undermines itself in a Gödelian way, and I am incomplete for that reason. Cases (1) and (2) are predicated on my being 100 per cent consistent—a very unlikely state of affairs. More likely is that I am inconsistent—but that's worse, for then inside me there are contradictions, and how can I ever understand that?

Consistent or inconsistent, no one is exempt from the mystery of the self. Probably we are all inconsistent. The world is just too complicated for a person to be able to afford the luxury of reconciling all of his beliefs with each other. Tension and confusion are important in a world where many decisions must be made quickly. Miguel de Unamuno once said, "If a person never contradicts himself, it must be that he says nothing." I would say that we all are in the same boat as the Zen master who, after contradicting himself several times in a row, said to the confused Doko, "I cannot understand myself."

Gödel's Theorem and Personal Nonexistence

Perhaps the greatest contradiction in our lives, the hardest to handle, is the knowledge "There was a time when I was not alive, and there will come a time when I am not alive." On one level, when you "step out of yourself" and see yourself as "just another human being", it makes complete sense. But on another level, perhaps a deeper level, personal nonexistence makes no sense at all. All that we know is embedded inside our minds, and for all that to be absent from the universe is not comprehensible. This is a basic undeniable problem of life; perhaps it is the best metaphorical analogue of Gödel's Theorem. When you try to imagine your own nonexistence, you have to try to jump out of yourself, by mapping yourself onto someone else. You fool yourself into believing that you can import an outsider's view of yourself into you, much as TNT "believes" it mirrors its own metatheory inside itself. But TNT only contains its own metatheory up to a certain extent—not fully. And as for you, though you may imagine that you have jumped out of yourself, you never can actually do so—no more than Escher's dragon can jump out of its native two-dimensional plane into three dimensions. In any case, this contradiction is so great that most of our lives we just sweep the whole mess under the rug, because trying to deal with it just leads nowhere.

Zen minds, on the other hand, revel in this irreconcilability. Over and over again, they face the conflict between the Eastern belief: "The world and I are one, so the notion of my ceasing to exist is a contradiction in terms" (my verbalization is undoubtedly too Westernized—apologies to Zenists), and the Western belief: "I am just part of the world, and I will die, but the world will go on without me."

Science and Dualism

Science is often criticized as being too "Western" or "dualistic"—that is, being permeated by the dichotomy between subject and object, or observer

[...] it is true that up until this century, science was [...]d with things which can be readily distinguished from [...]vers—such as oxygen and carbon, light and heat, stars [...]rations and orbits, and so on—this phase of science was a [...] to the more modern phase, in which life itself has come [...]. Step by step, inexorably, "Western" science has moved [...]ion of the human mind—which is to say, of the observer. [...]nce research is the furthest step so far along that route. [...] along, there were two major previews of the strange consequences of the mixing of subject and object in science. One was the revolution of quantum mechanics, with its epistemological problems involving the interference of the observer with the observed. The other was the mixing of subject and object in metamathematics, beginning with Gödel's Theorem and moving through all the other limitative Theorems we have discussed. Perhaps the next step after AI will be the self-application of science: science studying itself as an object. This is a different manner of mixing subject and object—perhaps an even more tangled one than that of humans studying their own minds.

By the way, in passing, it is interesting to note that all results essentially dependent on the fusion of subject and object have been limitative results. In addition to the limitative Theorems, there is Heisenberg's uncertainty principle, which says that measuring one quantity renders impossible the simultaneous measurement of a related quantity. I don't know why all these results are limitative. Make of it what you will.

Symbol *vs.* Object in Modern Music and Art

Closely linked with the subject-object dichotomy is the symbol-object dichotomy, which was explored in depth by Ludwig Wittgenstein in the early part of this century. Later the words "use" and "mention" were adopted to make the same distinction. Quine and others have written at length about the connection between signs and what they stand for. But not only philosophers have devoted much thought to this deep and abstract matter. In our century both music and art have gone through crises which reflect a profound concern with this problem. Whereas music and painting, for instance, have traditionally expressed ideas or emotions through a vocabulary of "symbols" (i.e. visual images, chords, rhythms, or whatever), now there is a tendency to explore the capacity of music and art to *not* express anything—just to *be*. This means to exist as pure globs of paint, or pure sounds, but in either case drained of all symbolic value.

