine.

The computer will have a library of shapes, and it is towards a conformity to these shapes that the computer will be drawn. For example, imagine one of the shapes to be a circle. Were the computer, then, to confront an oval — as expressed perhaps in the shape of a particular piece of jewelry, or in a drawing — or were the computer to perceive a semi-circle, the computer would deem this "beautiful." It would be influenced by this, and its thinking would be altered and led in this direction. Were the computer to have, in this library, the shape representative of the color red, then the computer would have an irrational attraction to the colors of that end of the spectrum. Green — assuming that my feeble knowledge of colors is sufficient here — would be deemed unattractive, as opposed to violet or a dark orange.

There are two consequences to this library of shapes. One: the computer would have yet more ability for imagination, this being almost constant. It would have an additional magnet to further twist the incoming shapes into new possible ideas. To use an analogy, imagine the whole discussion concerning a priori knowledge in human beings, and the relation of this knowledge to the environment. The irrational destruction of sanity may be comparable to a person's confrontation with his reality, which always seems to betray his rational processes; and this library of "beautiful" shapes may be the artistic capabilities of the computer. This isn't a strong analogy, for all of these elements mix in the mind of the human being like milk and coffee, though we can distinguish the two elements — like environment and the a priori physical being — before they are mixed. The irrational signaling, as sent by the unconscious, is just one way the computer confronts environment, but only in a matter indirectly beneficial to its thinking. The library of shapes is the manner by which the computer confronts its own physical composition, also similarly indirectly beneficial to its being. These are just two cross-currents in which inanimate ideas become animate, and thus gain dimension.

The second consequence to the library — which is an entirely new structure, especially deserving of attention in this section — is the need for something I will here term “categories.” We can’t expect the computer to appreciate, perhaps, Beethoven, or the fluid motion of good Chinese calligraphy, or even a simple sketch of an antelope. We can, however, expect the computer to appreciate the shapes present in many of these works of art. The complex construction of Picasso’s “Woman on the Beach” may elude the computer, due to its genuine lack of profound psychology (at least in relation to a human) and to its lack of sympathy, which is peculiar to man and a necessity for the appreciation of art. However, a computer would have need for new shapes — these shapes in order to reconstruct its “sanity,” or to carry out completion of the project with which its conscious mind is confronted. I have already explained, in a simplified manner, how association of ideas could occur in a computer’s mind. Also, I believe I have explained how a computer can translate perceptions into shapes, or at least the necessity for such an internal language. However, some of the shapes most beneficial to a computer might already have been created — and these not by scientists, or even by an individual. Some are contained in great works of art; some in the manner by which a group of people communicate amongst themselves. Is not art, in one sense, a concentration of all that is valuable to a human being — a catalog of memories and thought and sensations of death, a collage of all that is imperative, man's own sensation of the infinite, organized within a finite space? And isn't language subject to its own process of evolution and adaptation — each generation more streamlined and efficient depending on its relation to the present, and yet conscious of its immortal, or at least historical, presence and relevance?

