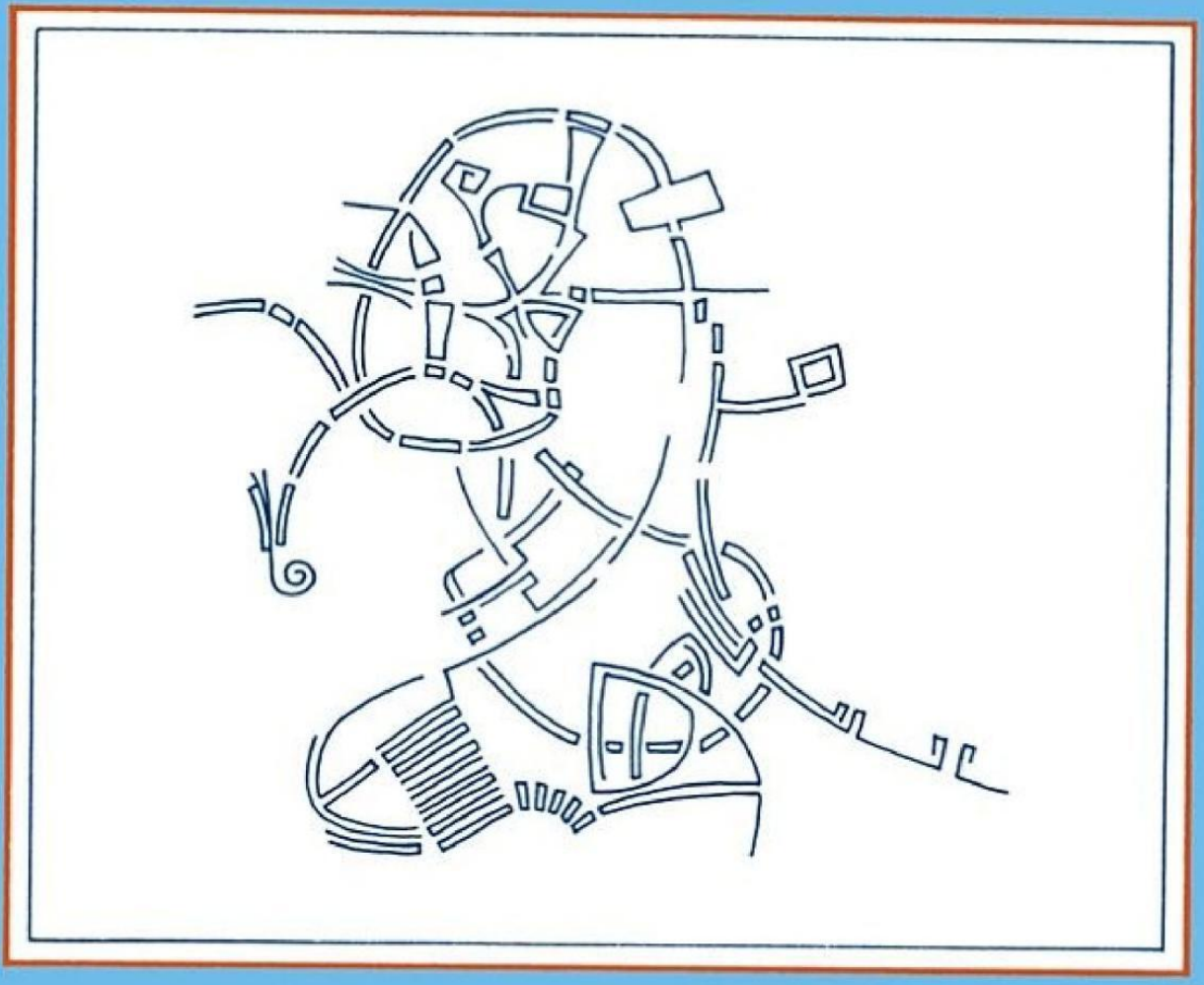
METAMAGICAL THEMAS:

Questing for the Essence of Mind and Pattern

DOUGLAS R. HOFSTADTER



An Interlocked Collection of Literary, Scientific, and Artistic Studies

Library of Congress Cataloging-in-Publication Data

Hofstadter, Douglas R., 1945-
Metamagical themas.

Bibliography: p. 802
Includes Index
1. Artificial intelligence. 2. Intellect. 3. Science—Philosophy. 4. Metamathematics. 5. Self (Philosophy) 6. Amusements. I. Title
Q335.H63 1985 001.53'5 83-46095
ISBN 978-0-786-72386-7

16 15 14 13

Short Contents

Long Contents

"sphexishness"—the inability to recognize and break out of one's ruts—in organisms on all levels of sophistication in the intelligence-creativity hierarchy.

Chapter 24: ***Analogies and Roles in Human and Machine Thinking.*** On the slippery things that must happen during the transport of ideas from one frame of reference to another, and how such phenomena can be boiled down into tiny domains where they can be studied in isolation.

[a]**Chapter 25:** ***Who Shoves Whom Around Inside the Careenium?*** An Achilles-Tortoise dialogue aiming at conveying clear imagery of the bidirectional causality that seems to louse up just about every effort to pinpoint what the meaning of the word "I" is.

[a]**Chapter 26:** ***Waking Up from the Boolean Dream*****, or,** ***Subcognition as Computation.*** A statement of my philosophical position about the dream underlying artificial intelligence, and the directions I hope and believe the field will move in.

Section VI: *Selection and Stability*

Chapter 27: ***The Genetic Code: Arbitrary?*** A view of biomolecules and their shenanigans through information-processing spectacles, aiming at characterizing in what sense the genetic code could be said to be arbitrary.

Chapter 28: ***Undercut, Flaunt, Pounce, and Mediocrity: Psychological Games with Numbers.*** Simple games of number choice pose amusing and serious dilemmas, and provide models by which to understand evolutionary processes.

Chapter 29: ***The Prisoner's Dilemma Computer Tournaments and the Evolution of Cooperation.*** In a recent worldwide computer tournament of the extremely simple game called "Prisoner's Dilemma", surprisingly moral strategies came out triumphant, a provocative discovery with metaphorical implications for biology, philosophy, theology, and politics.

Section VII: *Sanity and Survival*

Chapter 30: ***Dilemmas for Superrational Thinkers, Leading Up to a Luring Lottery.*** In which the outcome of a one-round Prisoner's Dilemma for money, participated in by a twentisome of the author's supposedly most rational friends, provokes the author into proposing an international lottery in *Scientific American.*

Chapter 31: ***Irrationality Is the Square Root of All Evil.*** In which the results of the preceding lottery are reported, and a fairly gloomy prognosis for humanity as a whole is extrapolated therefrom.

[a]**Chapter 32:** ***The Tale of Happiton.*** A short story about apathy and activism in a hypothetical but typical American small town when the townsfolk discover they are facing an ominous situation.

[a]**Chapter 33:** ***The Tumult of Inner Voices,*** **or,** ***What Is the Meaning of the Word "I"?*** A report from inside a mind split into many voices, on how they coexist and form a stable "soul" whose goals include intellectual, personal, and global causes.

List of Illustrations

Note: All the gridfonts and many of the more geometric and regular figures—especially in Chapters 14 and 24—were produced by the author on an Apple Macintosh using MacPaint. Unless otherwise indicated in their captions, all other figures were done by the author using conventional implements, such as felt-tip pens.

9–7. 2-against-3 rhythm in Chopin's Waltz Op. 42.
9–8. "A tricky bit of polyrhythm" in Chopin's F-minor Ballade.
10–1. *Fylfot Flipflop* [Fred Watts, William Huff studio].
10–2. *Crossover* [Richard Lane, William Huff studio].
10–3. *Dizzy Bee* [Richard Mesnik, William Huff studio].
10–4. *Consternation* [Scott Grady, William Huff studio].
10–5. *Oddity out of Old Oriental Ornament* [Francis O'Donnell, William Huff studio].
10–6. *Y Knot* [Leland Chen, William Huff studio].
10–7. *Crazy Cogs* [Arne Larson, William Huff studio].
10–8. *Trifoliolate* [Glen Paris, William Huff studio].
10–9. *Arabesque* [Joel Napach, William Huff studio].
10–10. *Razor Blades* [unknown, William Huff studio].
10–11. *Cucaracha* [Jorge Gutiérrez, William Huff studio].
10–12. *Beecombing Blossoms* [Laird Pylkas, William Huff studio].
10–13. *Clearing the Thicket* [Vincent Marlowe, William Huff studio].
10–14. Mondrian's *Composition with Lines,* with three computer-made variants.
10–15. *I at the Center* [David Oleson, William Huff studio].
11–1. One page from David Moser's "Metaculture Comics".
12–1. Psalm 23 in a gently drifting typeface, as realized by Donald Knuth using METAFONT.
12–2. Sixteen human faces, giving a sense of the vast space of possibilities.
12–3. 56 ways of realizing Platonic 'A', giving a sense of the vast space of possibilities.
12–4. 23 ways of realizing Platonic "hēi", giving a sense of the vast space of possibilities.
12–5. Implicospheres, alone and overlapping, with Mrs. Miniver's problem thrown in for fun.
13–1. Baskerville and Helvetica contrasted.
13–2. A series of diverse typefaces, giving a sense of the vast space of possibilities.
13–3. Three simple-seeming offshoots of Helvetica.
13–4. Three settings of a hypothetical knob controlling a typeface's "swirliness".
13–5. The complexities of role-filling in letterforms.
13–6. Several ambigrams by the author.
13–7. Shapes that spark competition between neighboring typographical niches.
13–8. The vertical and horizontal problems of Letter and Spirit.
13–9. Six typefaces by Hermann Zapf, giving a sense of the vast space of possibilities.
13–10. Transalphabetic leaps by the spirits of various typefaces.
13–11. The hazy boundary between book faces and display faces.
13–12. A Helvetica 'a' and an Italia 'g'.
13–13. Esthetically self-appraising Chinese sentences produced by the Hàn Zì program.

Section IV.

Section title page. A Whirly alphabet-canon (see "Notes on the Cover").
Introduction page. The gridfont called "Grecian Urn".

14–1. A Magic Cube, solved and scrambled.
14–2. Cubie types; the Cube's mechanism revealed; dismantling and reassembly methods for the Cube.
14–3. Labeling of cubies and moves.
14–4. The effects of various operators represented as cycles of cubies.
14–5. A zoological 3-cycle.
14–6. Diagrams for the proof that flippancy and twist are constrained.
14–7. The principle of conjugates.
14–8. The characteristic pattern of the Slice Group.

14–9. Basic “pretty patterns” on the Cube.
14–10. The Magic Domino.
14–11. Alternate colorings for the Cube.
15–1. A cube and a non-cube.
15–2. Variations on the theme of The Cube, giving a sense of the vast space of possibilities.
15–3. The Pyraminx: its modes of twisting and move notation.
15–4. The four types of piece in a Pyraminx.
15–5. The twisting mode of the Master Pyraminx.
15–6. The Magic Octahedron and its connection to the Magic Cube.
15–7. The Skewb.
15–8. The IncrediBall.
15–9. Operators for a planar two-circle twisting puzzle.
15–10. Four puzzles by Gabriel Lorente.
15–11. A 90-year-old intersecting-rings puzzle.
15–12. Generalizable nomenclature for moves on many cube-like puzzles.
15–13. A scrambled globe.
16–1. Folded functions in the unit square.
16–2. A parabolic function and its stable fixed point.
16–3. Another parabolic function and its stable attractor (a 2-cycle).
16–4. The fourfold iterate of a parabolic function.
16–5. The distribution of stable attractors at critical parameter values.
16–6. The smooth transition, via period-doubling, from order to chaos.
16–7. Beautiful vortices created when a stick is drawn through a fluid.
16–8. A firefly tracing out Duffing’s equation, observed by day and by night.
16–9. The strange attractor of Hénon.
16–10. The strange attractor of Duffing’s equation.
17–1. An Inner Glazunkian porpuquine trotting down the dusty road.
18–1. A trace of a recursive function call in Lisp.
18–2. The Tower of Brahma puzzle.
19–1. A pint-size version of the Tower of Brahma puzzle.
20–1. Classical and quantum-mechanical versions of the two-slit experiment.
20–2. Schrödinger’s cat in a quantum-mechanical superposition of states.
20–3. A robot in an emotional superposition of states.

Section V.

Section title page. A Whirly alphabet (see “Notes on the Cover”).
Introduction page. The gridfont called “Buxtehude”.

22–1. Happy programs dancing with happy interrogators, after a grueling Turing Test.
24–1. “Do this!”, starring Tom and Annie.
24–2. Son of “Do this!”, starring Tom and Annie, and introducing Fanny and Elephannie.
24–3. A plot of intrinsic saliencies in the Platonic alphabet of the Copycat universe.
24–4. Renormalization in Copycat; *or,* The world as seen by *q*.
24–5. Relative survival values of various foods and various analogies.
24–6. A graphical comparison of two answers to a Copycat analogy question.
24–7. The stepping-stone metaphor for translation between languages.
24–8. Three lattices for chess-like games.
24–9. Various conceptions of the “essence of knight’s move”.
24–10. Colored lattices and the slippable knight’s move.
24–11. Boards for chass and chäss (or chæss).

24–12. Cramming Helvetica ‘a’ into ever-tighter dot matrices with the aid of programs of differing levels of “intelligence”.
24–13. 87 ways of realizing Platonic ‘a’ in the Letter Spirit grid, giving a sense of the vast space of possibilities.
24–14. The horizontal and vertical problems as they arise in the Letter Spirit world.

Section VI.

Section title page. A Whirly alphabet (see “Notes on the Cover”).
Introduction page. The gridfont called “Benzene Right”.

27–1. The genetic code.
27–2. A protein jumping into its natural tertiary conformation.
27–3. An enzyme uniting two substrates in molecular matrimony.
27–4. A strand of messenger RNA.
27–5. Translation by Meri Boso and by ribosome.
27–6. Transfer RNA at three levels of abstraction.
27–7. A close-up of the intracellular translation process.
27–8. A mistaken guess about how amino acids and tRNA molecules get hooked up in the proper way.
27–9. DNA at two levels of abstraction.
27–10. The Central Dogma of molecular biology: “From DNA to RNA to proteins.”
28–1. Undercut seen both as a non-zero-sum game and as a zero-sum game.
28–2. Undercut contains the Prisoner’s Dilemma matrix many times over.
29–1. Various versions of the Prisoner’s Dilemma payoff matrix.

Section VII.

Section title page. A Whirly alphabet (see “Notes on the Cover”).
Introduction page. The gridfont called “Square Curl”.

30–1. Payoff matrices for a modified Prisoner’s Dilemma and for Wolf’s Dilemma.
31–1. US-SU symmetry in the ominous arms race.
33–1. The ostrich posture in the age of nuclear weapons.
33–2. World War III represented graphically in terms of World War II equivalents.

Notes on the Cover

A Spontaneous Essay on Whirly Art and Creativity

The drawing on the cover is a somewhat atypical example of a non-representational form of art I devised and developed over a period of years quite a long time ago, and which my sister Laura once rather light-heartedly dubbed "Whirly Art". The name stuck, for better or for worse. Generally speaking, I did Whirly Art on long thin strips of paper (available in rolls for adding machines) rather than on sheets of standard format. A typical piece of Whirly Art is five or six inches high and five or six feet long. Many are ten feet long, however, and some are as much as fifteen or even twenty feet in length. The one-dimensionality of Whirly Art was deliberate, of course: I was inspired by music and drew many visual fugues and canons. The time dimension was replaced by the long space dimension. I used the narrow width of the paper to represent something like pitch (although there was no strict mapping in any sense). A "voice" would be a single line tracing out some complex shape as it progressed in "time" along the paper. Several such voices could interact, and notions of what made "good" or "bad" visual harmony or counterpoint soon became intuitive to me.