In music, in particular, John Cage has been very influential in bringing a Zen-like approach to sound. Many of his pieces convey a disdain for "use" of sounds—that is, using sounds to convey emotional states—and an exultation in "mentioning" sounds—that is, concocting arbitrary juxtapositions of sounds without regard to any previously formulated code by which a listener could decode them into a message. A typical example is "Imaginary Landscape no. 4", the polyradio piece described in Chapter VI. I may not

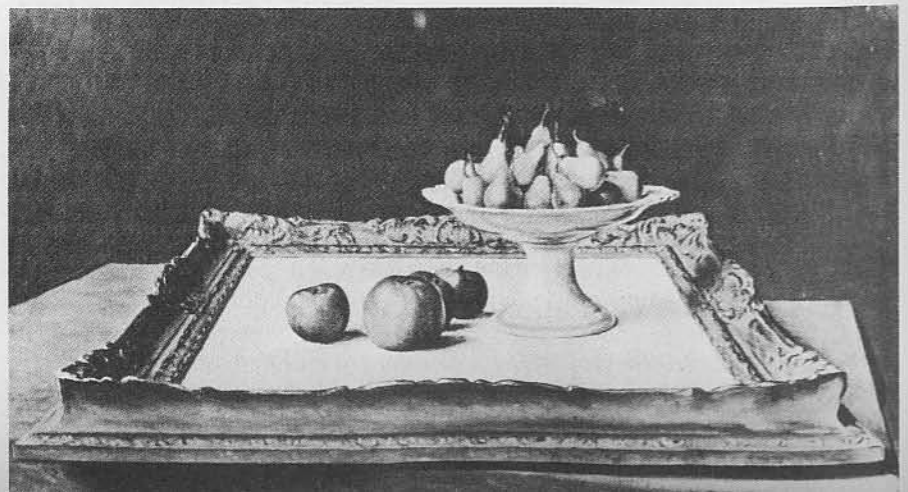
be doing Cage justice, but to me it seems that much of his work has been directed at bringing meaninglessness into music, and in some sense, at making that meaninglessness have meaning. Aleatoric music is a typical exploration in that direction. (Incidentally, chance music is a close cousin to the much later notion of "happenings" or "be-in"'s.) There are many other contemporary composers who are following Cage's lead, but few with as much originality. A piece by Anna Lockwood, called "Piano Burning", involves just that—with the strings stretched to maximum tightness, to make them snap as loudly as possible; in a piece by LaMonte Young, the noises are provided by shoving the piano all around the stage and through obstacles, like a battering ram.

Art in this century has gone through many convulsions of this general type. At first there was the abandonment of representation, which was genuinely revolutionary: the beginnings of abstract art. A gradual swoop from pure representation to the most highly abstract patterns is revealed in the work of Piet Mondrian. After the world was used to nonrepresentational art, then surrealism came along. It was a bizarre about-face, something like neoclassicism in music, in which extremely representational art was "subverted" and used for altogether new reasons: to shock, confuse, and amaze. This school was founded by André Breton, and was located primarily in France; some of its more influential members were Dalí, Magritte, de Chirico, Tanguy.

Magritte's Semantic Illusions

Of all these artists, Magritte was the most conscious of the symbol-object mystery (which I see as a deep extension of the use-mention distinction). He uses it to evoke powerful responses in viewers, even if the viewers do not verbalize the distinction this way. For example, consider his very strange variation on the theme of still life, called *Common Sense* (Fig. 137).