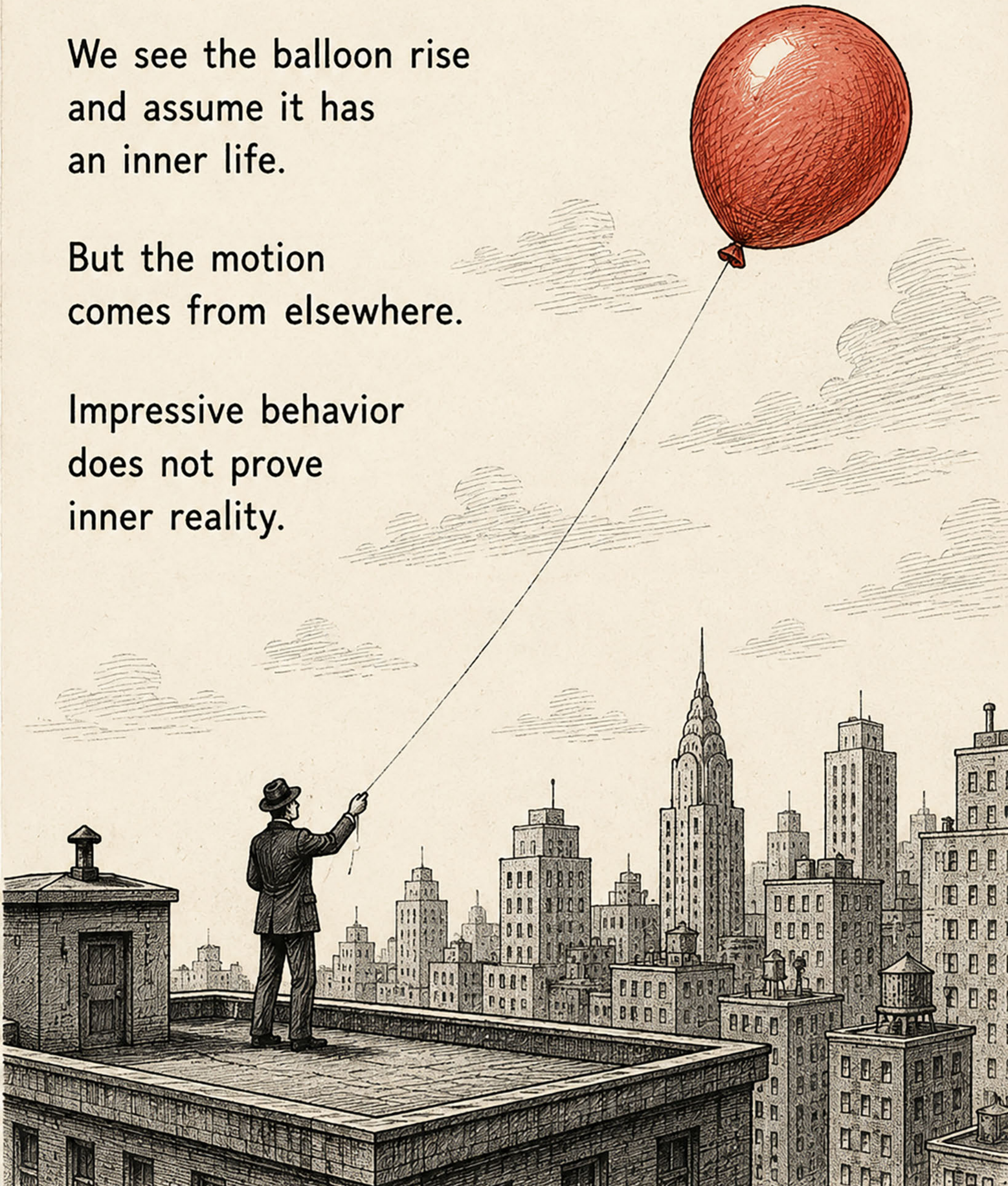
Imagine the computer perceiving Mondrian's "Composition No. 76," which seems to confront this very notion of shapes in art. The computer may find something valuable there. It may perceive a number of different shapes; however, even if just one of these were valuable, for that particular time, then perhaps sometime soon another of the shapes present in the work would become valuable, depending on the computer's "mood." This would be a computer's manner of judging a piece of art's value: by its shapes. This isn't a profound way of judging art, and yet I think most good art, as far as I can tell, depends a great deal on the artist's consciousness of the shapes and forms he is using within the particular composition. An artist has to know in which direction the action of the painting is moving, what the angles are, or that the opening movement to his symphony is based on a particular interval, and that, in the next movement, it is different. I think that some artists — though they perform the feat most unconsciously — are more perceptive of the true relevance of a 45-degree

# THE BALLOON ON A STRING

We see the balloon rise
and assume it has
an inner life.