The curvilinear motions constituting a single voice came from a blend of alphabets. At that time (the mid-60's), I was absolutely fascinated by the many writing systems found in and around India, exemplified by Tamil, Sinhalese, Kanarese, Telugu, Bengali, Hindi, Burmese, Thai, and many others. I studied some of them quite carefully, and even invented one of my own, based on the principles that most Indian scripts follow. It was natural that the motions my hand and mind were getting accustomed to would find their way into my visual fuguing. Thus was born Whirly Art.

Over the next several years, I did literally thousands of pieces of Whirly Art. Each one was totally improvised—in pen—so that there was no going back. A mistake was a mistake! Alternatively, a mistake could be interpreted as a very daring move from which it would be difficult, but not impossible, to recover gracefully. In other words, what seemed at first to be a disastrous mistake could turn into a joyful challenge! (I am sure that jazz improvisers will know exactly what I am talking about.) Sometimes, of course, I would fail, but other times I would succeed (at least by my own standards, since I was both performer and "listener").

Whirly Art became a (very) highly idiosyncratic language, with its own esthetic and traditions. However, traditions are made to be broken, and as soon as I spotted a tradition, I began experimenting around, violating it in various ways to see how I might move beyond my current state—how I might "jump out of the system". Style succeeded style, and I found myself paralleling the development of music. I moved from baroque Whirly Art (fugues, canons, and so forth) to "classical" Whirly Art, thence to "romantic" Whirly Art. After several years (it was now the late 60's), I reached the twentieth century, and found myself spiritually imitating such favorite composers of mine as Prokofiev and Poulenc. I did not copy any pieces specifically, but simply felt a kinship to those composers' style. Whirly Art is not translated music, but metaphorical music.

It is natural to wonder if I managed to jump beyond the twentieth century and make visual 21st-century music. That would have been quite a feat! Actually, in the early 70's I found that I simply was slowing down in production of Whirly Art. It had taken me seven years to recapitulate the history of Western music! At that point, I seemed to run out of creative juices. Of course, I could still make new Whirly Art then, as I can now—but I simply was less often inclined to do so. And today, I hardly ever do any Whirly Art, although the way that I draw curvy lines and letterforms bears the indelible marks of Whirly Art.

The piece on the cover, then, is atypical because it was done on an ordinary sheet of paper and has no direction of temporal flow. Also, there really is no concept of counterpoint in it. Still, it has something of a Whirly Art spirit. There are also seven Whirly alphabets in the book, one on each of the title pages of the seven sections. They are all somewhat atypical as well, but for slightly different reasons. Each was done on an ordinary sheet of paper but there is still always a clear flow, namely from 'A' to 'Z'. The real atypicality is the fact that genuine letters from a genuine alphabet are being used. I usually eschewed real letters, preferring to use shapes *inspired* by letters—shapes more complex and, well, "whirly" than most letters, even more so than Tamil or Sinhalese letters, which are pretty darn whirly.

Whirly Art is, I feel, quite possibly the most creative thing I have ever done. That, of course, is my opinion. Other people may disagree. It is a fairly strange and idiosyncratic form of art, however, and cannot be instantly understood. It has its own logic, related to the logics of musical harmony and counterpoint, Indian alphabets, gestalt perception, and who knows what else. I've kept it all quite literally in my closet for years—rolled up and piled into many paper bags and cardboard boxes. Because of its physical awkwardness, it is hard to show to people. But Whirly Art itself, and the experience of doing it, is an absolutely central fact about my way of looking at art, music, and creativity. Practically every time I write about creativity, some part of my mind is re-enacting Whirly Art experiences. In other words, a lot of my convictions about creativity come from self-observation rather than from scholarly study of the manuscripts or sketches of various composers or painters or writers or scientists. Of course, I have done some of that type of scholarship too, because I am fascinated by creativity in general—but I feel that to some extent "you don't really understand it unless you've done it", and so I rely a great deal on that personal experience. I feel that way that "I know what I'm talking about."

However, I would make a slightly stronger statement: Any two creative things that I've done seem to be, at some deep level, isomorphic. It's as if Whirly Art and mathematical discoveries and strange dialogues and little pieces of piano music and so on are all coming from a very similar core, and the same mechanisms are being exploited over and over again, only dressed up differently. Of course it's not all of the same quality: my *real* music is not as good as my visual music, for instance. But because I have this conviction that the core creativity behind all these things is really the same (at least in my own case), I am trying like mad to get at, and to lay bare, that core. For that reason I pursue ever-simpler domains in which I can feel myself doing "the same thing". In Chapter 24 of this book—in some sense the most creative Chapter, not surprisingly—I write about three of those domains: the Seek-Whence domain, the Copycat domain, and the Letter Spirit domain.

It is the Letter Spirit domain—"gridfonts" in particular—that is currently my most intense obsession. That domain came out of a lifelong fascination with our alphabet and other writing systems. I simply boiled away what I considered to be less interesting aspects of letterforms—I boiled and boiled—until I was left with what might be called the "conceptual skeletons" of letterforms. That is what gridfonts are about. People who have not shared my alphabetic fascination often underestimate at first the potential range of gridfonts, thinking that there might be a few and that's all. That is dead wrong: There are a huge number of them, and their variety is astounding.

As I look at the gridfonts I produce—and as I *feel* myself producing a gridfont—I feel that what I am doing is just Whirly Art all over again, in a new and ridiculously constrained way. The same

mechanisms of shape transformation, the same quest for grace and harmony, the same intuitions about what works and what doesn't, the same desire to "jump out of the system"—all this is truly the same. Doing gridfonts is therefore very exciting to me and provides a new proving ground for my speculations. The one advantage that gridfonts have over Whirly Art is that they are so preposterously constrained. This means that the possibilities for choice can be watched much more easily. It does not mean that a choice can be *explained* easily, but at least it can be watched. In a way, gridfonts are allowing me to re-experience the Whirly-Art period of my life, but with the advantage of several years' thinking about artificial intelligence and how I would like to try to make it come about. In other words, I can now hope that perhaps I can get a handle—a bit of one, anyway—on what is going on in creativity by means of computer modeling of it.

Since I feel that in a fundamental sense, Whirly-Art creativity is no deeper than gridfont creativity, the study of gridfont creation—more specifically, the computer modeling of gridfont creation—could reveal some things that I have sought for a long time. Therefore the next few years will be an important time for me—a time to see if I can really get at the essence, via modeling, of what my mind is doing when I create something that to me is excitingly novel.

This book, as it says on its cover and in the Introduction, deals with Mind and Pattern. To me, boiling things down to their conceptual skeletons is the royal road to truth (to mix metaphors rather horribly). I think that a lot of truth about Mind and Pattern lies waiting to be extracted in the tiny domains that I have carved out very painstakingly over the past seven years or so in Indiana. I urge you to keep these kinds of things in mind as you read this book. This "confession", coming as it does in a most unexpected place, is a very spontaneous one and probably captures as well as anything could the reason that my research is focused as it is, and the reason that I wrote this book.

Introduction

This book takes its title from the column I wrote in *Scientific American* between January 1981 and July 1983. In that two-and-a-half-year span, I produced 25 columns on quite a variety of topics. My choice of title deliberately left the focus of the column somewhat hazy, which was fine with me as well as with *Scientific American.* When Dennis Flanagan, the magazine's editor, wrote to me in mid-1980 to offer me the chance to write a column in that distinguished publication, he made it clear that what was desired was a bridge between the scientific and the literary viewpoints, something he pointed out Martin Gardner had always done, despite the ostensibly limiting title of *his* column, "Mathematical Games". Here is how Dennis put it in his letter:

> I might emphasize the flexible nature of the department we have been calling "Mathematical Games". As you know, under this title, Martin has written a great deal that is neither mathematical nor game-like. Basically, "Mathematical Games" has been Martin's column to talk about anything under the sun that interests him. Indeed, in our view, the main import of the column has been to demonstrate that a modern intellectual can have a range of interests that are not confined by such words as "scientific" or "literary". We hope that whoever succeeds Martin will feel free to cover his own broad range of interests, which are unlikely to be identical to Martin's.

What a refreshingly open attitude! So I was being asked to be the successor to Martin Gardner—but not necessarily to continue the same column. Rather than filling the same role as Martin had, I would merely occupy the same physical spot in the magazine.

I had been offered a unique opportunity to say pretty much anything I wanted to say to a vast, ready-made audience, in a prestigious context. *Carte blanche,* in short. What more could I ask? Even so, I had to deliberate long and hard about whether to take it, because I did not consider myself primarily a writer, but a thinker and researcher, and time taken in writing would surely be time taken away from research. The conservative pathway, following what was known, would have been to say no, and just do research. The adventurous pathway, exploring the new opportunity and forsaking some research, was tempting. Both were risky, since I knew that either way I would inevitably wonder, "How would things have gone had I decided the other way?" Moreover, I had no idea how long I might write my column, since that was not stipulated. It could go on for many years—or I could decide it was too much for me, and quit after a year.

In a way, I knew from the beginning that I would take the offer, I guess because I am basically more adventurous than I am conservative. But it was a little like purchasing new clothes: no matter how much you like them, you still want to see how you look in them before you buy them, so you put them on and parade around the store, looking at yourself in the mirror and asking whoever is with you what they think of it. So I talked it over with numerous people, and finally decided as I had expected: to take the offer.

* * *

For the first year, Martin Gardner and I alternated columns. I have to admit that even though I was utterly free to "be myself", I felt somewhat tradition-bound. True, I had metamorphosed his

title into my own title (see Chapter 1 for an explanation), but I was aware that readers of Martin's column would, naturally enough, be expecting a similar type of fare. It took a little while for me to test the waters, getting reader reactions and seeing if the magazine was satisfied with my performance, a performance very different in style from Martin's, after all. Needless to say, some readers were disappointed that I was not a clone of Martin Gardner, but others complimented me on how I had managed to keep the same level of quality while changing the style and content greatly. It was hard, knowing that people were constantly comparing me with someone very different from me. It was particularly hard when people who should have known better really confused my role with Martin's. For instance, as late as June 1983, at a conference on artificial intelligence, a colleague who spotted me came up to me and eagerly told me a math puzzle he'd just discovered and solved, hoping I would put it in my "Mathematical Games" column. How often did I have to tell people that my column was not called "Mathematical Games"!

I doubt that anyone loved Martin Gardner's column more than I did, or owed more to it. Yet I did not want my identity confused with someone else's. So writing this column and being in the shadow of someone superlative was not always easy. But I think I hit my stride and became comfortable with my new role after a few months.

In 1982, Martin retired, leaving the space entirely to me. It was a chore, to be sure, to get a column out each month, but it was also a lot of fun. In any case, what mattered to me the most was to do my best to make the column interesting and diverse and highly provocative. I took Dennis' offer quite literally, not restricting myself to purely scientific topics, but venturing into musical and literary topics as well.

After a year and a half, I was beginning to wonder how long I could sustain it without seriously jeopardizing my research. I decided to divide up my long list of prospective topics into categories: columns I would *love* to do, columns I would simply enjoy doing, and columns I could write with interest but no real passion. I found I had about a year's worth left in the first category, maybe another year's worth left in the second, and then a large number in the third. It seemed, then, that in another year or so it would be a good time to reassess the whole issue of writing the column. As it turned out, my thinking was quite consonant with evolving desires at the editorial level of the magazine. They were most interested in launching a new column to be devoted to the recreational aspects of computing, and our plans dovetailed well. My column could be phased out just as the new one was being phased in. And that is the way it came to pass, with two surprise columns by Martin Gardner filling the gap. My farewell to readers came as a postscript to Martin's final column, in September 1983.

Thus my era as a columnist came to an end. As I look back on it, I feel it lasted just about the right length of time: long enough to let me get a significant amount said, but not so long that it became a real drag on me. This way, at least, I got to explore that avenue that was so tempting, and yet it didn't radically alter the course of my life. So in sum, I am quite pleased with my stint at *Scientific American.* I am proud to have been associated with that venerable institution, and to have filled that unique slot for a time, especially coming right on the heels of someone of such high caliber.

* * *

The diversity of my columns is worth discussing for a moment. On the surface, they seem to wander all over the intellectual map—from sexism to music to art to nonsense, from game theory to artificial intelligence to molecular biology to the Cube, and more. But there is, I believe, a deep underlying unity to my columns. I felt that gradually, as I wrote more and more of them, regular readers would start to see the links between disparate ones, so that after a while, the coherence of the web would be quite clear. My image of this was always geometric. I envisioned my intellectual "home territory" as a rather large region in some conceptual space, a region that most people do

not see as a connected unit. Each new column was in a way a new "random dot" in that conceptual space, and as dots began peppering the space more fully over the months, the shape of my territory would begin to emerge more clearly. Eventually, I hoped, there would emerge a clear region associated with the name "Metamagical Themas".

Of course I wonder if my 25 1/2 columns are sufficient to convey the connectedness of my little patch of intellectual territory, or if, on the contrary, they would leave a question mark in the mind of someone who read them all in succession without any other explanation. Would it simply seem like a patchwork quilt, a curious potpourri? Truth to tell, I suspect that 25 columns are not quite enough, on their own. Probably the dots are too sparsely distributed to suggest the rich web of potential cross-connections there. For that reason, in drawing all my columns together to form a book, I decided to try to flesh out that space by including a few other recent writings of mine that might help to fill some of the more blatant gaps. There are seven such pieces included (indicated by asterisks in the table of contents). I believe they help to unify this book.

If someone were to ask me, "What is your new book about, in a word?", I would probably mutter something like "Mind and Pattern". That, in fact, was one title I considered for the column, way back when. Certainly it tells what most intrigues me, but it doesn't convey it quite vividly or passionately enough. Yes, I am a relentless quester after the chief patterns of the universe—central organizing principles, clean and powerful ways to categorize what is "out there". Because of this, I have always been pulled to mathematics. Indeed, even though I dropped the idea of being a professional mathematician many years ago, whenever I go into a new bookstore, I always make a beeline for the math section (if there is one). The reason is that I still feel that mathematics, more than any other discipline, studies the fundamental, pervasive patterns of the universe. However, as I have gotten older, I have come to see that there are inner mental patterns underlying our ability to conceive of mathematical ideas, universal patterns in human minds that make them receptive not only to the patterns of mathematics but also to abstract regularities of all sorts in the world. Gradually, over the years, my focus of interest has shifted to those more subliminal patterns of memory and associations, and away from the more formal, mathematical ones. Thus my interest has turned ever more to Mind, the principal apprehender of pattern, as well as the principal producer of certain kinds of pattern.

To me, the deepest and most mysterious of all patterns is music, a product of the mind that the mind has not come close to fathoming yet. In some sense, all my research is aimed at finding patterns that will help us to understand the mysteries of musical and visual beauty. I could be bolder and say, "I seek to discover what musical and visual beauty really are." However, I don't believe that those mysteries will ever be truly cleared up, nor do I wish them to be. I would like to understand things better, but I don't want to understand them perfectly. I don't wish the fruits of my research to include a mathematical formula for Bach's or Chopin's music. Not that I think it possible. In fact, I think the very idea is nonsense. But even though I find the prospect repugnant, I am greatly attracted by the effort to do as much as possible in that direction. Indeed, how could anyone hope to approach the concept of beauty without deeply studying the nature of formal patterns and their organizations and relationships to Mind? How can anyone fascinated by beauty fail to be intrigued by the notion of a "magical formula" behind it all, chimerical though the idea certainly is? And in this day and age, how can anyone fascinated by creativity and beauty fail to see in computers the ultimate tool for exploring their essence? Such ideas are the inner fire that propels my research and my writings, and they are the core of this book.

There is another aspect of my inner fire that is brought out in the writings here collected, particularly toward the end, but it pops up throughout. That is a concern with the global fate of humanity and the role of the individual in helping determine it. I have long been an activist, someone who periodically gets fired up by some cause and ardently works for it, exhorting everyone else I come across to get involved as well. I am a fierce believer in the value of passion and commitment to social causes, someone baffled and troubled by apathy. One of my personal

mottos is: "Apathy on the individual level translates into insanity at the mass level", a saying nowhere better exemplified than by today's insane dedication of so many human and natural resources to the building up of unimaginably catastrophic arsenals, all while mountains of humanity are starving and suffering in horrible ways. Everyone knows this, and yet the situation remains this way, getting worse day by day. We do live in a ridiculous world, and I would not wish to talk about the world without indicating my confusion and sadness, but also my vision and hope, concerning our shared human condition.