FIGURE 137. Common Sense, *by René Magritte (1945-46).*



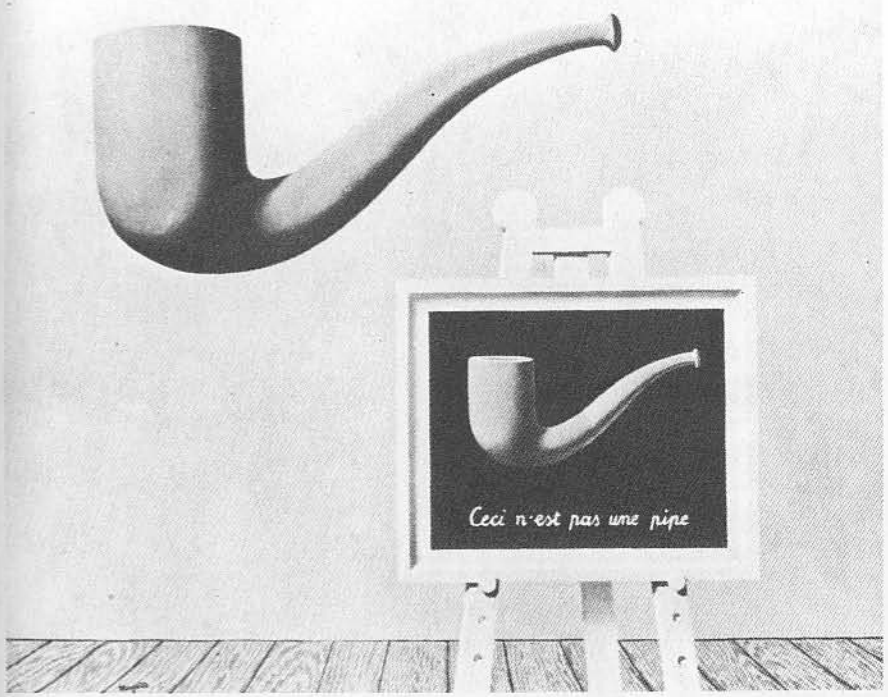


FIGURE 138. The Two Mysteries, *by René Magritte (1966).*

Here, a dish filled with fruit, ordinarily the kind of thing represented inside a still life, is shown sitting on top of a blank canvas. The conflict between the symbol and the real is great. But that is not the full irony, for of course the whole thing is itself just a painting—in fact, a still life with nonstandard subject matter.

Magritte's series of pipe paintings is fascinating and perplexing. Consider *The Two Mysteries* (Fig. 138). Focusing on the inner painting, you get the message that symbols and pipes are different. Then your glance moves upward to the "real" pipe floating in the air—you perceive that it is real, while the other one is just a symbol. But that is of course totally wrong: both of them are on the same flat surface before your eyes. The idea that one pipe is in a twice-nested painting, and therefore somehow "less real" than the other pipe, is a complete fallacy. Once you are willing to "enter the room", you have already been tricked: you've fallen for image as reality. To be consistent in your gullibility, you should happily go one level further down, and confuse image-within-image with reality. The only way not to be sucked in is to see both pipes merely as colored smudges on a surface a few inches in front of your nose. Then, and only then, do you appreciate the full meaning of the written message "Ceci n'est pas une pipe"—but ironically, at the very instant everything turns to smudges, the writing too turns to smudges, thereby losing its meaning! In other words, at that instant, the verbal message of the painting self-destructs in a most Gödelian way.

FIGURE 139. Smoke Signal. [*Drawing by the author.*]

The Air and the Song (Fig. 82), taken from a series by Magritte, accomplishes all that *The Two Mysteries* does, but in one level instead of two. My drawings *Smoke Signal* and *Pipe Dream* (Figs. 139 and 140) constitute "Variations on a Theme of Magritte". Try staring at *Smoke Signal* for a while. Before long, you should be able to make out a hidden message saying, "Ceci n'est pas un message". Thus, if you find the message, it denies itself—yet if you don't, you miss the point entirely. Because of their indirect self-snuffing, my two pipe pictures can be loosely mapped onto Gödel's G—thus giving rise to a "Central Pipemap", in the same spirit as the other "Central Xmaps": Dog, Crab, Sloth.

A classic example of use-mention confusion in paintings is the occurrence of a palette in a painting. Whereas the palette is an illusion created by the representational skill of the painter, the paints on the painted palette are literal daubs of paint from the artist's palette. The paint plays itself—it does not symbolize anything else. In *Don Giovanni,* Mozart exploited a related trick: he wrote into the score explicitly the sound of an orchestra tuning up. Similarly, if I want the letter 'I' to play itself (and not symbolize me), I put 'I' directly into my text; then I enclose 'I' between quotes. What results is ''I'' (not 'I', nor '''I'''). Got that?