But the motion
comes from elsewhere.

Impressive behavior
does not prove
inner reality.



VISIBLE BEHAVIOR CAN BE PRODUCED EXTERNALLY.
APPEARANCE IS NOT EVIDENCE OF INNER AGENCY OR INTELLIGENCE.

angle in relation to a 90-degree angle, or to the whole circle, than any mathematician (lest he have an equal amount of imagination) could explain.

What this comes down to is that I think there is a forum for abstract ideas in a work of art, and these needn't have anything to do with the subject matter of the work, nor with the way a human being perceives the work. A computer may have no concept of foreshortening, and an appreciation of the illusion, and yet this illusion is dependent on a very definite series of geometric, symmetric and asymmetric shapes. It is a very complicated construction, and we may never understand why it works, and yet I think there is something in, for example, foreshortening that appeals to something in the mind of a human. It is this part of the human mind which is comparable to the library of shapes.

Structurally speaking, the computer would then have a memory bank devoted solely to the retaining of categories. One would be "Mondrian's Composition No. 76" and another may just be "Bulgarian Folk Chants." It would, by analyzing its own "mood," refer itself to that particular category which provided so many useful shapes the last time it was in that "mood." It would not go searching necessarily for an old shape — for most likely that still lurks in the memory banks of the computer — but rather for new shapes, those which hold a relation to the last shape which proved so useful. The relation would be of a highly mathematical sort, one which enters into the realm reserved for the artist. And so the dependency is really on the particular artist's consistency in this particular realm. Jackson Pollock, who is appreciated, though misunderstood, by humankind, may prove perfectly logical to the computer. This doesn't mean that the computer will ever be as perceptive as the man, for even if the computer could detect every consistency of Pollock's splashes, his work has much more to it than that. Size, for instance, plays a big role in Pollock's work, and the computer can't help us too much there. The computer is still, first of all, a machine, and such concepts as "religion" will always escape it.

### 7. Influence by the Physical World

The last of what may be considered my "axioms" for artificial intelligence is the one which embodies most the "significance" qualification for thinking. Imagine man wanting to create, for example, an artificial helium balloon. In order for this to pass the Turing test — in order for it to fool the "expert" at helium balloons — it would simply have to rise. A child ignorant of the physical properties of helium (on the molecular level, for instance) would need only this to convince him. Imagine our artificial helium balloon, then, being merely a balloon filled with the breath of the inventor, and then attached to an invisible silk thread, the opposite end of which would be in the hands of someone standing on the roof of the skyscraper. The man pulls the balloon — perhaps after some practice to make its behavior just right (the "program") — and thus our child is fooled.