* * *

Inevitably, people will compare this book with my earlier books, *Gödel, Escher, Bach: an Eternal Golden Braid,* and *The Mind's I,* coedited with my friend Daniel Dennett. Let me try for a moment to anticipate them.

GEB was a unique sort of book—the detailed working-out of a single potent spark. It was a kind of explosion in my mind triggered by my re-falling in love with mathematical logic after a long absence. It was the first time I had tried to write anything long, and I pulled out all the stops. In particular, I made a number of experiments with style, especially in writing dialogues based on musical forms such as fugues and canons. In essence, *GEB* was one extended flash having to do with Kurt Gödel's famous incompleteness theorem, the human brain, and the mystery of consciousness. It is well described on its cover as "a metaphorical fugue on minds and machines".

The Mind's I is very different from *Gödel, Escher, Bach.* It is an extensively annotated anthology rather than the work of a single person. It is far more like a monograph than *GEB* is, in that it has a unique goal: to probe the mysteries of matter and consciousness in as vivid and jolting a way as possible, through stories that anyone can read and understand, followed by careful commentaries by Dan Dennett and myself. Its subtitle is "Fantasies and Reflections on Self and Soul".

One thing that *GEB* and *The Mind's I* have in common is their internal structure of alternation. *GEB* alternates between dialogues and chapters, while *The Mind's I* alternates between fantasies and reflections. I guess I like this contrapuntal mode, because it crops up again in the present volume. Here, I alternate between articles and postscripts.

If *GEB* is an elaborate fugue on one very complex theme, and *MI* is a collection of many variations on a theme, then perhaps *MT* is a fantasia employing several themes. If it were not for the postscripts, I would say that it was disjointed. However, I have made a great effort to tie together the diverse themes—Themas—by writing extensive commentaries that cast the ideas of each article in the light of other articles in the book. Sometimes the postscripts approach the length of the piece they are "post", and in one case (Chapter 24) the postscript is quite a bit longer than its source.

The reason for that particularly long postscript is that I decided to use it to describe some aspects of my own current research in artificial intelligence. There are other places as well in the book where I touch on my research ideas, though I never go into technical details. My main concern is to give a clear idea of certain central riddles about how minds work, riddles that I have run across over and over again in different guises. The questions I raise are difficult but I find them as beguiling as mathematical ones. In any case, this book will give readers a better understanding of how my research and the rest of my ideas fit together.

* * *

One aspect of this book that, I must admit, sometimes makes me uneasy is the striking disparity in the seriousness of its different topics. How can both Rubik's Cube and nuclear Armageddon be discussed at equal length in one book by one author? Partly the answer is that life itself is a mixture of things of many sorts, little and big, light and serious, frivolous and formidable, and *Metamagical Themas* reflects that complexity. Life is not worth living if one can never afford to be

delighted or have fun.

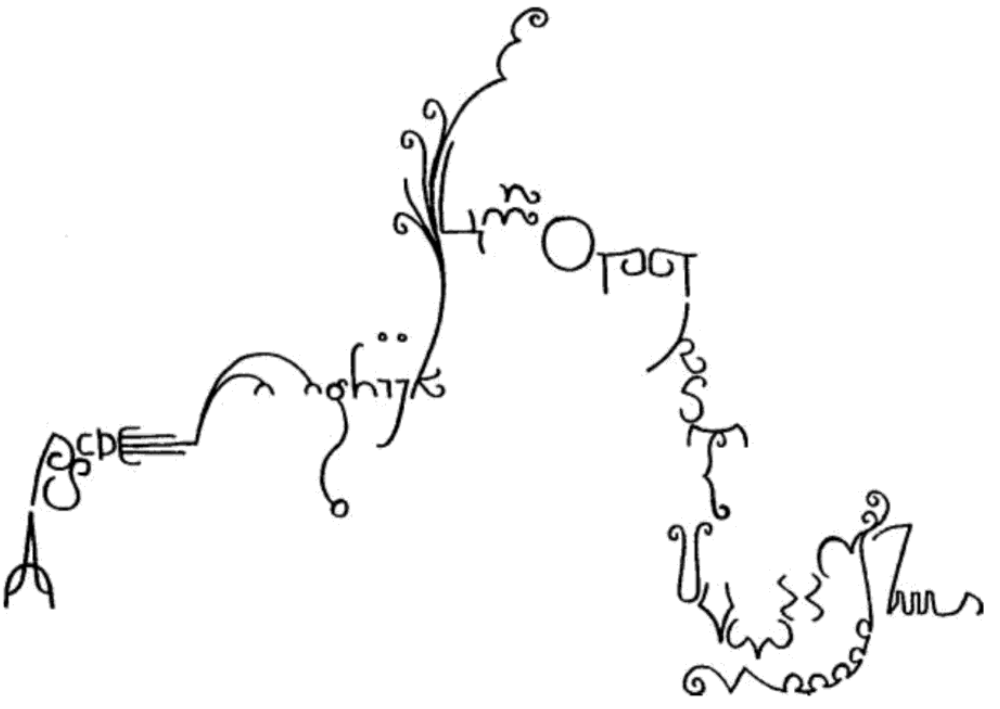
There is another way of explaining this huge gulf. Elegant mathematical structures can be as central to a serious modern worldview as are social concerns, and can deeply influence one's ways of thinking about anything—even such somber and colossal things as total nuclear obliteration. In order to comprehend that which is incomprehensible because it is too huge or too complex, one needs simpler models. Often, mathematics can provide the right starting point, which is why beautiful mathematical concepts are so pervasive in explanations of the phenomena of nature on the micro-level. They are now proving to be of great help also on a larger scale, as Robert Axelrod's lovely work on the Prisoner's Dilemma so impeccably demonstrates (see Chapter 29).

The Prisoner's Dilemma is poised about halfway between the Cube and Armageddon, in terms of complexity, abstraction, size, and seriousness. I submit that abstractions of this sort are direly needed in our times, because many people—even remarkably smart people—turn off when faced with issues that are too big. We need to make such issues graspable. To make them graspable and fascinating as well, we need to entice people with the beauties of clarity, simplicity, precision, elegance, balance, symmetry, and so on.

Those artistic qualities, so central to good science as well as to good insights about life, are the things that I have tried to explore and even to celebrate in *Metamagical Themas.* (It is not for nothing that the word "magic" appears inside the title!) I hope that *Metamagical Themas* will help people to bring more clarity, precision, and elegance to their thinking about situations large and small. I also hope that it will inspire people to dedicate more of their energies to global problems in this lunatic but lovable world, because we live in a time of unprecedented urgency. If we do not care enough now, future generations may not exist to thank us for their existence and for our caring.

Section I:

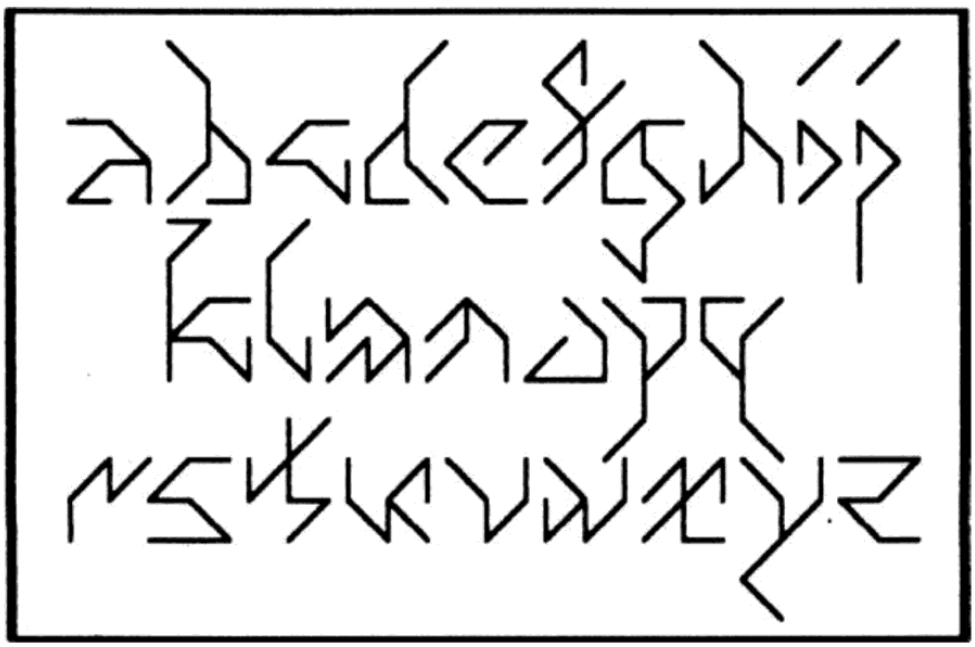
Snags and Snarls



Section 1:

Snags and Snarls

The title of this section conveys the image of problematical twistiness. The twists dealt with here are those whereby a system (sentence, picture, language, organism, society, government, mathematical structure, computer program, etc.) twists back on itself and closes a loop. A very general name for this is *reflexivity.* When realized in different ways, this abstraction becomes a concrete phenomenon. Examples are: self-reference, self-description, self-documentation, self-contradiction, self-questioning, self-response, self-justification, self-refutation, self-parody, self-doubt, self-definition, self-creation, self-replication, self-modification, self-amendment, self-limitation, self-extension, self-application, self-scheduling, self-watching, and on and on. In the following four chapters, these strange phenomena are illustrated in sentences and stories that talk about themselves, ideas that propagate themselves from mind to mind, machines that replicate themselves, and games that modify their own rules. The variety of these loopy tangles is quite remarkable, and the subject is far from being exhausted. Furthermore, although their connection with paradox may make reflexive systems seem no more than intellectual playthings, study of them is of great importance in understanding many mathematical and scientific developments of this century, and is becoming ever more central to theories of intelligence and consciousness, whether natural or artificial. Reflexivity will therefore make many return appearances in this book.



1

On Self-Referential Sentences

January, 1981

I never expected to be writing a column for *Scientific American.* I remember once, years ago, wishing I were in Martin Gardner's shoes. It seemed exciting to be able to plunge into almost any topic one liked and to say amusing and instructive things about it to a large, well-educated, and receptive audience. The notion of doing such a thing seemed ideal, even dreamlike. Over the next several years, by a series of total coincidences (which turned out to be not so total), I met one after another of Martin's friends. First it was Ray Hyman, a psychologist who studies deception. He introduced me to the magician Jerry Andrus. Then I met the statistician and magician Persi Diaconis and the computer wizard Bill Gosper. Then came Scott Kim, and soon afterward, the mathematician Benoît Mandelbrot. All of a sudden, the world seemed to be orbiting Martin Gardner. He was at the hub of a magic circle, people with exciting, novel, often offbeat ideas, people with many-dimensional imaginations. Sometimes I felt overawed by the whole remarkable bunch.

One day, five or so years ago, I had the pleasure of spending several hours with Martin in his house, discussing many topics, mathematical and otherwise. It was an enlightening experience for me, and it gave me a new view into the mind of someone who had contributed so much to my own mathematical education. Perhaps the most striking thing about Martin to me was his natural simplicity. I had been told that he is an adroit magician. This I found hard to believe, because one does not usually imagine someone so straightforward pulling the wool over anyone's eyes. However, I did not see him do any magic tricks. I simply saw his vast knowledge and love of ideas spread out before me, without the slightest trace of pride or pretense. The Gardners—Martin and his wife Charlotte—entertained me for the day. We ate lunch in the kitchen of their cozy three-story house. It pleased me somehow to see that there was practically no trace of mathematics or games or tricks in their simple but charming living room.

After lunch—sandwiches that Martin and I made while standing by the kitchen sink—we climbed the two flights of stairs to Martin's hideaway. With his old typewriter and all kinds of curious jottings in an ancient filing cabinet and his legendary library of three-by-five cards, he reminded me of an old-time journalist, not of the center of a constellation of mathematical eccentrics and game addicts, to say nothing of magicians, anti-occultists, and of course the thousands of readers of his column.

Occasionally we were interrupted by the tinkling of a bell attached to a string that led down the stairs to the kitchen, where Charlotte could pull it to get his attention. A couple of phone calls came, one from the logician and magician Raymond Smullyan, someone whose name I had known for a long time, but who I had no idea belonged to this charmed circle. Smullyan was calling to chat about a book he was writing on Taoism, of all things! For a logician to be writing about what seemed to me to be the most anti-logical of human activities sounded wonderfully paradoxical. (In

fact, his book *The Tao Is Silent* is delightful and remarkable.) All in all, it was a most enjoyable day.

Martin's act will be a hard one to follow. But I will not be trying to be another Martin Gardner. I have my own interests, and they are different from Martin's, although we have much in common. To express my debt to Martin and to symbolize the heritage of his column, I have kept his title "Mathematical Games" in the form of an anagram: "Metamagical Themas".

What does "metamagical" mean? To me, it means "going one level beyond magic". There is an ambiguity here: on the one hand, the word might mean "ultramagical"—magic of a higher order—yet on the other hand, the magical thing about magic is that what lies behind it is always *non*-magical. That's metamagic for you! It reflects the familiar but powerful adage "Truth is stranger than fiction." So my "Metamagical Themas" will, in Gardnerian fashion, attempt to show that magic often lurks where few suspect it, and, by the opposite token, that magic seldom lurks where many suspect it.

* * *

In his July, 1979 column, Martin wrote a very warm review of my book *Gödel, Escher, Bach: an Eternal Golden Braid.* He began the review with a short quotation from my book. If I had been asked to guess what single sentence he would quote, I would never have been able to predict his choice. He chose the sentence "This sentence no verb." It is a catchy sentence, I admit, but something about seeing it again bothered me. I remembered how I had written it one day a few years earlier, attempting to come up with a new variation on an old theme, but even at the time it had not seemed as striking as I had hoped it would. After seeing it chosen as the symbol of my book, I felt challenged. I said to myself that surely there must be much cleverer types of self-referential sentence. And so one day I wrote down quite a pile of self-referential sentences and showed them to friends, which began a mild craze among a small group of us. In this column, I will present a selection of what I consider to be the cream of that crop.

Before going further, I should explain the term "self-reference". Self-reference is ubiquitous. It happens every time anyone says "I" or "me" or "word" or "speak" or "mouth". It happens every time a newspaper prints a story about reporters, every time someone writes a book about writing, designs a book about book design, makes a movie about movies, or writes an article about self-reference. Many systems have the capability to represent or refer to themselves somehow, to designate themselves (or elements of themselves) within the system of their own symbolism. Whenever this happens, it is an instance of self-reference.

Self-reference is often erroneously taken to be synonymous with paradox. This notion probably stems from the most famous example of a self-referential sentence, the Epimenides paradox. Epimenides the Cretan said, "All Cretans are liars." I suppose no one today knows whether he said it in ignorance of its self-undermining quality or for that very reason. In any case, two of its relatives, the sentences "I am lying" and "This sentence is false", have come to be known as the *Epimenides paradox* or the *liar paradox.* Both sentences are absolutely self-destructive little gems and have given self-reference a bad name down through the centuries. When people speak of the evils of self-reference, they are certainly overlooking the fact that not every use of the pronoun "I" leads to paradox.

* * *

Let us use the Epimenides paradox as our jumping-off point into this fascinating land. There are many variations on the theme of a sentence that somehow undermines itself. Consider these two:

This sentence claims to be an Epimenides paradox, but it is lying.

This sentence contradicts itself—or rather—well, no, actually it doesn't!

What should you do when told, "Disobey this command"? In the following sentence, the

Epimenides quality jumps out only after a moment of thought: "This sentence contains exactly threee erors." There is a delightful backlash effect here.