FIGURE 140. Pipe Dream. [*Drawing by the author.*]

The "Code" of Modern Art

A large number of influences, which no one could hope to pin down completely, led to further explorations of the symbol-object dualism in art. There is no doubt that John Cage, with his interest in Zen, had a profound influence on art as well as on music. His friends Jasper Johns and Robert Rauschenberg both explored the distinction between objects and symbols by using objects as symbols for themselves—or, to flip the coin, by using symbols as objects in themselves. All of this was perhaps intended to break down the notion that art is one step removed from reality—that art speaks in "code", for which the viewer must act as interpreter. The idea was to eliminate the step of interpretation and let the naked object simply *be,* period. ("Period"—a curious case of use-mention blur.) However, if this was the intention, it was a monumental flop, and perhaps had to be.

Any time an object is exhibited in a gallery or dubbed a "work", it acquires an aura of deep inner significance—no matter how much the viewer has been warned *not* to look for meaning. In fact, there is a backfiring effect whereby the more that viewers are told to look at these objects without mystification, the more mystified the viewers get. After all, if a

wooden crate on a museum floor is just a wooden crate on a museum floor, then why doesn't the janitor haul it out back and throw it in the garbage? Why is the name of an artist attached to it? Why did the artist want to demystify art? Why isn't that dirt clod out front labeled with an artist's name? Is this a hoax? Am I crazy, or are artists crazy? More and more questions flood into the viewer's mind; he can't help it. This is the "frame effect" which art—Art—automatically creates. There is no way to suppress the wonderings in the minds of the curious.

Of course, if the purpose is to instill a Zen-like sense of the world as devoid of categories and meanings, then perhaps such art is merely intended to serve—as does intellectualizing about Zen—as a catalyst to inspire the viewer to go out and become acquainted with the philosophy which rejects "inner meanings" and embraces the world as a whole. In this case, the art is self-defeating in the short run, since the viewers *do* ponder about its meaning, but it achieves its aim with a few people in the long run, by introducing them to its sources. But in either case, it is not true that there is no code by which ideas are conveyed to the viewer. Actually, the code is a much more complex thing, involving statements about the absence of codes and so forth—that is, it is part code, part metacode, and so on. There is a Tangled Hierarchy of messages being transmitted by the most Zen-like art objects, which is perhaps why so many find modern art so inscrutable.

Ism Once Again

Cage has led a movement to break the boundaries between art and nature. In music, the theme is that all sounds are equal—a sort of acoustical democracy. Thus silence is just as important as sound, and random sound is just as important as organized sound. Leonard B. Meyer, in his book *Music, the Arts, and Ideas,* has called this movement in music "transcendentalism", and states:

> If the distinction between art and nature is mistaken, aesthetic valuation is irrelevant. One should no more judge the value of a piano sonata than one should judge the value of a stone, a thunderstorm, or a starfish. "Categorical statements, such as right and wrong, beautiful or ugly, typical of the rationalistic thinking of tonal aesthetics," writes Luciano Berio [a contemporary composer], "are no longer useful in understanding why and how a composer today works on audible forms and musical action."

Later, Meyer continues in describing the philosophical position of transcendentalism:

> . . . all things in all of time and space are inextricably connected with one another. Any divisions, classifications, or organizations discovered in the universe are arbitrary. The world is a complex, continuous, single event.[2] [Shades of Zeno!]

I find "transcendentalism" too bulky a name for this movement. In its place, I use "ism". Being a suffix without a prefix, it suggests an ideology

FIGURE 141. The Human Condition I, *by René Magritte (1933).*

without ideas—which, however you interpret it, is probably the case. And since "ism" embraces whatever is, its name is quite fitting. In "ism" the word "is" is half mentioned, half used; what could be more appropriate? Ism is the spirit of Zen in art. And just as the central problem of Zen is to unmask the self, the central problem of art in this century seems to be to figure out what art is. All these thrashings-about are part of its identity crisis.