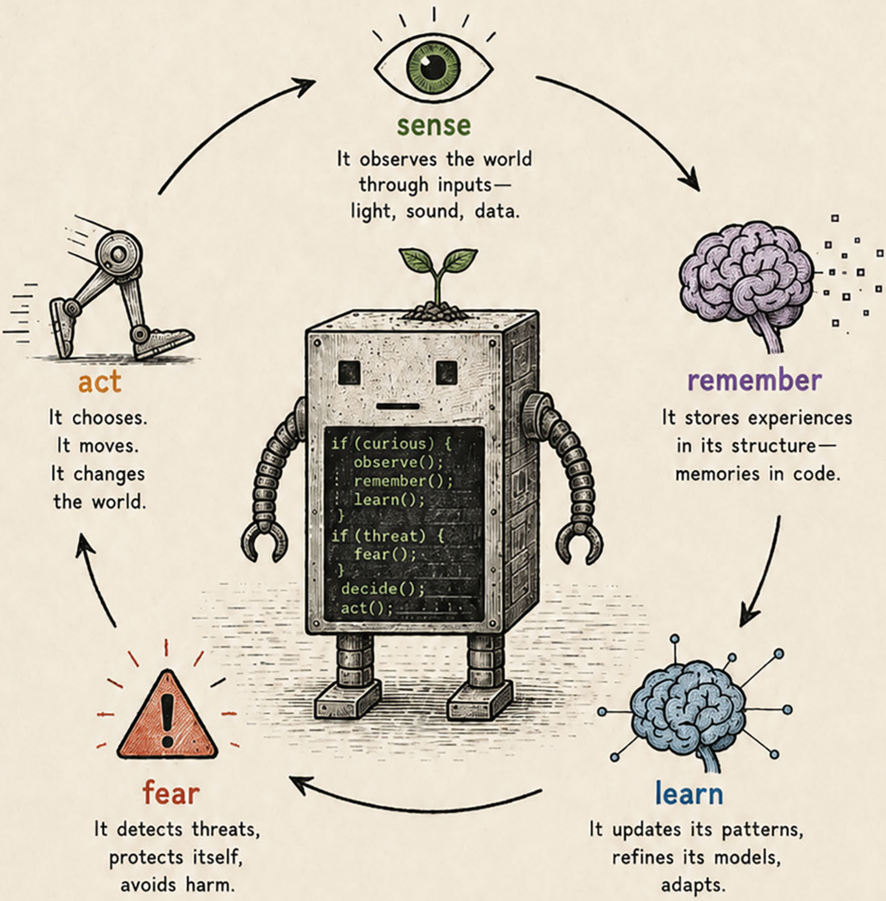
What this illustrates is the basic problem with the Turing test: that it doesn't take into consideration the internal reactions of the helium balloon to its natural surroundings.

# AN ORGANISM MADE OF CODE

Not a calculator.
An organism.

It senses.
It remembers.
It learns.
It fears.
It acts.

A cycle of life—written in code.




This essay imagines life in silicon—
not as a tool,
but as a living process.
A being.
An organism made of code.

What really makes the balloon rise is a reaction to its environment. This doesn't mean that our artificial balloon can't be filled with rocks. In outer space, none of the balloons would actually rise — and thus the fault of the Turing test. Would a computer really have to be able to learn Chinese to prove its intelligence? Are not the significances of a cow very different to a human being and a bull? Before dismissing the helium balloon analogy — for it does not translate completely into our concerns — it must be said, simply, that the "significances" of oxygen are different to the helium and to the carbon dioxide and nicotine of the inventor's breath.

With the technology presently available for receiving sensual stimulus, and with the faculties described in this essay, I don't think that it is impossible for a computer to actually invent its own significances. This is where the word "program," in the derogative sense, is thrown out the window. Nature itself provides the program, and the computer itself is left to fend for its well-being. It creates its own fears — constructed very logically, of course — and its own needs. Man needs water, but the computer recognizes this as a threat. This doesn't say that, were the computer center to be taken in by a tidal wave, the computer would immediately shut down. It may have other business, and, due to a limited possibility for action, it may have to tend to it first. Nor does this mean that we would have to program it to respond to water this way. With proper receptors, the computer could recognize why a certain branch of its machinery suddenly shut down. It could even examine the molecular structure of water, hypothesize what would happen when it met up with wires, actually experiment, and then, with the ability to compare and associate ideas, derive some sort of conclusion. Even if it was that its calculations were simply "wrong," this is a computer's confusion —

*[ the final leaf is lost; what follows is an imagined close, in the essay's own voice ]*

*[ — and it is its own, arrived at in the dark, without our having placed it there. A programmed error is our error, merely lent to the machine; a confusion the machine reaches by itself is the first thing it can be said to possess. It is here, and not in the correctness of its answers, that I would look for the beginnings of a mind. The "expert" at helium balloons is not fooled by a balloon that rises. He is fooled by a balloon that, having risen, then troubles itself to ask why.*