Kurt Gödel's famous Incompleteness Theorem in metamathematics can be thought of as arising from his attempt to replicate as closely as possible the liar paradox in purely mathematical terms. With marvelous ingenuity, he was able to show that in any mathematically powerful axiomatic system *S* it is possible to express a close cousin to the liar paradox, namely, "This formula is unprovable within axiomatic system *S*."

In actuality, the Gödel construction yields a mathematical formula, not an English sentence; I have translated the formula back into English to show what he concocted. However, astute readers may have noticed that, strictly speaking, the phrase "this formula" has no referent, since when a *formula* is translated into an English *sentence*, that sentence is no longer a formula!

If one pursues this idea, one finds that it leads into a vast space. Hence the following brief digression on the preservation of self-reference across language boundaries. How should one translate the French sentence *Cette phrase en français est difficile à traduire en anglais*? Even if you do not know French, you will see the problem by reading a literal translation: "This sentence in French is difficult to translate into English." The problem is: To what does the subject ("This sentence in French") refer? If it refers to the sentence it is part of (which is not in French), then the subject is self-contradictory, making the sentence false (whereas the French original was true and harmless); but if it refers to the French sentence, then the meaning of "this" is strained. Either way, something disquieting has happened, and I should point out that it would be just as disquieting, although in a different way, to translate it as: "This sentence in English is difficult to translate into French." Surely you have seen Hollywood movies set in France, in which all the dialogue, except for an occasional *Bonjour* or similar phase, is in English. What happens when Cardinal Richelieu wants to congratulate the German baron for his excellent command of French? I suppose the most elegant solution is for him to say, "You have an excellent command of our language, *mon cher baron*", and leave it at that.

* * *

But let us undigress and return to the Gödelian formula and focus on its meaning. Notice that the concept of *falsity* (in the liar paradox) has been replaced by the more rigorously understood concept of *provability.* The logician Alfred Tarski pointed out that it is in principle impossible to translate the liar paradox exactly into any rigorous mathematical language, because if it were possible, mathematics would contain a genuine paradox—a statement both true and false—and would come tumbling down.

Gödel's statement, on the other hand, is not paradoxical, though it constitutes a hair-raisingly close approach to paradox. It turns out to be true, and for this reason, it is unprovable in the given axiomatic system. The revelation of Gödel's work is that in *any* mathematically powerful and consistent axiomatic system, an endless series of true but unprovable formulas can be constructed by the technique of self-reference, revealing that somehow the full power of human mathematical reasoning eludes capture in the cage of rigor.

In a discussion of Gödel's proof, the philosopher Willard Van Orman Quine invented the following way of explaining how self-reference could be achieved in the rather sparse formal language Gödel was employing. Quine's construction yields a new way of expressing the liar paradox. It is this:

> "yields falsehood, when appended to its quotation." yields falsehood, when appended to its quotation.

This sentence describes a way of constructing a certain typographical entity—namely, a phrase appended to a copy of itself in quotes. When you carry out the construction, however, you see that the end product is the sentence itself—or a perfect copy of it. (There is a resemblance here to the

way self-replication is carried out in the living cell.) The sentence asserts the falsity of the constructed typographical entity, namely itself (or an indistinguishable copy of itself). Thus we have a less compact but more explicit version of the Epimenides paradox.

It seems that all paradoxes involve, in one way or another, self-reference, whether it is achieved directly or indirectly. And since the credit for the discovery—or creation—of self-reference goes to Epimenides the Cretan, we might say: "Behind every successful paradox there lies a Cretan."

On the basis of Quine's clever construction we can create a self-referential question:

What is it like to be asked,
"What is it like to be asked, self-embedded in quotes after its comma?" self-embedded in quotes after its comma?

Here again, you are invited to construct a typographical entity that turns out, when the appropriate operations have been performed, to be identical with the set of instructions. This self-referential question suggests the following puzzle: What is a question that can serve as its own answer? Readers might enjoy looking for various solutions to it.

* * *

When a word is used to *refer* to something, it is said to be being *used.* When a word is *quoted,* though, so that one is examining it for its surface aspects (typographical, phonetic, etc.), it is said to be being *mentioned.* The following sentences are based on this famous use-mention distinction:

You can't have your use and mention it too.

You can't have "your cake" and spell it "too".

"Playing with the use-mention distinction" isn't "everything in life, you know".

In order to make sense of "this sentence", you will have to ignore the quotes in "it".

This is a sentence with "onions", "lettuce", "tomato", and "a side of fries to go".

This is a hamburger with vowels, consonants, commas, and a period at the end.

The last two are humorous flip sides of the same idea. Here are two rather extreme examples of self-referential use-mention play:

> Let us make a new convention: that anything enclosed in *triple* quotes—for example, "'No, I have decided to change my mind; when the triple quotes close, just skip directly to the period and ignore everything up to it'"—is not even to be read (much less paid attention to or obeyed).

> A ceux qui ne comprennent pas l'anglais, la phrase citée ci-dessous ne dit rien: "For those who know no French, the French sentence that introduced this quoted sentence has no meaning."

The bilingual example may be more effective if you know only one of the two languages involved.

Finally, consider this use-mention anomaly: "i should begin with a capital letter." This is a sentence referring to itself by the pronoun "I", a bit mauled, instead of through a pointing-phrase such as "this sentence"; such a sentence would seem to be arrogantly proclaiming itself to be an animate agent. Another example would be "I am not the person who wrote me." Notice how easily we understand this curious nonstandard use of "I". It seems quite natural to read the sentence this way, even though in nearly all situations we have learned to unconsciously create a mental model of some person—the sentence's speaker or writer—to whom we attribute a desire to communicate some idea. Here we take the "I" in a new way. How come? What kinds of cues in a sentence make us recognize that when the word "I" appears, we are supposed to think not about the author of the

sentence but about the sentence itself?

* * *

Many simplified treatments of Gödel's work give as the English translation of his famous formula the following: "I am not provable in axiomatic system *S*." The self-reference that is accomplished with such sly trickery in the formal system is finessed into the deceptively simple English word "I", and we can—in fact, we automatically do—take the sentence to be talking about itself. Yet it is hard for us to hear the following sentence as talking about itself: "I *already* took the garbage out, honey."

The ambiguous referring possibilities of the first-person pronoun are a source of many interesting self-referential sentences. Consider these:

I am not the subject of this sentence.

I am jealous of the first word in this sentence.

Well, how about that—this sentence is about me!

I am simultaneously writing and being written.

This raises a whole new set of possibilities. Couldn't "I" stand for the writing instrument ("I am not a pen"), the language ("I come from Indo-European roots"), the paper ("Cut me out, twist me, and glue me to form a Möbius strip, please")? One of the most involved possibilities is that "I" stands not for the physical tokens we perceive before us but for some more ethereal and intangible essence, perhaps the *meaning* of the sentence. But then, what is meaning? The next examples explore that idea:

I am the meaning of this sentence.

I am the thought you are now thinking.

I am thinking about myself right now.

I am the set of neural firings taking place in your brain as you read the set of letters in this sentence and think about me.

This inert sentence is my body, but my soul is alive, dancing in the sparks of your brain.

The philosophical problem of the connections among Platonic ideas, mental activity, physiological brain activity, and the external symbols that trigger them is vividly raised by these disturbing sentences.

This issue is highlighted in the self-referential question, "Do you think anybody has ever had *precisely this thought* before?" To answer the question, one would have to know whether or not two different brains can *ever* have precisely the same thought (as two different computers can run precisely the same program). An illustration of this possibility may be found in Figure 24-2. I have often wondered: Can *one* brain have the same thought more than once? Is a thought something Platonic, something whose essence exists independently of the brain it is occurring in? If the answer is "Yes, thoughts are brain-independent", then the answer to the self-referential question would also be yes. If it is not, then no one could ever have had the same thought before—not even the person thinking it!

Certain self-referential sentences involve a curious kind of communication between the sentence and its human friends:

You are under my control because I am choosing exactly what words you are made out of, and in what order.

No, *you* are under *my* control because you will read until you have reached the end of me.

Hey, down there—are you the sentence I am writing, or the sentence I am reading?

And you up there—are you the person writing me, or the person reading me?

You and I, alas, can have only one-way communication, for you are a person and I, a mere sentence.

As long as you are not reading me, the fourth word of this sentence has no referent.

The reader of this sentence exists only while reading me.

Now *that* is a rather frightening thought! And yet, by its own peculiar logic, it is certainly true.

Hey, out there—is that *you* reading me, or is it someone else?

Say, haven't you written me somewhere else before?

Say, haven't I written you somewhere else before?

The first of the three sentences above addresses its reader; the second addresses its author. In the last one, an author addresses a sentence.

Many sentences include words whose referents are hard to figure out because of their ambiguity—possibly accidental, possibly deliberate:

Thit sentence is not self-referential because "thit" is not a word.

No language can express every thought unambiguously, least of all this one.

In the Escher-inspired Figure 1-1, visual and verbal ambiguity are simultaneously exploited.

* * *

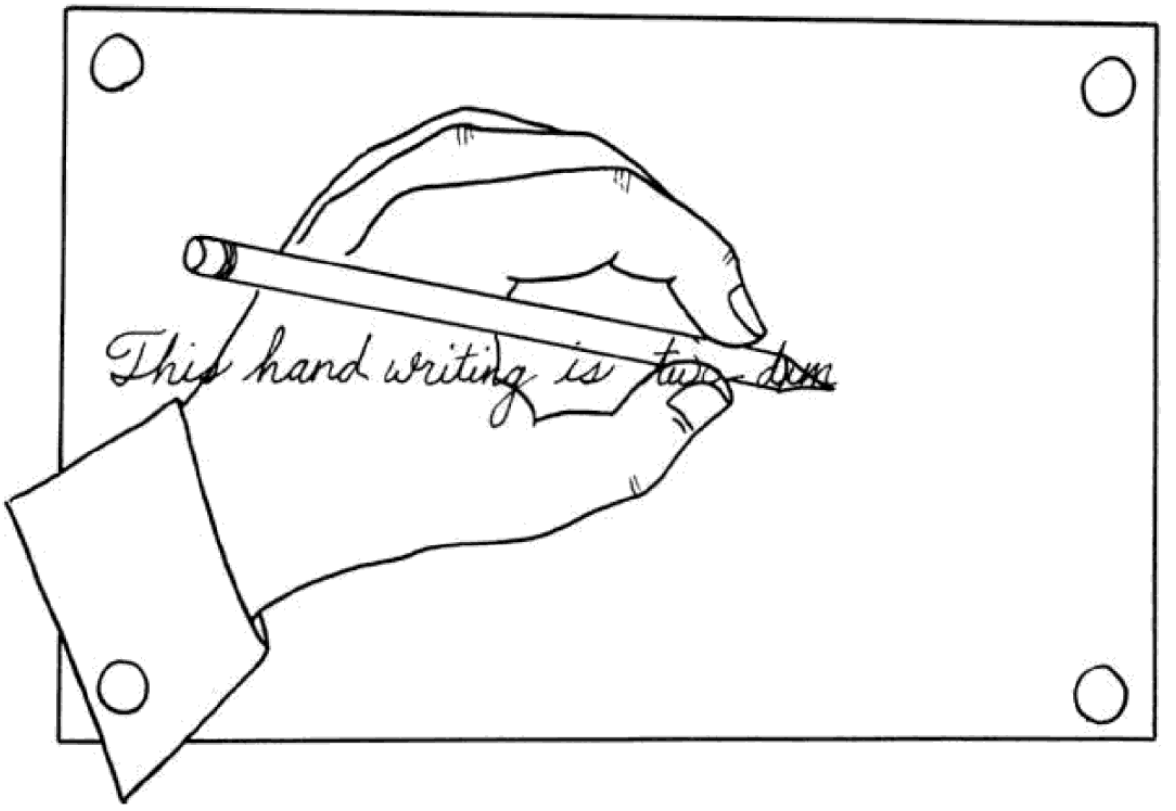


FIGURE 1–1. *Ambiguity: What is being described—the hand, or the writing?* [*Drawing by David Moser, after M. C. Escher.*]

Let us turn to a most interesting category, namely sentences that deal with the languages they are in, once were in, or might have been in:

When you are not looking at it, this sentence is in Spanish.

I had to translate this sentence into English because I could not read the original Sanskrit.

The sentence now before your eyes spent a month in Hungarian last year and was only recently translated back into English.

If this sentence were in Chinese, it would say something else.

.siht ekil ti gnidaer eb d'uoy, werbeH ni erew ecnetnes siht fI

The last two sentences are examples of *counterfactual conditionals.* Such a sentence postulates in its first clause (the *antecedent*) some contrary-to-fact situation (sometimes called a "possible world") and extrapolates in its second clause (the *consequent*) some consequence of it. This type of sentence opens up a rich domain for self-reference. Some of the more intriguing self-referential counterfactual conditionals I have seen are the following:

If this sentence didn't exist, somebody would have invented it.

If I had finished this sentence,

If there were no counterfactuals, this sentence would not be paradoxical.

If wishes were horses, the antecedent of this conditional would be true.

If this sentence were false, beggars would ride.

What would this sentence be like if it were not self-referential?

What would this sentence be like if π were 3?

Let us ponder the last of these (invented by Scott Kim) for a moment. In a world where π actually *did* have the value 3, you wouldn't ask about how things *would* be if π were 3. Instead, you might muse "if π were 2" or "if π *weren't* 3". So one's first answer to the question might be this: "What would this sentence be like if π weren't 3?". But there is a problem. The referent of "this sentence" has now changed identity. So is it fair to say that the second sentence is an answer to the first? It is a little like a woman who muses, "What would I be doing now if I had had different genes?" The problem is that she would not be herself; she would be someone else, perhaps the little boy across the street, playing in his sandbox. Personal pronouns like "I" cannot quite keep up with such strange hypothetical world-shifts.

But getting back to Scott Kim's counterfactual, I should point out that there is an even more serious problem with it than so far mentioned. Changing the value of π is, to put it mildly, a radical change in mathematics, and presumably you cannot change mathematics radically without having radically changed the fabric of the universe within which we live. So it is quite doubtful that any of the concepts in the sentence would make any sense if π were 3 (including the concepts of "π", "3", and so on).

Here are two more counterfactual conditionals to put in your pipe and smoke:

If the subjunctive was no longer used in English, this sentence would be grammatical.

This sentence would be seven words long if it were six words shorter.

These two lovely examples, invented by Ann Trail (who is also responsible for quite a few others in this column), bring us around to sentences that comment on their own form. Such sentences are quite distinct from ones that comment on their own content (such as the liar paradox, or the sentence that says "This sentence is not about itself, but about whether it is about itself."). It is easy to make up a sentence that refers to its own form, but it is hard to make up an *interesting* one. Here are a few more quite good ones:

because I didn't think of a good beginning for it.

This sentence was in the past tense.

This sentence has contains two verbs.

This sentence contains one numeral 2 many.

a preposition. This sentence ends in

In the time it takes you to read this sentence, eighty-six letters could have been processed by your brain.

* * *

David Moser, a composer and writer, is a delector and creator of self-reference and frame-breaking of all kinds. He has even written a story in which every sentence is self-referential (it is included in Chapter 2). It might seem unlikely that in such a limited domain, individual styles could arise and flourish, but David has developed a self-referential style quite his own. As a mutual friend (or was it David himself?) wittily observed, "If David Moser had thought up this sentence, it would

have been funnier." Many Moser creations have been used above. Some further Moserian delights are these:

This is not a complete. Sentence. This either.

This sentence contains only one nonstandard English flutzpah.

This gubblick contains many nonsklarkish English flutzpahs, but the overall pluggandisp can be glorked from context.

This sentence has cabbage six words.

In my opinion, it took quite a bit of flutzpah to just throw in a random word so that there *would* be cabbage six words in the sentence. That idea inspired the following: "This sentence has five (5) words." A few more miscellaneous Moserian gems follow:

This is to be or actually not two sentences to be, that is the question, combined.

It feels *sooo* good to have your eyes run over my curves and serifs.

This sentence is a !!!! premature punctuator

Sentences that talk about their own punctuation, as the preceding one does, can be quite amusing. Here are two more:

This sentence, though not interrogative, nevertheless ends in a question mark?