We have seen that the use-mention dichotomy, when pushed, turns into the philosophical problem of symbol-object dualism, which links it to the mystery of mind. Magritte wrote about his painting *The Human Condition I* (Fig. 141):

> I placed in front of a window, seen from a room, a painting representing exactly that part of the landscape which was hidden from view by the painting. Therefore, the tree represented in the painting hid from view the tree situated behind it, outside the room. It existed for the spectator, as it were, simultaneously in his mind, as both inside the room in the painting, and outside in the real landscape. Which is how we see the world: we see it as being outside ourselves even though it is only a mental representation of it that we experience inside ourselves.[3]

Understanding the Mind

First through the pregnant images of his painting, and then in direct words, Magritte expresses the link between the two questions "How do symbols work?" and "How do our minds work?" And so he leads us back to the question posed earlier: "Can we ever hope to understand our minds/brains?"

Or does some marvelous diabolical Gödelian proposition preclude our ever unraveling our minds? Provided you do not adopt a totally unreasonable definition of "understanding", I see no Gödelian obstacle in the way of the eventual understanding of our minds. For instance, it seems to me quite reasonable to desire to understand the working principles of brains in general, much the same way as we understand the working principles of car engines in general. It is quite different from trying to understand any single brain in every last detail—let alone trying to do this for one's own brain! I don't see how Gödel's Theorem, even if construed in the sloppiest way, has anything to say about the feasibility of this prospect. I see no reason that Gödel's Theorem imposes any limitations on our ability to formulate and verify the general mechanisms by which thought processes take place in the medium of nerve cells. I see no barrier imposed by Gödel's Theorem to the implementation on computers (or their successors) of types of symbol manipulation that achieve roughly the same results as brains do. It is entirely another question to try and duplicate in a program some particular human's mind—but to produce an intelligent program at all is a more limited goal. Gödel's Theorem doesn't ban our reproducing our own level of intelligence via programs any more than it bans our reproducing our own level of intelligence via transmission of hereditary information in DNA, followed by education. Indeed, we have seen, in Chapter XVI, how remarkable Gödelian mechanism—the Strange Loop of proteins an DNA—is precisely what allows transmission of intelligence!

Does Gödel's Theorem, then, have absolutely nothing to offer us thinking about our own minds? I think it does, although not in the mystic and limitative way which some people think it ought to. I think that th process of coming to understand Gödel's proof, with its construction in volving arbitrary codes, complex isomorphisms, high and low levels o interpretation, and the capacity for self-mirroring, may inject some ric undercurrents and flavors into one's set of images about symbols an symbol processing, which may deepen one's intuition for the relationshi between mental structures on different levels.

Accidental Inexplicability of Intelligence?

Before suggesting a philosophically intriguing "application" of Gödel proof, I would like to bring up the idea of "accidental inexplicability" o intelligence. Here is what that involves. It could be that our brains, unlik car engines, are stubborn and intractable systems which we cannot neatl decompose in any way. At present, we have no idea whether our brains wi yield to repeated attempts to cleave them into clean layers, each of whic can be explained in terms of lower layers—or whether our brains will fo all our attempts at decomposition.

But even if we do fail to understand ourselves, there need not be an Gödelian "twist" behind it; it could be simply an accident of fate that ou brains are too weak to understand themselves. Think of the lowly giraffe for instance, whose brain is obviously far below the level required fo self-understanding—yet it is remarkably similar to our own brain. In fact the brains of giraffes, elephants, baboons—even the brains of tortoises o unknown beings who are far smarter than we are—probably all operate o basically the same set of principles. Giraffes may lie far below the threshol of intelligence necessary to understand how those principles fit together t produce the qualities of mind; humans may lie closer to that threshold—perhaps just barely below it, perhaps even above it. The point is that ther may be no *fundamental* (i.e., Gödelian) reason why those qualities are in comprehensible; they may be completely clear to more intelligent beings