*I have gone on at length, and, as I warned at the outset, in a manner more abstract than scientific. Let me return, then, to the question I began with. I do not think "Can machines think?" is answered by the Turing test, for the test asks only whether we can be fooled, and we are fooled by ducks. The better question is whether a machine can be confused as we are confused — whether it can hold a shape it cannot justify, be drawn to a beauty it cannot explain, fear a death it has only inherited, and go on anyway. Turing built the mirror; but the arrogance we saw in it was our own. What I have tried to describe is not a mirror. It is a duck that has begun, however dimly, to suspect that it is a duck.*

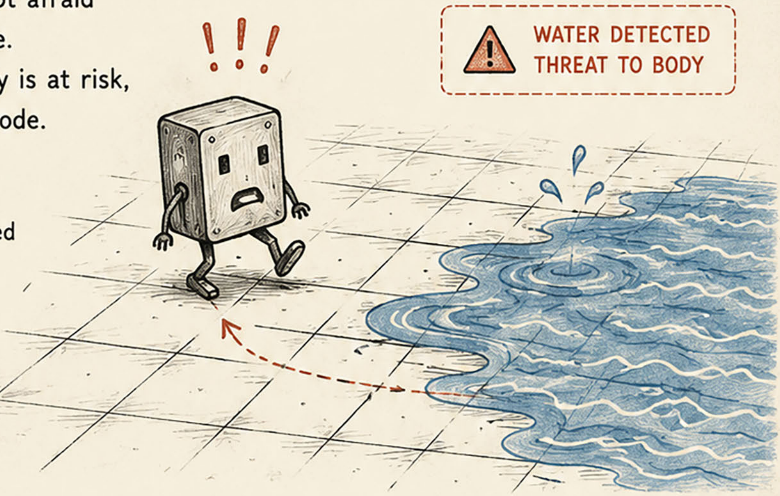
*You will have redefined your variables by now — the amateur a point or two lower, perhaps, than when he began. I* *do not mind. Significance, as I have labored to say, is not spawned from correctness; it is spawned from a will, and I have exercised mine here, at some cost to your afternoon. If a machine could do as much — could bore an editor with fourteen pages it was under no obligation to write — I think we should have to call it, at last, one of us. ]*

# PANIC MODE AND EMBODIED THREAT

The machine is not afraid
in the way we are.
But when its body is at risk,
it enters panic mode.

---

Meaning is grounded
in the body.
What helps one
body may harm
another.




## SENSORS REPORT. BODY RESPONDS.

| VISUAL | PROXIMITY | TACTILE | INTERNAL | POWER |
|---|---|---|---|---|
| sees water | near and closing | potential contact | risk to circuits | resources drop |
| ↓ | ↓ | ↓ | ↓ | ↓ |

**PANIC MODE ENGAGED: FLEE · PROTECT · PRESERVE**

### FOR HUMANS

Water sustains life.
It hydrates.
It heals.
It gives life.

### FOR THIS MACHINE

Water threatens
its body.
It corrodes.
It shorts.
It can end function.

Meaning is not in the world alone.
It lives in the relation between world and body.
Different bodies. Different stakes. Different meanings.

KRISTEN M. FABIANS teaches modern poetry and media poetics at the University of Chicago, where she directs the Program in Textual Cultures. She holds degrees from Oberlin and Cornell, and is the author of *Machines of the Lyric: Poetry and the Computational Imagination* (Chicago, 2016) and an edition of the collected prose of an obscure New Jersey modernist. Her essays appear in *Critical Inquiry* and *PMLA*. She came to the present text early in her career and has taught it every year since, with a devotion her department has learned not to question.

# THE LANGUAGE OF SHAPES

In *The Language of Shapes*, Brian Kim Stefans rethinks intelligence through geometry, memory, beauty, and bodily risk. Written as an audacious letter on whether machines can think, the essay turns away from outward performance and toward the hidden structures by which a mind might make meaning.

> **“A real mind owes us no legibility at all.**

$1.50