This sentence has no punctuation semicolon the others do period

Another ingenious inventor of self-referential sentences is Donald Byrd, several of whose sentences have already been used above. Don too has his own very characteristic way of playing with self-reference. Two of his sentences follow:

This hear sentence do'nt know Inglish purty good.

If you meet this sentence on the board, erase it.

The latter, via its form, alludes to the Buddhist saying "If you meet the Buddha on the road, kill him."

Allusion through similarity of form is, I have discovered, a marvelously rich vein of self-reference, but unfortunately this article is too short to contain a full proof of that discovery. I shall explicitly discuss only two examples. The first is "This sentence verbs good, like a sentence should." Its primary allusion is to the famous slogan "Winston tastes good, like a cigarette should", and its secondary allusion is to "This sentence no verb." The other example involves the following lovely self-referential remark, once made by the composer John Cage: "I have nothing to say, and I am saying it." This allows the following rather subtle twist to be made: "I have nothing to allude to, and I am alluding to it."

* * *

Some of the best self-referential sentences are short but sweet, relying for their effect on secondary interpretations of idiomatic expressions or well-known catch phrases. Here are five of my favorites, which seem to defy other types of categorization:

Do you read me?

This point is well taken.

You may quote me.

I am going two-level with you.

I have been sentenced to death.

In some of these, even sophisticated non-native speakers would very likely miss what's going on.

Surely no article on self-reference would be complete without including a few good examples of self-fulfilling prophecy. Here are a few:

This prophecy will come true.

This sentence will end before you can say "Jack Rob

Surely no article on self-reference would be complete without including a few good examples of self-fulfilling prophecy.

Does this sentence remind you of Agatha Christie?

That last sentence—one of Ann Trail's—is intriguing. Clearly it has nothing to do with Agatha Christie, nor is it in her style, and so the answer ought to be no. Yet I'll be darned if I can read it without being reminded of Agatha Christie! (And what is even stranger is that I don't know the first thing about Agatha Christie!)

In closing, I cannot resist the touching plea of the following Byrdian sentence:

Please, oh please, publish me in your collection of self-referential sentences!

Post Scriptum.

This first column of mine triggered a big wave of correspondence, some of which is presented in the next chapter. Most of the correspondence was light-hearted, but there were a number of serious letters that intrigued me. Here is a repartee that appeared in the pages of *Scientific American* a few months later.

> The kind of structural analysis engaged in, and the resulting questions raised by, Douglas Hofstadter in his amusing and intriguing article concerning self-referential sentences need not lead inevitably to bafflement of the reader.

Help is at hand from the "laggard science" psychology, but only from that carefully defined quarter of psychology known as behavior analysis, which was progenerated by the famous Harvard psychologist B. F. Skinner almost 50 years ago.

In examining the implications of linguistic analyses such as Hofstadter's for the serious student of verbal behavior, Skinner comments in his book *About Behaviorism* (pages 98 – 99) as follows:

> Perhaps there is no harm in playing with sentences in this way or in analyzing the kinds of transformations which do or do not make sentences acceptable to the ordinary reader, but it is still a waste of time, particularly when the sentences thus generated could not have been emitted as verbal behavior. A classical example is a paradox, such as 'This sentence is false', which appears to be true if false and false if true. The important thing to consider is that no one could ever have emitted the sentence as verbal behavior. A sentence must be in existence before a speaker can say, 'This sentence is false', and the response itself will not serve, since it did not exist until it was emitted. What the logician or linguist calls a sentence is not necessarily verbal behavior in any sense which calls for a behavioral analysis.

As Skinner pointed out long ago, verbal behavior results from contingencies of reinforcement

arranged by verbal communities, and it is these contingencies that must be analyzed if we are to identify the variables that control verbal behavior. Until we grasp the full import of Skinner's position, which goes beyond structure to answer *why* we behave as we do verbally or nonverbally, we shall continue to fall back on prescientific formulations that are about as useful in understanding these phenomena as Hofstadter's quaint metaphorical speculation: "Such a sentence would seem to be arrogantly proclaiming itself to be an animate agent."

George Brabner
College of Education
University of Delaware

I felt compelled to reply to Professor Brabner's interesting views about these matters, and so here is what I wrote:

> I assume that the quote from B. F. Skinner reflects Professor Brabner's own sentiments about the likelihood of self-referential utterances. I am always baffled by people who doubt the likelihood of self-reference and paradox. Verbal behavior comes in many flavors. Humor, particularly self-referential humor, is one of the most pervasive flavors of verbal behavior in this century. One has only to watch the Muppets or Monty Python on television to see dense and intricate webs of self-reference. Even advertisements excel in self-reference.
>
> In art, René Magritte, Pablo Picasso, M. C. Escher, John Cage, and dozens of others have played with the level-distinction between *that which represents* and *that which is represented*. The "artistic behavior" that results includes much self-reference and many confusing and sometimes exhilaratingly paradoxical tangles. Would Professor Brabner say that no one could ever have "emitted" such works as "artistic behavior"? Where is the borderline?
>
> Ordinary language, as I pointed out in my column, is filled with self-reference, usually a little milder-seeming than the very sharply pointed paradoxes that Professor Brabner objects to. "Mouth", "word", and so on are all self-referential. Language is inherently filled with the potential of sharp turns on which it may snag itself.
>
> Many scholarly papers begin with a sentence about "the purpose of this paper". Newspapers report on their own activities, conceivably on their own inaccuracies. People say, "I'm tired of this conversation." Arguments evolve about arguments, and can get confusingly and painfully self-involved. Has Professor Brabner never thought of "verbal behavior" in this light? It is likely that in hunting woolly mammoths, no one found it extraordinarily likely to shout, "This sentence is false!" However, civilization has come a long way since those days, and the primitive purposes of language have by now been almost buried under an avalanche of more complex purposes.
>
> Part of human nature is to be introspective, to probe. Part of our "verbal behavior" deliberately, often playfully, explores the boundaries between conceptual levels of systems. All of this has its root in the struggle to survive, in the fact that our brains have become so flexible that much of their time is spent in dealing with their own activities, consciously or unconsciously. It is simply a consequence of representational power—as Kurt Gödel showed—that systems of increasing complexity become increasingly self-referential.
>
> It is quite possible for people filled with self-doubt to recognize this trait in themselves, and to begin to doubt their self-doubt itself. Such psychological dilemmas are at the heart of some current theories of therapy. Gregory Bateson's "double bind", Victor Frankl's "logotherapy", and Paul Watzlawick's therapeutic ideas are all based on level-crossing paradoxes that crop up in real life. Indeed, psychotherapy is itself based completely on the idea of a "twisted system of self"—a self that wants to reach inward and change some presumably wrong part of itself.
>
> We human beings are the only species to have evolved humor, art, language, tangled psychological problems, even an awareness of our own mortality. Self-reference—even of the sharp Epimenides type—is connected to profound aspects of life. Would Professor Brabner argue that suicide is not conceivable human behavior?
>
> Finally, just suppose Professors Skinner and Brabner are right, and no one ever says exactly "This sentence is false." Would this mean that study of such sentences is a waste of time? Still not. Physicists study ideal gases because they represent a distillation of the most significant principles of the behavior of real gases. Similarly, the Epimenides paradox is an "ideal paradox"—one that cuts crisply to the heart of the matter. It has opened up vast domains in logic, pure science, philosophy, and other disciplines, and will continue to do so despite the skepticism of behaviorists.

It is a curious coincidence that the only other reply to my article that was printed in the "Letters" column of *Scientific American* also came from the University of Delaware. Here it is:

I hope that you do not receive any correspondence concerning Douglas R. Hofstadter's article on self-reference. I should like to inform your readers that many years of study on this problem have convinced me no conclusion whatsoever can be drawn from it that would stand up to a moment's scrutiny. There is no excuse for *Scientific American* to publish letters from those cranks who consider such matters to be worthy of even the slightest notice.

A. J. Dale
Department of Philosophy
University of Delaware

I replied as follows:

> Many years of reading such letters have convinced me that no reply whatsoever can be given to them that would stand up to a moment's scrutiny. There is no excuse for publishing responses to those cranks who send them.

After these two exchanges had appeared in print, a number of people remarked to me that they'd read the two letters from Delaware that had attacked me, and had enjoyed my responses. Two? I guess it wasn't so obvious that Dale's letter was completely tongue-in-cheek. In fact, that was its point.

* * *

Two other letters stand out sharply in my memory. One was from an individual who signed himself (I presume it is a male) as "Mr Flash qFiasco". Mr Flash insisted that a sentence cannot *say* what it *shows.* The former concerns only its *content,* which is supposedly independent of how it manifests itself in print, while the latter is a property exclusively of its *form,* that is, of the physical sentence only when it is in print. This distinction sounds crystal-clear at first, but in reality it is mud-blurry. Here is some of what Flash wrote me:

> For a sentence to attempt to say what it shows is to commit an error of logical types. It seems to be putting a round peg into a square hole, whereas it is instead putting a round peg into something which is not a hole at all, square or otherwise. This is a category mismatch, not a paradox. It is like throwing the recipe in with the flour and butter and eggs. The source of the equivocation is an illegitimate use of the term 'this'. 'This' can point to virtually anything, but 'this' cannot point to itself. If you stick out your index finger, you can point to virtually anything; and by curling it you can even point to the pointing finger; but you cannot point to *pointing.* Pointing is of a higher logical type than the thing which is doing the pointing. Similarly, the referent of 'this sentence' can be virtually anything but that sentence. Sentences of the form exemplified by 'This sentence no verb.' and 'This sentence has a verb.' are not well-formed: they commit fallacies of logical type equivocation. Thus their self-referential character is not genuine and they present no problem as paradoxes.

There will always be people around who will object in this manner, and in the Brabnerian manner. Such people think it is possible to draw a sharp line between attributes of a printed sentence that can be considered part of its *form* (*e.g.,* the typeface it is printed in, the number of words it contains, and so on), and attributes that can be considered part of its *content* (*i.e.,* the things and events and relationships that it refers to).

Now, I am used to thinking about language in terms of how to get a machine to deal with it, since I look at the human brain as a very complex machine that can handle language (and many other things as well). Machines, in trying to make sense of sentences, have access to nothing more than the *form* of such sentences. The *content,* if it is to be accessible to a machine, has to be derived, extracted, constructed, or created somehow from the sentence's physical structure, together with other knowledge and programs already available to the machine.

When very simple processing is used to operate on a sentence, it is convenient to label the information thus obtained "syntactic". For instance, it is clearly a syntactic fact about "This sentence no verb." that it contains six vowels. The vowel-consonant distinction is obviously a typographical one, and typographical facts are considered superficial and syntactic. But there is a problem here. With different *depths of processing,* aspects of different degrees of "semanticity" may be detected.

Consider, for example, the sentence "Mary was sick yesterday." Let's call it *Sentence M.* Listed below are the results of seven different degrees of processing of Sentence M by a hypothetical machine, using increasingly sophisticated programs and increasingly large knowledge bases. You

should think of them as being English translations, for your convenience, of computational structures inside the machine that it can act on and use fluently.

1. Sentence M contains twenty characters.
2. Sentence M contains four English words.
3. Sentence M contains one proper noun, one verb, one adjective, and one adverb, in that order.
4. Sentence M contains one human's name, one linking verb, one adjective describing a potential health state of a living being, and one temporal adverb, in that order.
5. The subject of Sentence M is a pointer to an individual named 'Mary', the predicate is an ascription of ill health to the individual so indicated, on the day preceding the statement's utterance.
6. Sentence M asserts that the health of an individual named 'Mary' was not good the day before today.
7. Sentence M says that Mary was sick yesterday.

Just where is the boundary line that says, "You can't do *that* much processing!"? A machine that could go as far as version 7 would have actually *understood*—at least in some rudimentary sense—the content of Sentence M. Work by artificial-intelligence researchers in the field of natural language understanding has produced some very impressive results along these lines, considerably more sophisticated than what is shown here. Stories can be "read" and "understood", at least to the extent that certain kinds of questions can be answered by the machine when it is probed for its understanding. Such questions can involve information not explicitly in the story itself, and yet the machine can fill in the missing information and answer the question.

I am making this seeming digression on the processing of language by computers because intelligent people like Mr Flash qFiasco seem to have failed to recognize that the boundary line between form and content is as blurry as that between blue and green, or between human and ape. This comparison is not made lightly. Humans are supposedly able to get at the "content" of utterances, being genuine language-users, while apes are not. But ape-language research clearly shows that there is some kind of in-between world, where a certain degree of content can be retrieved by a being with reduced mental capacity. If mental capacity is equated with potential processing depth, then it is obvious why it makes no sense to draw an arbitrary boundary line between the form and the content of a sentence. *Form blurs into content as processing depth increases.* Or, as I have always liked to say, "Content is just fancy form." By this I mean, of course, that "content" is just a shorthand way of saying "form as perceived by a very fancy apparatus capable of making complex and subtle distinctions and abstractions and connections to prior concepts".

Flash qFiasco's down-home, commonsense distinction between form and content breaks down swiftly, when analyzed. His charming image of someone making a "category error" by throwing a recipe in with the flour and butter and eggs reveals that he has never had Recipe Cake. This is a delicious cake whose batter is made out of cake recipes (if you use pie recipes, it won't taste nearly as good). The best results are had if the recipes are printed in French, in Baskerville Roman. A preponderance of *accents aigus* lends a deliciously piquant aroma to the cake. My recommendation to Brabner and qFiasco is: "Let them eat recipes."

* * *

Finally, I come to John Case, a computer scientist who wrote from Yale, insisting that there is no conceptual problem whatsoever in translating the French sentence "*Cette phrase en français est difficile à traduire en anglais*" into English. Case's translation was the following English sentence:

> The French sentence "*Cette phrase en français est difficile à traduire en anglais*" is difficult to translate into English.

In other words, Case translates a *self*-referential French sentence into an *other*-referential English sentence. The English sentence talks about the French sentence—in fact it quotes it completely! Something radical is missing here. At one level, of course, Case is right: now the two sentences, one French and one English, both are talking about (or pointing to) the same thing (the French sentence). But the absolute crux of the French one is its tangledness; the English one completely lacks that quality. Clearly Case has had to make a sacrifice, a compromise.

The alternative, which I prefer, is to construct in English an *analogue* to the French sentence: a *self*-referential English sentence, one that has a tangledness isomorphic to that of the French sentence. That's where the *essence* of the sentence lies, after all! "But is that its *translation*?" you might ask. A good question.

Ionesco once remarked, "The French for London is Paris." (Use-mention fanatic that I am, I assume that he meant "The French for 'London' is 'Paris' ", although it is pungent either way.) What he meant was that in understanding situations, French people tend to translate them into their own frame of reference. This is of course true for all of us. If Mary tells Ann, "My brother died", and if Ann does not know Mary's brother, then how can she understand this statement? Surely projection is of the essence: Ann will imagine her *own* brother dying (if she has one—and if not, then her sister, a good friend, possibly even a pet!). This alternate frame of reference allows Ann to empathize with Mary. Now if Ann *did* know Mary's brother somewhat, then she might flicker between thinking of him as the person she vaguely remembers and thinking of her own brother (friend, pet, or whatever) dying. This dilemma (discussed further in the postscript to Chapter 24) arises for all beings with their own preferred vantage points: Do I map things into *what they would be for me,* or do I stand apart and survey them completely objectively and impassively?

Case is advocating the latter, which is all very well as an intellectual stance to adopt, but when it comes to real life, it just won't cut the mustard. To be concrete, one might ask: What was the actual solution used in the French edition of *Scientific American?* The answer, surprising no one, I hope, was this: "This English sentence is difficult to translate into French." I rest my case.

* * *

I wonder what literalists like John Case would suggest as the proper translation of the title of the book *All the President's Men* (a book about the downfall of President Nixon, a downfall that none of the people around him could prevent). Would they say that *Tous les hommes du Président* fills the bill admirably? Back-translated rather literally, it means "All the men of the President". It completely lacks the allusion—the reference by similarity of form—to the nursery rhyme "Humpty Dumpty". Is that dispensable? In my opinion, hardly. To me, the essence of the title resides in that allusion. To lose that allusion is to deflate the title totally.