Undecidability Is Inseparable from a High-Level Viewpoint

Barring this pessimistic notion of the accidental inexplicability of the brain what insights might Gödel's proof offer us about explanations of ou minds/brains? Gödel's proof offers the notion that a high-level view of a system may contain explanatory power which simply is absent on the lowe levels. By this I mean the following. Suppose someone gave you G, Gödel' undecidable string, as a string of TNT. Also suppose you knew nothing o Gödel-numbering. The question you are supposed to answer is: "Why isn'

Now it is possible to go considerably further in removing the pillars by which orientation is achieved. One step at a time . . . We begin by collapsing the whole array of boards into a single board. What is meant by this? There will be two ways of interpreting the board: (1) as pieces to be moved; (2) as rules for moving the pieces. On your turn, you move pieces—and perforce, you change rules! Thus, the rules constantly change themselves. Shades of Typogenetics—or for that matter, of real genetics. The distinction between game, rules, metarules, metametarules, has been lost. What was once a nice clean hierarchical setup has become a Strange Loop, Or Tangled Hierarchy. The moves change the rules, the rules determine the moves, round and round the mulberry bush . . . There are still different levels, but the distinction between "lower" and "higher" has been wiped out.

Now, part of what was inviolate has been made changeable. But there is still plenty that is inviolate. Just as before, there are conventions between you and your opponent by which you interpret the board as a collection of rules. There is the agreement to take turns—and probably other implicit conventions, as well. Notice, therefore, that the notion of different levels has survived, in an unexpected way. There is an Inviolate level—let's call it the *I-level*—on which the interpretation conventions reside; there is also a Tangled level—the *T-level*—on which the Tangled Hierarchy resides. So these two levels are still hierarchical: the I-level governs what happens on the T-level, but the T-level does not and cannot affect the I-level. No matter that the T-level itself is a Tangled Hierarchy—it is still governed by a set of conventions outside of itself. And that is the important point.

As you have no doubt imagined, there is nothing to stop us from doing the "impossible"—namely, tangling the I-level and the T-level by making the interpretation conventions themselves subject to revision, according to the position on the chess board. But in order to carry out such a "supertangling", you'd have to agree with your opponent on some further conventions connecting the two levels—and the act of doing so would create a *new* level, a new sort of inviolate level on top of the "supertangled" level (or underneath it, if you prefer). And this could continue going on and on. In fact, the "jumps" which are being made are very similar to those charted in the *Birthday Cantatatata,* and in the repeated Gödelization applied to various improvements on TNT. Each time you think you have reached the end, there is some new variation on the theme of jumping out of the system which requires a kind of creativity to spot.

The Authorship Triangle Again

But I am not interested in pursuing the strange topic of the ever more abstruse tanglings which can arise in self-modifying chess. The point of this has been to show, in a somewhat graphic way, how in any system there is always some "protected" level which is unassailable by the rules on other levels, no matter how tangled their interaction may be among themselves. An amusing riddle from Chapter IV illustrates this same idea in a slightly different context. Perhaps it will catch you off guard:

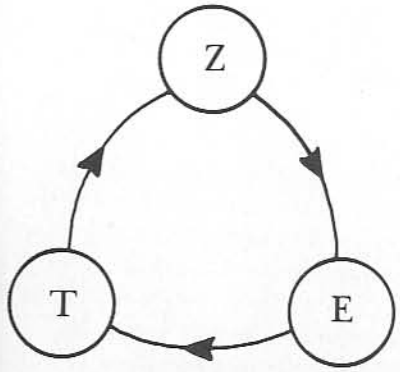


FIGURE 134. An "authorship triangle"

There are three authors—Z, T, and E. Now it happens that Z exists only in a novel by T. Likewise, T exists only in a novel by E. And strangely, E, too, exists only in a novel—by Z, of course. Now, is such an "authorship triangle" *really* possible? (See Fig. 134.)

Of course it's possible. But there's a trick . . . All three authors Z, T, E, ar themselves characters in another novel—by H! You can think of the Z-T-E triangle as a Strange Loop, Or Tangled Hierarchy; but author H is outside of the space in which that tangle takes place—author H is in an inviolate space. Although Z, T, and E all have access—direct or indirect—to each other, and can do dastardly things to each other in their various novels, none of them can touch H's life! They can't even imagine him—no more than you can imagine the author of the book *you're* a character in. If I were to draw author H, I would represent him somewhere off the page. Of course that would present a problem, since drawing a thing necessarily puts it *onto* the page . . . Anyway, H is really outside of the world of Z, T, and E, and should be represented as being so.