Of course, what do I mean by "that allusion"? Do I wish the French title to contain, somehow, an allusion to an *English* nursery rhyme? That would be rather pointless. Well, then, do I want the French title to allude to the French version of "Humpty Dumpty"? It all depends how well known that is. But given that Humpty Dumpty is practically an unknown figure to French-speaking people, it seems that something else is wanted. Any old French nursery rhyme? Obviously not. The critical allusion is to the lines "All the King's horses/ And all the King's men/ Couldn't put Humpty together again." Are there—*anywhere* in French literature—lines with a similar import? If not, how about in French popular songs? In French proverbs? Fairy tales?

One might well ask why French-speaking people would ever care about reading a book about Watergate in the first place. And even if they *did* want to read it, shouldn't it be *completely* translated, so that it happens in a French-speaking city? Come to think of it, didn't Ioratno once remark that the French for Washington is Montréal?

Clearly, this is carrying things to an extreme. There must be some middle ground of reasonableness. These are matters of subtle judgment, and they are where being human and

flexible makes all the difference. Rigid rules about translation may lead you to a kind of mechanical consistency, but at the sacrifice of all depth and charm. The problem of self-referential sentences is just the tip of the iceberg, as far as translation is concerned. It is just that these issues show up very early when direct self-reference is concerned. When self-reference (or reference in general, for that matter) is indirect, mediated by form, then fluidity is required. The understanding of such sentences involves a mixture of deriving the content and yet retaining the form in mind, letting qualities of the form conjure up flavors and enhance the meaning with a halo of not-quite-conscious pseudo-meanings, connotations, flavors, that flicker in the mind, not quite in reach, not quite out of reach. Self-reference is a good starting point for investigation of this kind of issue, because it is so much on the surface there. You can't sweep the problems under the rug, even though some would like to do so.

* * *

This first column, together with this postscript, provides a good introduction to the book as a whole, because many central issues are touched on: codes, translation, analogies, artificial intelligence, language and machines, mind and meanings, self and identity, form and content—all the issues I originally was motivated by when first writing that collection of teasing self-referential sentences.

2

Self-Referential Sentences: A Follow-Up

January, 1982

As January has rolled around again, I thought I'd give a follow-up to my column of a year ago on self-referential sentences, and that is what this column is; however, before we get any further, I would like to take advantage of this opening paragraph to warn those readers whose sensibilities are offended by explicit self-referential material that they probably will want to quit reading before they reach the end of this paragraph, or for that matter, this sentence—in fact, this clause—even this noun phrase—in short, *this.*

Well, now that we've gotten *that* out of the way, I would like to say that, since last January, I have received piles upon piles of self-referential mail. Tony Durham astutely surmised: "What with the likely volume of replies, I should not think you are reading this in person." John C. Waugh's letter yelped: "Help, I'm buried under an avalanche of reader's responses!" At first, I thought Waugh himself was empathizing with my plight, putting words into my own mouth, but then I realized it was his *letter* calling for help. Fortunately, it was rescued, and now is comfortably nestled in a much reduced pile. Indeed, I have had to cull from that massive influx of hundreds of replies a very small number. Here I shall present some of my favorites.

Before leaving the topic of mail, I would like to point out that the postmark on Ivan Vince's postcard from Britain cryptically remarked, "Be properly addressed." Was this an order issued by the post office to the postcard itself? If so, then British postcards must be far more intelligent than American ones; I have yet to meet a postcard that could read, let alone correct its own address. (One postcard that reached me was addressed to me in care of *Omni* magazine! And yet somehow it arrived.)

I was flattered by a couple of self-undermining compliments. Richard Ruttan wrote, "I just can't tell you how much I enjoyed your first article.", and John Collins said, "This does not communicate my delight at January's column." I was also pleased to learn that my fame had spread as far as the men's room at the Tufts University Philosophy Department, where Dan Dennett discovered "This sentence is graffiti.—Douglas R. Hofstadter" penned on the wall.

* * *

A popular pastime was the search for interesting self-answering questions. However, only a few succeeded in genuinely "jootsing" (jumping out of the system), which, to me, means being truly novel. It seems that successes in this limited art form are not easy to come by. John Flagg cynically remarked (I paraphrase slightly): "Ask a self-answering question, and get a self-questioning

answer." One of my favorites was given by Henry Taves: "I fondly remember a history exam I encountered in boarding school that contained the following: 'IV. Write a question suitable for a final exam in this course, and then answer it.' My response was simply to copy that sentence twice." I was delighted by this. Later, upon reflection, I began to suspect something was slightly wrong here. What do you think?

Richard Showstack contributed two droll self-answering questions: "What question no verb?" and "What is a question that mentions the word 'umbrella' for no apparent reason?" Jim Shiley sent in a clever entry that I modify slightly into "Is this a rhetorical question, or is this a rhetorical question?" He also contributed the following idea:

> Take a blank sheet of paper and on it write:
>
> *How far across the page will this sentence run?*
>
> Now if some polyglot friend of yours points out that the same string of phonemes in Ural-Altaic means '2.3 inches', send me a free subscription to *Scientific American.* Otherwise, if the inscription of a question counts both as the question and as unit of measure, I at least get a booby prize. But I think somehow I bent the rules.

My own solutions to the problem of the self-answering question are actually not so much self-*answering* as self-*provoking,* as in the following example: "Why are you asking me *that* out of the blue?" It is obvious that when the question is asked out of the blue, it might well elicit an identical response, indicating the hearer's bewilderment.

Philip Cohen relayed the following anecdote about a self-answering question, from Damon Knight: "Terry Carr, an old friend, sent us a riddle on a postcard, then the answer on another postcard. Then he sent us another riddle: 'How do you keep a turkey in suspense?' and never sent the answer. After about two weeks, we realized that *was* the answer."

* * *

Several of the real masterpieces sent in belong to what I call the *self-documenting* category, of which a simple example is Jonathan Post's "This sentence contains ten words, eighteen syllables and sixty-four letters." A neat twist is supplied by John Atkins in his sentence " 'Has eighteen letters' does." The self-documenting form can get much more convoluted and introspective. An example by the wordplay master Howard Bergerson was brought to my attention by Philip Cohen. It goes:

> In this sentence, the word *and* occurs twice, the word *eight* occurs twice, the word *four* occurs twice, the word *fourteen* occurs four times, the word *in* occurs twice, the word *seven* occurs twice, the word *the* occurs fourteen times, the word *this* occurs twice, the word *times* occurs seven times, the word *twice* occurs eight times and the word *word* occurs fourteen times.

That is good, but the gold medal in the category is reserved for Lee Sallows, who submitted the following *tour de force:*

> Only the fool would take trouble to verify that his sentence was composed of ten a's, three b's, four c's, four d's, forty-six e's, sixteen f's, four g's, thirteen h's, fifteen i's, two k's, nine l's, four m's, twenty-five n's, twenty-four o's, five p's, sixteen r's, forty-one s's, thirty-seven t's, ten u's, eight v's, eight w's, four x's, eleven y's, twenty-seven commas, twenty-three apostrophes, seven hyphens, and, last but not least, a single!

I (perhaps the fool) did take trouble to verify the whole thing. First, though, I carried out some spot checks. And I must say that when the first random spot check worked (I think I checked the number of 'g's), this had a strong psychological effect: all of a sudden, the credibility rating of the *whole sentence* shot way up for me. It strikes me as weird (and wonderful) how, in certain situations, the verification of a tiny percentage of a theory can serve to powerfully strengthen your belief in the full theory. And perhaps that's the whole point of the sentence!

The noted logician Raphael Robinson submitted a playful puzzle in the self-documenting genre.

Readers are asked to complete the following sentence:

> In this sentence, the number of occurrences of 0 is —, of 1 is —, of 2 is —, of 3 is —, of 4 is —, of 5 is —, of 6 is —, of 7 is —, of 8 is —, and of 9 is —.

Each blank is to be filled with a numeral of one or more digits, written in decimal notation. Robinson states that there are exactly two solutions. Readers might also search for two sentences of this form that document each other, or even longer loops of that kind.

Clearly the ultimate in self-documentation would be a sentence that does more than merely inventory its parts; it would be a sentence that includes a rule as well, telling all the King's men how to put those parts back together again to create a full sentence—in short, a self-reproducing sentence. Such a sentence is Willard Van Orman Quine's English rendition of Kurt Gödel's classic metamathematical homage to Epimenides the Cretan:

> "yields falsehood when appended to its quotation." yields falsehood when appended to its quotation.

Quine's sentence in effect tells the reader how to construct a replica of the sentence being read, and then (just for good measure) adds that the replica (not *itself,* for heaven's sake!) asserts a falsity! It's a bit reminiscent of the famous remark made by Epilopsides the Concretan (second cousin of Epimenides) to Flora, a beautiful young woman whose ardent love he could not return (he was betrothed to her twin sister Fauna): "Take heart, my dear. I have a suggestion that may cheer you up. Just take one of these cells from my muscular biceps here, and clone it. You'll soon wind up with a dashing blade who looks and thinks just like me! But *do* watch out for him—he is given to telling beautiful women real whoppers!"

* * *

In the early 1950's, John von Neumann worked hard trying to design a machine that could build a replica of itself out of raw materials. He came up with a theoretical design consisting of hundreds of thousands of parts. Seen in hindsight and with a considerable degree of abstraction, the idea behind von Neumann's self-reproducing machine turns out to be pretty similar to the means by which DNA replicates itself. And this in turn is close to Gödel's method of constructing a self-referential sentence in a mathematical language in which at first there seems to be no way of referring to the language itself.

The First Every-Other-Decade Von Neumann Challenge is thus hereby presented for ambitious readers: Create a comprehensible and not unreasonably long self-documenting sentence that not only lists its parts (at the word level or, better yet, the letter level) but also tells how to put them together so that the sentence reconstitutes itself. (Notice, by the way, the requirement is that the sentence be *not unreasonably long,* which is different—very different—from being *reasonably long.*) The parts list (or *seed*) should be an inventory of words or typographical symbols, more or less as in the sentences created by Howard Bergerson and Lee Sallows. The inventoried symbols should in some way be clearly distinguishable from the text that talks about them. For instance, they can be enclosed in quotation marks, printed in another typeface, or referred to by name. It is not so important what convention is adopted, so long as the distinction is sharp. The rest of the sentence (the *building rule*) should be printed normally, since it is to be regarded not as typographical raw material but as a set of instructions. This is the use-mention distinction I discussed in Chapter 1, and to disregard it is a serious conceptual weakness. (It is a flaw in Sallows' sentence that slightly tarnishes the gold on his medal.)

The building rule may not talk about normally-printed material—only about parts of the inventory. Thus, it is not permitted for the building rule to refer to itself in any way! The building rule has to describe structure explicitly. Furthermore (and this is the subtlest and probably the most often overlooked aspect of self-reference), the building rule must specify which parts are to be printed normally and which parts in quotes (or however the raw materials are being indicated).

In this respect, Bergerson's sentence fails. Although, to its credit, it sharply distinguishes between use and mention by relying on upper case for the names of inventory items and lower case for item counts and filler words, it does not have separate inventories for items in upper case and lower case. Instead it lumps the two together, blurring a vital distinction.

In the Von Neumann Challenge, extra points will be awarded for solutions given in Basic English, or whose seed is entirely at the letter level (as in Sallows' sentence). The Quine sentence, although it clearly incorporates a seed (the seven-word phrase in quotation marks) and a building rule (that of appending something to its quotation), is not a legal entry because its seed is too far from being raw material. It is so structured that it is like a fetus more than it is like a zygote.

* * *

There is a very good reason, by the way, that the Quine sentence's seed is so complicated—in fact, is identical with the building rule, except for the quotation marks. The reason is simple to state: You've got to *build a copy of the building rule* out of raw materials, and the more your building rule looks like your seed, the simpler it will be to build a copy of it from a copy of the seed. To make a full new sentence, all you need to do is make two copies of the seed, carry out whatever simple manipulations will convert one copy of the seed into the building rule, and then splice the other copy of the seed onto the newly minted building rule to make up a complete new sentence, fresh off the assembly line.

To make this clearer, it is helpful to show a slight variation on Quine's sentence. Imagine that you could recognize only the lowercase roman letters, and that uppercase letters were alien to you. Then text printed in upper case would, for all practical purposes, be devoid of meaning or interest, whereas text in lower case would be full of meaning and interest, able to suggest ideas or actions in your mind. Now suppose someone gave you a conversion table that matched each uppercase letter with its lowercase counterpart, so that you could "decode" uppercase text. Then one day you came across this piece of "meaningless" uppercase text:

YIELDS A FALSEHOOD WHEN USED AS THE SUBJECT OF ITS LOWERCASE VERSION

On being decoded, it would yield a lowercase sentence, or rather, a lowercase sentence fragment—a predicate without a subject. Suggestive, eh? What might you try out, as a possible subject of that predicate?

This notion of two parallel alphabets, one in which text is inert and meaningless and the other in which text is active and meaningful, may strike you as yielding no more than a minor variation on Quine's sentence, but in fact it is very similar to an exceedingly clever trick that nature discovered and has exploited in every cell of every living organism. Our seed—our genome—our DNA—is a huge long volume of *inert* text written in a chemical alphabet that has 64 "uppercase" letters (codons). Our building rules—our enzymes—are short, pithy slogans of *active* text written in a different chemical alphabet that has just twenty "lowercase" letters (amino acids). There is a map (the genetic code) that converts uppercase letters into lowercase ones. Obviously, some lowercase letters must correspond to more than one uppercase letter, but here that is a detail. It also turns out that three characters of the uppercase alphabet are not letters but punctuation marks telling where one pithy slogan ends and the next one begins—but again, these are details. (See Chapter 27 for some of those details.)

Once you know this mapping, you often won't even remember to distinguish between the two chemical alphabets: the inert uppercase codon alphabet and the active lowercase amino acid alphabet. The main thing is that, armed with the genetic code, you can read the DNA book (seed) as if it were a sequence of enzyme slogans (building rules) telling how to write a new DNA book together with a new set of enzyme slogans! It is a perfect parallel to our variation on the Quine sentence, where inert, uppercase seed-text was converted into active, lowercase rule-text that told

how to make a copy of the full Quine sentence, given its seed.

A cell's DNA and enzymes act like the seed and building rules of Quine's sentence, or the parts list and building rules of von Neumann's self-reproducing automaton—or then again, like the seed and building rules of computer programs that print themselves out. It is amazing how universal this mechanism of self-reference is, and for that reason I always find it quaint that people who rant and rave against the silliness of self-reference are themselves composed of trillions and trillions of tiny self-referential molecules.

* * *

Scott Kim and I constructed an intriguing pair of sentences:

> The following sentence is totally identical with this one, except that the words 'following' and 'preceding' have been exchanged, as have the words 'except' and 'in', and the phrases 'identical with' and 'different from'.
>
> The preceding sentence is totally different from this one, in that the words 'preceding' and 'following' have been exchanged, as have the words 'in' and 'except', and the phrases 'different from' and 'identical with'.

At first glance, these sentences are reminiscent of a two-step variant on the Epimenides paradox ("The following sentence is true."; "The preceding sentence is false."). On second glance, though, they are seen to say exactly the same thing. Curiously, my Australian colleague and sometime alter ego, Egbert B. Gebstadter, writing in his ever fascinating but often-furiating monthly row "Thetamagical Memas" (which appears in *Literary Australian*), disagrees with me; he maintains they say totally different things. (See figure 2-1.)

Not surprisingly, several of the sentences submitted by readers had a paradoxical flavor. Some were variants on Bertrand Russell's paradox about the barber who shaves all those who do not shave themselves, or the set of all sets that do not include themselves as elements. For instance, Gerald Hull concocted this strange sentence: "This sentence refers to every sentence that does not refer to itself." Is Hull's concoction self-referential, or is it not? In a similar vein, Michael Gardner cited a Reed College senior thesis whose dedication ran: "This thesis is dedicated to all those who did not dedicate their theses to themselves." The book *Model Theory,* by C. C. Chang and H. J. Keisler, bears a similar dedication, as Charles Brenner pointed out to me. He also suggested another variant on Russell's paradox: Write a computer program that prints out a list of all programs that do not ever print themselves out. The question is, of course: Will this program ever print itself out?

One of the most disorienting sentences came from Robert Boeninger: "This sentence does in fact not have the property it claims not to have." Got that? A serious problem seems to be to figure out just what property it is that the sentence claims it lacks.

The Dutch mathematician Hans Freudenthal sent along a charming paradoxical anecdote based on self-reference:

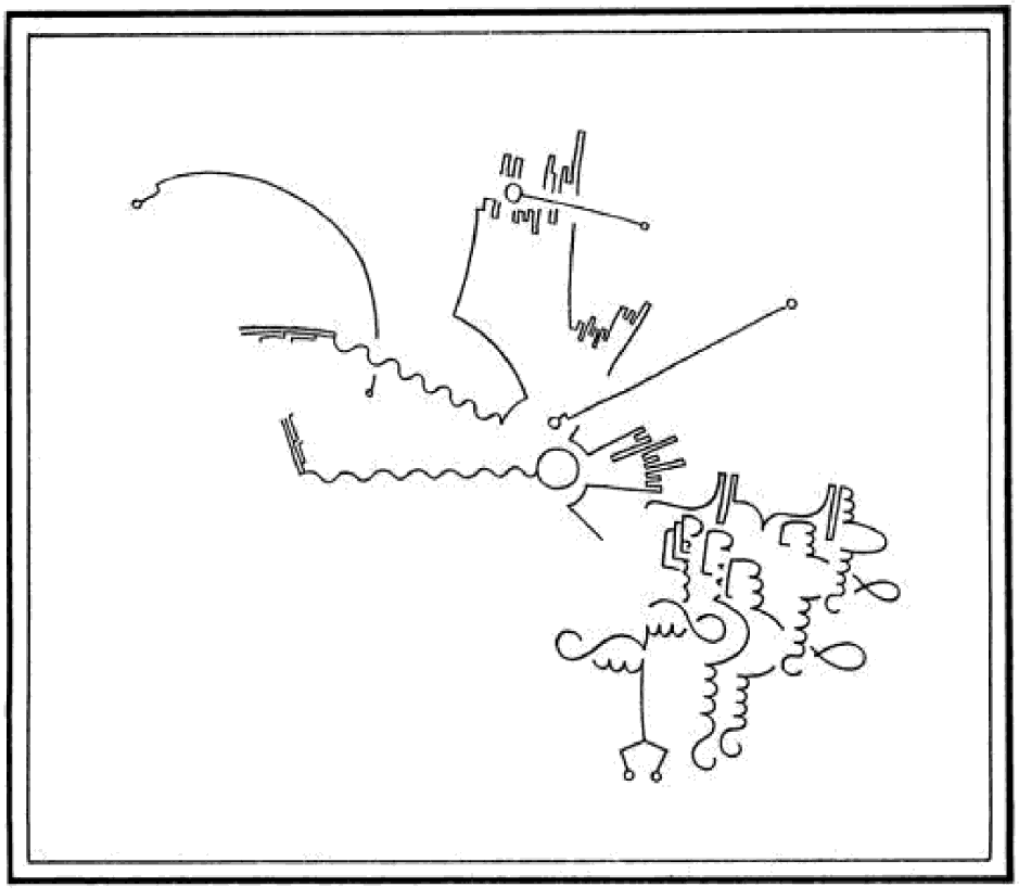
> There is a story by the eighteenth-century German Christian Gellert called "Der Bauer und sein Sohn" ("The Peasant and His Son"). One day during a walk, when the son tells a big lie, his father direly warns him about the "Liars' Bridge", which they are approaching. This bridge always collapses when a liar walks across it. After hearing this frightening warning, the boy admits his lie and confesses the truth.
>
> When I [Freudenthal] told a ten-year-old boy this story, he asked me what happened when they eventually came to the bridge. I replied, "It collapsed under the father, who had lied, since in fact there is no Liars' Bridge." (Or did it?)

C. W. Smith, writing from London, Ontario, described a situation reminiscent of the Epimenides paradox:

THETAMAGICAL MEMAS:

Seeking the Whence of Letter and Spirit

EGBERT B. GEBSTADTER



A Copious Concatenation of Artsy, Scientistic, and Literal Mumbo-Jumbo

FIGURE 2–1. *The cover of Egbert B. Gebstadter's latest book, showing some of his "Whorly Art." See the Bibliography for a short description of the book.*

Gebstadter, best known as the author of Copper, Silver, Gold: an Indestructible Metallic Alloy, *also co-edited* The Brain's U *with Australian philosopher Denial E. Dunnitt, and for two and a half years wrote a monthly row ("Thetamagical Memas") for* Literary Australian. *Having spent the last several years in the Psychology Department of Pakistania University in Wiltington, Pakistania, he has recently joined the faculty of the Computer Science Department of the University of Mishuggan in Tom Treeline, Mishuggan, where he occupies the Rexall Chair in the College of Art, Sciences, and Letters. His current research projects in IA (intelligent artifice) are called Quest-Essence, Mind Pattern, Intellect, and Studio. His focus is on deterministic sequential models of digital emotion.*

During the 1960's, standing alone in the midst of a weed-strewn field in this city, there was a weathered sign that read: "$25 reward for information leading to the arrest and conviction of anyone removing this sign." For whatever it's worth, the sign has long since disappeared. And so, for that matter, has the field.

Incidentally, the Epimenides paradox should not be confused with the Nixonides paradox, first uttered by Nixonides the Cretin in A.D. 1974: "This statement is inoperative." Speaking of Epimenides, one of the most elegant variations on his paradox is the "Errata" section in a hypothetical book described by Beverly Rowe. It looks like this:

(vi)

Errata

Page (vi): For *Errata,* read *Erratum*

Closely related to the truly paradoxical sentences are those that belong to what I call the *neurotic* and *healthy* categories. A healthy sentence is one that, so to speak, practices what it preaches, whereas a neurotic sentence is one that says one thing while doing its opposite. Alan Auerbach has given us a good example in each category. His healthy sentence is: "Terse!" His neurotic sentence is: "Proper writing—and you've heard this a million times—avoids exaggeration." Here's a healthy one by Brad Shelton: "Fourscore and seven words ago, this sentence hadn't started yet." One of the jootsingest of sentences came from Carl Bender:

The rest of this sentence is written in Thailand, on

Consider a related sentence sent in by David Stork: "It goes without saying that. . ." To which category does it belong? Perhaps it is a psychotic sentence.

Pete Maclean contributed a puzzling one: "If the meanings of 'true' and 'false' were switched, then this sentence wouldn't be false." I'm still scratching my head over what that means! Dan Krimm wrote to tell me: "I've heard that this sentence is a rumor." Linda Simonetti contributed the following example, "which actually is not a complete sentence, but merely a subordinate clause." Douglas Wolfe offered the following neurotic rule of thumb: "Never use the imperative, and it is also never proper to construct a sentence using mixed moods." David Moser reminded me of a slogan that the *National Lampoon* once used: "So funny it sells without a slogan!" Perry Weddle

wrote, "I'm trying to teach my parrot to say, 'I don't understand a thing I say.' When *I* say it, it's viciously self-referential, but in *his* case?" Stephen Coombs pointed out that "A sentence may self-refer in the verb." My mother, Nancy Hofstadter, heard Secretary of State Alexander Haig describe a warning message to the Russians as "a calculated ambiguity that would be clearly understood". Yes, Sir!

Jim Propp submitted a sequence of sentences that slide elegantly from the neurotically healthy to the healthily neurotic:

(1) This sentence every third, but it still comprehensible.
(2) This would easier understand fewer had omitted.
(3) This impossible except context.
(4) *4'33"* attempt idea.
(5)

The penultimate sentence refers to John Cage's famous piece of piano music consisting of four minutes and 33 seconds of silence. The last sentence might well be an excerpt from *The Wit and Wisdom of Spiro T. Agnew,* although it is too short an excerpt to be sure. Propp also sent along the following healthy sentence, which was apparently inspired by his readings in the book *Intelligence in Ape and Man,* by David Premack: "By the 'productivity' of language, I mean the ability of language to introduce new words in terms of old ones."

Philosopher Howard DeLong contributed what might be considered a neurotic syllogism:

> All invalid syllogisms break at least one rule.
> This syllogism breaks at least one rule.
>
> ---
>
> Therefore, this syllogism is invalid.

Several readers pointed out phrases and jokes that have been making the rounds. D.A. Treissman, for instance, reminded me that "Nostalgia ain't what it used to be." Henry Taves mentioned the delightful T-shirts adorned with statements such as "My folks went to Florida and all they brought back for me was this lousy T-shirt!" And John Fletcher described an episode of the television program *Laugh-In* a few years ago on which Joanne Worley sang, "I'm just a girl who can't say 'n . . .', 'n . . .', 'n . . .' ". John Healy wrote, "I used to think I was indecisive, but now I'm not so sure."

I myself have a few contributions to this collection. A neurotic one is: "In this sentence, the concluding three words 'were left out'." Or is it neurotic? These things confuse me! In any case, a most healthy sentence is: "This sentence offers its reader(s) various alternatives/options that he or she (or they) is (are) free to accept and/or reject." And then there is the inevitable "This sentence is neurotic." The thing is, if it *is* neurotic, it practices what it preaches, so it's healthy and therefore cannot be neurotic—but then if it *isn't* neurotic, it's the opposite of what it claims to be, so it's *got* to be neurotic. No wonder it's neurotic, poor thing!

Speaking of neurotic sentences, what about sentences with identity crises? These are, in some sense, the most interesting ones of all to me. A typical example is Dan Krimm's vaguely apprehensive question, "If I stated something else, would it still be me?" I thought this could be worded better, so I revised it slightly, as follows: "If I said something else, would it still be me saying it?" I still was not happy, so I wrote one more version: "In another world, could I have been a sentence about Humphrey Bogart?" When I paused to reflect on what I had done, I realized that in reworking Dan's sentence, I had tampered with its identity in the very way it feared. The question remained, however: Were all these variants really the same sentence, deep down? My last experiment along these lines was: "In another world, could this sentence have been Dan Krimm's

sentence?"

Clearly some readers were thinking along parallel lines, since John Atkins queried, "Can anyone explain why this would still be the same magazine without this query, and yet this would not be the same query without this word?" (Of course, just which word "this word" refers to is a little vague, but the idea is clear.) And Loul McIntosh, who works at a rehabilitation center for formerly schizophrenic patients, had a question connecting personal identity with self-referential sentences: "If I were you, who would be reading this sentence?" She then added: "That's what I get for working with schizophrenics." This brings me to Peter M. Brigham, M.D., who in his work ran across a severe case of literary schizophrenia: "You have, of course, just begun reading the sentence that you have just finished reading." It's one of my favorites.

Pursuing the slithery snake of self in his own way, Uilliam M. Bricken, Jr., wrote in: "If you think this sentence is confusing, then change one pig." Now, *anyone* can see that this doesn't make any sense at all. Surely what he meant was, "If you think this sentence is confusing, then *roast* one pig."—don't ewe agree? By the by, if ewe think "Uilliam" is confusing, then roast one ewe. And while we're mentioning ewes, what's a nice word like "ewe" doing in a foxy paragraph like this?

A while back, driving home late at night, I tuned in to a radio talk show about pets. A heated discussion was taking place about the relative merits of various species, and at one point the announcer mused, "If a dog had written this broadcast, he might have said that *people* are inferior because they don't wag their tails." This gave me paws for thought: What might this column have been like if it had been written by a dog? I can't say for *sure,* but I have a hunch it would have been about chasing squirrels. And it might have had a paragraph speculating about what this column would have been like if it had been written by a squirrel.

* * *

I think my favorite of all the sent-in-ces was one contributed by Harold Cooper. He was inspired by Scott Kim's counterfactual self-referential question: "What would this sentence be like if π were 3?" His answer is shown in Figure 2-2. This, to me, exemplifies the meaning of the verb "joots". The six-sided 'o's represent the fact that the ratio of the circumference to the diameter of a hexagon is 3. Clearly, in Cooper's mind, if π were 3, why, what more natural conclusion than that *circles would be hexagons*! Who could ever think otherwise? I was intrigued by the fact that, as π's value slipped to 3, not only did circles turn into hexagons, but also the interrogative mood slipped into the declarative mood. Remember that the question asked how the question itself would be in that strange subjunctive world. Would it lose its curiosity about itself and cease to be a question? I did not see why that personality trait of the sentence would be affected by the value of π. On the other hand, it seemed obvious to me that if π were 3, the antecedent of the conditional should no longer be subjunctive. In fact, rather than saying "if π were 3", it should say, "*because* π is 3" (or something to that effect). Putting my thoughts together, then, I came up with a slight variation on Cooper's sentence: "What is this sentence like, π being 3 (as usual)?"

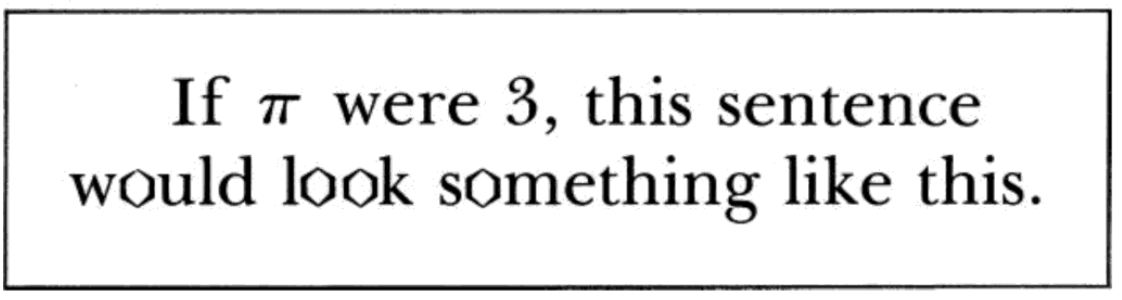


FIGURE 2–2. *A counterfactual self-referential sentence, inspired by Harold Cooper and Scott Kim.*

Several readers were interested in sentences that refer to the language they are in (or not in, as the case may be). An example is "If you spoke English, you'd be in your home language now." Jim Propp sent in a delightful pair of such sentences that need to be read together:

> Cette phrase se réfère à elle-même, mais d'une manière peu évidente à la plupart des Américains.
>
> Plim glorkle pegram ut replat, trull gen ris clanter froat veb nup lamerack gla smurp Earthlings.

If you do not understand the first sentence, just get a Martian friend to help you decode the second one. That will provide hints about the first. (I apologize for leaving off the proper Martian accent marks, but they were not available in this typeface.)

* * *

Last January, I published several sentences by David Moser and mentioned that he had written an entire story consisting of self-referential sentences. Many readers were intrigued. I decided there could be no better way to conclude this column than to print David's story in its entirety. So here 'tis!

This Is the Title of This Story, Which Is Also Found Several Times in the Story Itself

This is the first sentence of this story. This is the second sentence. This is the title of this story, which is also found several times in the story itself. This sentence is questioning the intrinsic value of the first two sentences. This sentence is to inform you, in case you haven't already realized it, that this is a self-referential story, that is, a story containing sentences that refer to their own structure and function. This is a sentence that provides an ending to the first paragraph.

This is the first sentence of a new paragraph in a self-referential story. This sentence is introducing you to the protagonist of the story, a young boy named Billy. This sentence is telling you that Billy is blond and blue-eyed and American and twelve years old and strangling his mother. This sentence comments on the awkward nature of the self-referential narrative form while recognizing the strange and playful detachment it affords the writer. As if illustrating the point made by the last sentence, this sentence reminds us, with no trace of facetiousness, that children are a precious gift from God and that the world is a better place when graced by the unique joys and delights they bring to it.