Escher's *Drawing Hands*

Another classic variation on our theme is the Escher picture of *Drawing Hands* (Fig. 135). Here, a left hand (LH) draws a right hand (RH), while at the same time, RH draws LH. Once again, levels which ordinarily are seen as hierarchical—that which draws, and that which is drawn—turn back on each other, creating a Tangled Hierarchy. But the theme of the Chapter is borne out, of course, since behind it all lurks the undrawn but drawing hand of M. C. Escher, creator of both LH and RH. Escher is outside of the two-hand space, and in my schematic version of his picture (Fig. 136), you can see that explicitly. In this schematized representation of the Escher picture, you see the Strange Loop, Or Tangled Hierarchy at the top; also, you see the Inviolate Level below it, enabling it to come into being. One could further Escherize the Escher picture, by taking a photograph of a hand drawing it. And so on.

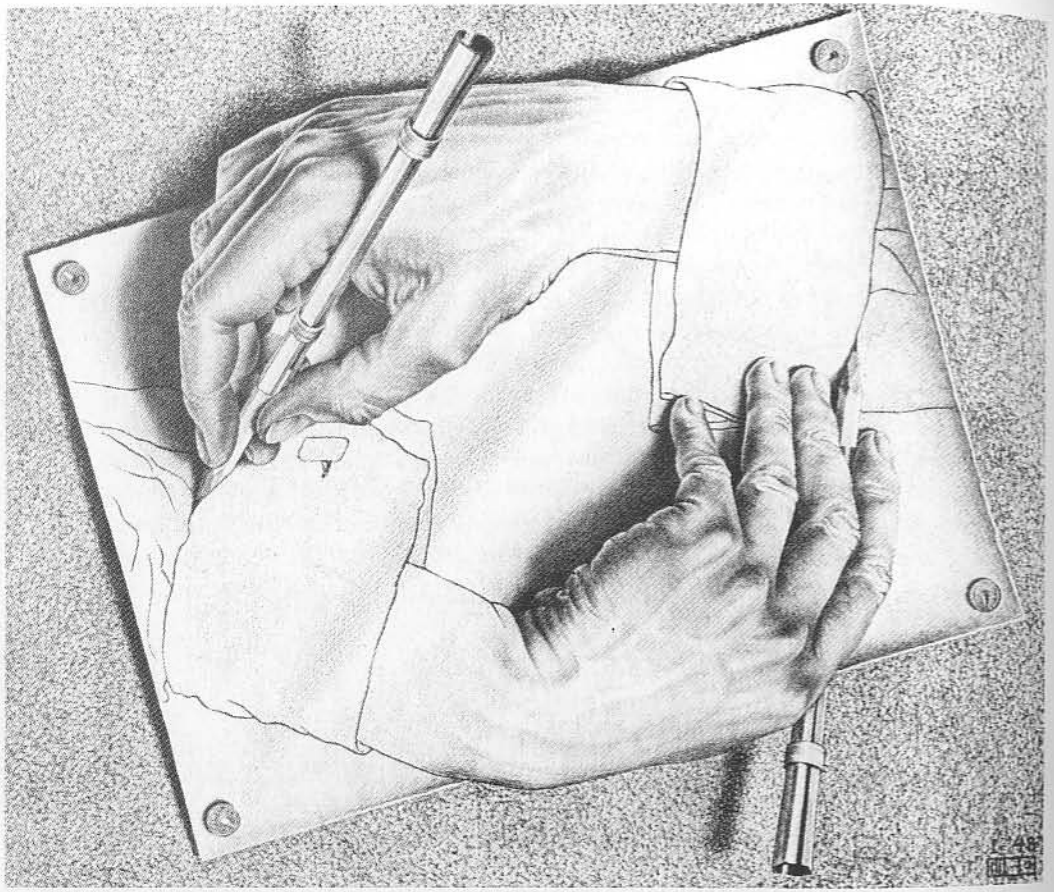


FIGURE 135. Drawing Hands, *by M. C. Escher (lithograph, 1948).*

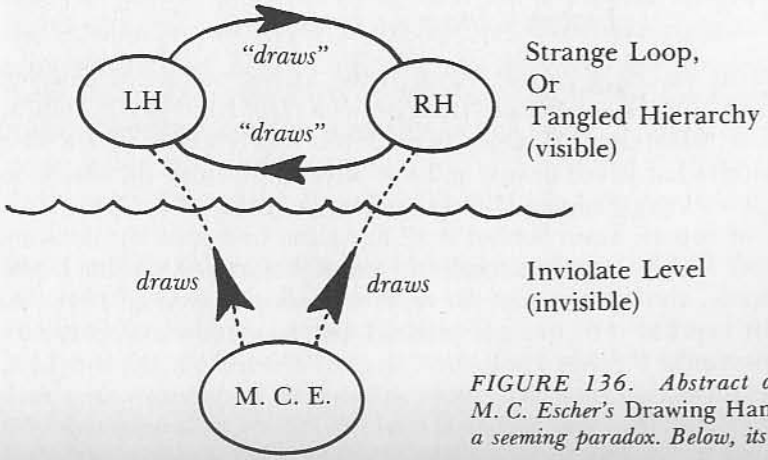


FIGURE 136. Abstract diagram of M. C. Escher's Drawing Hands. *On top, a seeming paradox. Below, its resolution.*

Brain and Mind: A Neural Tangle Supporting a Symbol Tangle

Now we can relate this to the brain, as well as to AI programs. In our thoughts, symbols activate other symbols, and all interact heterarchically. Furthermore, the symbols may cause each other to change internally, in the fashion of programs acting on other programs. The illusion is created, because of the Tangled Hierarchy of symbols, that *there is no inviolate level.* One thinks there is no such level because that level is shielded from our view.

If it were possible to schematize this whole image, there would be a gigantic forest of symbols linked to each other by tangly lines like vines in a tropical jungle—this would be the top level, the Tangled Hierarchy where thoughts really flow back and forth. This is the elusive level of *mind:* the analogue to LH and RH. Far below in the schematic picture, analogous to the invisible "prime mover" Escher, there would be a representation of the myriad neurons—the "inviolate substrate" which lets the tangle above it come into being. Interestingly, this other level is itself a tangle in a literal sense—billions of cells and hundreds of billions of axons, joining them all together.

This is an interesting case where a software tangle, that of the symbols, is supported by a hardware tangle, that of the neurons. But only the symbol tangle is a Tangled Hierarchy. The neural tangle is just a "simple" tangle. This distinction is pretty much the same as that between Strange Loops and feedback, which I mentioned in Chapter XVI. A Tangled Hierarchy occurs when what you presume are clean hierarchical levels take you by surprise and fold back in a hierarchy-violating way. The surprise element is important; it is the reason I call Strange Loops "strange". A simple tangle, like feedback, doesn't involve violations of presumed level distinctions. An example is when you're in the shower and you wash your left arm with your right, and then vice versa. There is no strangeness to the image. Escher didn't choose to draw hands drawing hands for nothing!

Events such as two arms washing each other happen all the time in the world, and we don't notice them particularly. I say something to you, then you say something back to me. Paradox? No; our perceptions of each other didn't involve a hierarchy to begin with, so there is no sense of strangeness.

On the other hand, where language does create strange loops is when it talks about itself, whether directly or indirectly. Here, something *in* the system jumps out and acts *on* the system, as if it were *outside* the system. What bothers us is perhaps an ill-defined sense of topological wrongness: the inside-outside distinction is being blurred, as in the famous shape called a "Klein bottle". Even though the system is an abstraction, our minds use spatial imagery with a sort of mental topology.

Getting back to the symbol tangle, if we look only at it, and forget the neural tangle, then we seem to see a self-programmed object—in just the same way as we seem to see a self-drawn picture if we look at *Drawing Hands* and somehow fall for the illusion, by forgetting the existence of Escher. For