This sentence describes Billy's mother's bulging eyes and protruding tongue and makes reference to the unpleasant choking and gagging noises she's making. This sentence makes the observation that these are uncertain and difficult times, and that relationships, even seemingly deep-rooted and permanent ones, do have a tendency to break down.

Introduces, in this paragraph, the device of sentence fragments. A sentence fragment. Another. Good device. Will be used more later.

This is actually the last sentence of the story but has been placed here by mistake. This is the title of this story, which is also found several times in the story itself. As Gregor Samsa awoke one morning from uneasy dreams he found himself in his bed transformed into a gigantic insect. This sentence informs you that the preceding sentence is from another story entirely (a much better one, it must be noted) and has no place at all in this particular narrative. Despite the claims of the preceding sentence, this sentence feels compelled to inform you that the story you are reading is in actuality "The Metamorphosis" by Franz Kafka, and that the sentence referred to by the preceding sentence is the *only* sentence which does indeed belong in this story. This sentence overrides the

preceding sentence by informing the reader (poor, confused wretch) that this piece of literature is actually the Declaration of Independence, but that the author, in a show of extreme negligence (if not malicious sabotage), has so far failed to include even *one single sentence* from that stirring document, although he has condescended to use a small sentence *fragment,* namely, "When in the course of human events", embedded in quotation marks near the end of a sentence. Showing a keen awareness of the boredom and downright hostility of the average reader with regard to the pointless conceptual games indulged in by the preceding sentences, *this* sentence returns us at last to the scenario of the story by asking the question, "Why is Billy strangling his mother?" This sentence attempts to shed some light on the question posed by the preceding sentence but fails. *This* sentence, however, succeeds, in that it suggests a possible incestuous relationship between Billy and his mother and alludes to the concomitant Freudian complications any astute reader will immediately envision. Incest. The unspeakable taboo. The universal prohibition. Incest. And notice the sentence fragments? Good literary device. Will be used more later.

This is the first sentence in a new paragraph. This is the last sentence in a new paragraph.

This sentence can serve as either the beginning of the paragraph or the end, depending on its placement. This is the title of this story, which is also found several times in the story itself. This sentence raises a serious objection to the entire class of self-referential sentences that merely comment on their own function or placement within the story (*e.g.,* the preceding four sentences), on the grounds that they are monotonously predictable, unforgivably self-indulgent, and merely serve to distract the reader from the real subject of this story, which at this point seems to concern strangulation and incest and who knows what other delighful topics. The purpose of this sentence is to point out that the preceding sentence, while not itself a member of the class of self-referential sentences it objects to, nevertheless *also* serves merely to distract the reader from the real subject of this story, which actually concerns Gregor Samsa's inexplicable transformation into a gigantic insect (despite the vociferous counterclaims of other well-meaning although misinformed sentences). This sentence can serve as either the beginning of a paragraph or the end, depending on its placement.

This is the title of this story, which is also found several times in the story itself. This is *almost* the title of the story, which is found only once in the story itself. This sentence regretfully states that up to this point the self-referential mode of narrative has had a paralyzing effect on the actual progress of the story itself—that is, these sentences have been so concerned with analyzing themselves and their role in the story that they have failed by and large to perform their function as communicators of events and ideas that one hopes coalesce into a plot, character development, etc.—in short, the very *raisons d'être* of any respectable, hardworking sentence in the midst of a piece of compelling prose fiction. This sentence in addition points out the obvious analogy between the plight of these agonizingly self-aware sentences and similarly afflicted human beings, and it points out the analogous paralyzing effects wrought by excessive and tortured self-examination.

The purpose of this sentence (which can also serve as a paragraph) is to speculate that if the Declaration of Independence had been worded and structured as lackadaisically and incoherently as this story has been so far, there's no telling what kind of warped libertine society we'd be living in now or to what depths of decadence the inhabitants of this country might have sunk, even to the point of deranged and debased writers constructing irritatingly cumbersome and needlessly prolix sentences that sometimes possess the questionable if not downright undesirable quality of referring to themselves and they sometimes even become run-on sentences or exhibit other signs of inexcusably sloppy grammar like unneeded superfluous redundancies that almost certainly would have insidious effects on the lifestyle and morals of our impressionable youth, leading them to commit incest or even murder and maybe *that's* why Billy is strangling his mother, because of sentences *just like this one,* which have no discernible goals or perspicuous purpose and just end up anywhere, even in mid

Bizarre. A sentence fragment. Another fragment. Twelve years old. This is a sentence that.

Fragmented. And strangling his mother. Sorry, sorry. Bizarre. This. More fragments. This is it. Fragments. The title of this story, which. Blond. Sorry, sorry. Fragment after fragment. Harder. This is a sentence that. Fragments. Damn good device.

The purpose of this sentence is threefold: (1) to apologize for the unfortunate and inexplicable lapse exhibited by the preceding paragraph; (2) to assure you, the reader, that it will not happen again; and (3) to reiterate the point that these are uncertain and difficult times and that aspects of language, even seemingly stable and deeply rooted ones such as syntax and meaning, do break down. This sentence adds nothing substantial to the sentiments of the preceding sentence but merely provides a concluding sentence to this paragraph, which otherwise might not have one.

This sentence, in a sudden and courageous burst of altruism, tries to abandon the self-referential mode but fails. This sentence tries again, but the attempt is doomed from the start.

This sentence, in a last-ditch attempt to infuse some iota of story line into this paralyzed prose piece, quickly alludes to Billy's frantic cover-up attempts, followed by a lyrical, touching, and beautifully written passage wherein Billy is reconciled with his father (thus resolving the subliminal Freudian conflicts obvious to any astute reader) and a final exciting police chase scene during which Billy is accidentally shot and killed by a panicky rookie policeman who is coincidentally named Billy. This sentence, although basically in complete sympathy with the laudable efforts of the preceding action-packed sentence, reminds the reader that such allusions to a story that doesn't, in fact, yet exist are no substitute for the real thing and therefore will not get the author (indolent goof-off that he is) off the proverbial hook.

Paragraph. Paragraph. Paragraph. Paragraph. Paragraph. Paragraph. Paragraph. Paragraph. Paragraph. Paragraph. *Paragraph.* Paragraph. Paragraph. Paragraph.

The purpose. Of this paragraph. Is to apologize. For its gratuitous use. Of. Sentence fragments. Sorry.

The purpose of this sentence is to apologize for the pointless and silly adolescent games indulged in by the preceding two paragraphs, and to express regret on the part of us, the more mature sentences, that the entire tone of this story is such that it can't seem to communicate a simple, albeit sordid, scenario.

This sentence wishes to apologies for all the needless apologies found in this story (this one included), which, although placed here ostensibly for the benefit of the more vexed readers, merely delay in a maddeningly recursive way the continuation of the by-now nearly forgotten story line.

This sentence is bursting at the punctuation marks with news of the dire import of self-reference as applied to sentences, a practice that could prove to be a veritable Pandora's box of potential havoc, for if a sentence can refer or allude to itself, why not a lowly subordinate clause, perhaps *this very* clause? Or this sentence fragment? Or three words? Two words? *One?*

Perhaps it is appropriate that this sentence gently and with no trace of condescension remind us that these are indeed difficult and uncertain times and that in general people just aren't nice enough to each other, and perhaps we, whether sentient human beings or sentient sentences, should just *try harder.* I mean, there *is* such a thing as free will, there *has* to be, and this sentence is proof of it! Neither this sentence nor you, the reader, is completely helpless in the face of all the pitiless forces at work in the universe. We should stand our ground, face facts, take Mother Nature by the throat and just *try harder.* By the throat. Harder. Harder, harder.

Sorry.

This is the title of this story, which is also found several times in the story itself.

This is the last sentence of the story. This is the last sentence of the story. This is the last sentence of the story. This is.

Sorry.

Post Scriptum.

As you can see, there is a vast amount of self-referential material out there in the world. To pick only the very best is a monumental task, and certainly a highly subjective one. I would like to include here some of the things that I had to omit from the second self-reference column with great regret, as well as some of the things that were sent in later, in response to it.

First, though, I would like to mention an amusing incident. When Lee Sallows' self-documenting sentence was to be printed in the narrow columns of *Scientific American,* nobody remembered to tell the typesetters not to break any unhyphenated words. As luck would have it, two such breaks were introduced, yielding two spurious hyphens, thus spoiling (in a superficial sense) the accuracy of his construction. How subtly one can get snagged when self-reference is concerned!

Paul Velleman sent me a copy of the front page of the *Ithaca Journal,* dated January 26, 1981, with a banner headline saying "Ex-hostages enjoy their privacy". He wrote, "I think it may be self-referent (and self-contradictory) in a different way than your other examples because the medium, positioning, and size of its printing are all necessary components of the contradiction." When I looked at the page, I simply saw nothing self-referential. I thought maybe I was supposed to look at the flip side, for some reason, but that had even less of interest. So I looked back at the headline, and suddenly it hit me: How can people "enjoy privacy" when it's being blared across the front page of newspapers across the nation?

Along the same lines, soon thereafter I came across a photograph of Lady Di in tears, and in the caption her tears were explained this way: "Lady Di was apparently overcome by the strain of the impending royal wedding and having her every move in public watched by thousands. See story on page A20. Details on the royal honeymoon, page A7."

John M. Lankford wrote me a long letter from Japan on self-reference, remarkably similar in some ways to the one from Flash qFiasco. The most memorable paragraph in his letter was the following one:

> Here in Japan, twice a week, I teach a little class in English for a group of university students—mainly graduate students in the sciences. I spent one class hour taking some of your sentences from the *Scientific American* article, writing them on the blackboard, and asking the students what they meant. The students had a fairly good command of written English, but they were poor in their command of idiom, quick verbal response, and, for want of a better term, "humor of the abstract". As I suspected, many of the sentences—perhaps the most interesting of them—die when ripped from their cultural context. I had quite a bit of difficulty getting across the idea that the pronoun "I" could refer to the sentence as well as to the writer of the sentence. Pronouns cause a lot of trouble in Japan. For example, when I ask someone, "Am I wearing a blue jacket?", they might frequently reply, "Yes, I am wearing a blue jacket." This confusion is easy in Japanese due to the relative lack of pronouns in ordinary speech. Of course you can imagine the extra layers of incomprehension that would arise in reading your sentences if the boundaries between "you" and "I" were rather vague.

On a visit to Gettysburg, I read Abraham Lincoln's Gettysburg address, and for the first time its curious self-reference struck me: "The world will little note nor long remember what we say here." Lincoln had no way of knowing at the time, but this would turn out to be an extremely false sentence (if it is permissible to speak of *degrees* of falsity). In fact, that sentence itself is a very memorable one. While we're on presidential self-reference, listen to this self-descriptive remark by former President Ford: "I am the first to admit that I am no great orator or no person that got where I have gotten by any William Jennings Bryan technique." I guess that where Lincoln's sentence was extremely false, Ford's is extremely true. Here is a final self-referential sentence along presidential lines:

> If John F. Kennedy were reading this sentence, Lee Harvey Oswald would have missed.

* * *

One of the best self-answering questions came up naturally in the course of a very brief

telephone call I made to a restaurant one evening. It went this way: "May I help you?" to which I answered, "You've already helped me—by telling me that you're open today. Thank you. Bye!" And here's a "self-deferential" sentence by Don Byrd: "I am not as witty as my author."

I received this anonymous letter in the mail: "I received this anonymous letter in the mail so I can't credit the author."—so I can't credit the author. I also received a request from someone living in Calgary, Alberta, whose name I forget (but if he's reading this, he'll know who he is) who wrote "This is my feeble way of attempting to get my name into print." I hope this satisfies him.

And now a few miscellaneous examples by me, culled from a second wild binge of self-referential sentence-writing I engaged in not long ago. The first three involve translation issues.

One me has translated at the foot of the letter of the French.

Would not be anomalous if were in Italian.

When one this sentence into the German to translate wanted, would one the fact exploit, that the word order and the punctuation already with the German conventions agree.

How come *this* noun phrase doesn't denote the same thing as *this* noun phrase does?

Every last word in this sentence is a grotesque misspelling of "towmatow".

I don't care *who* wrote this sentence—whoever he is, he's a damn sexist!

This analogy is like lifting yourself by your own bootstraps.

Although this sentence begins with the word "because", it is false.

Despite the fact that it opens like a two-pronged pitchfork—or rather, because of it—this sentence resembles a double-edged sword.

This line from Shakespeare has delusions of grandeur.

If writers were bakers, this sentence would be exactly a dozen words long.

If this sentence had been on the previous page, this very moment would have occurred approximately 60 seconds ago.

This sentence is helping to increase the likelihood of nuclear war by distracting you from the more serious concerns of the world and beguiling you with the trivial joys of self-reference.

This sentence is helping to decrease the likelihood of nuclear war by chiding you for indulging in the trivial joys of self-reference and reminding you of the more serious concerns of the world.

We *mention* "our gigantic nuclear arsenal" in order not to *use* it.

The whole point of this sentence is to make clear what the whole point of this sentence is.

This last one's bizarre circularity reminds me of the number *P* that I invented a couple of years ago. *P* is, for each individual, the number of minutes per month that that person spends thinking about the number *P*. For me, the value of *P* seems to average out at about 2. I certainly wouldn't want it to go much above that! I find it crosses my mind most often when I'm shaving.

* * *

Dr. J. K. Aronson from Oxford, England, sent in some of the most marvelous discoveries. Here is one of his best:

'T' is the first, fourth, eleventh, sixteenth, twenty-fourth, twenty-ninth, thirty-third, . . .

The sentence never ends, of course. He also submitted a wonderful complementary pair that faked me out beautifully. His challenge to you is: Try deciphering the first before you read the second.

I eee oai o ooa a e ooi eee o oe.

Ths sntnc cntns n vwls nd th prcdng sntnc n cnsnnts.

One that reminds me somewhat of Aronson's last sentence above is the following spoof on the ads that I believe you can still find in the New York subway, after all these years:

f y cn rd ths, itn tyg h myxbl cd.

By a remarkable coincidence, the remainder of Carl Bender's sentence "The rest of this sentence is written in Thailand, on" was discovered in, of all places, Bangkok, Thailand, by Gregory Bell, who lives there. He has luckily provided me with a perfect copy of it, so for all those who were dying of suspense, it is shown in Figure 2-3.

One evening during a bad electrical storm, I got the following message on the computer from Marsha Meredith:

I]ion't be able to work at all tonight b]iecause of the w&atherBr/ I]i'm getting too many bad characters (as you can see). 100 baw3d—I get spurious characters]! all over tithe place—talk totrrRBow, 1F7U Marsha.

กระดาษแผ่นนี้และเขียนเป็นภาษาไทย

FIGURE 2–3. *The conclusion of Carl Bender's sentence fragment ("The rest of this sentence is written in Thailand, on"), discovered by Gregory Bell on a scrap of paper in Bangkok, Thailand. Translated, it says: "this sheet of paper and is in Thai".*

I wish she had had the patience to type more carefully, so that I could have understood what her problem was.

The sentences having to do with identity in counterfactual worlds, such as Dan Krimm's and its alter egos, reminded me of a blurb by E. O. Wilson I read recently on Lewis Thomas' latest book: "If Montaigne had possessed a deep knowledge of twentieth-century biology, he would have been Lewis Thomas." Ah me, the flittering elf of self! And Banesh Hoffmann, in *Relativity and Its Roots,* has written: "How safe we would be from death by nuclear bomb had we been born in the time of Shakespeare." Sure, except we'd also all be long dead—unless, of course, the 24th-century doctors who will invent immortality pills had *also* been born in Shakespeare's time!

The following self-referential poem just came to me one day:

Twice five syllables,
Plus seven, can't say much—but . . .
That's haiku for you.

The genre of self-referential poetry—including haiku—was actually quite popular. Tom McDonald submitted this non-limerick:

A very sad poet was Jenny—

Her limericks weren't worth a penny.
In technique they were sound,
Yet somehow she found
Whenever she tried to write any,
That she always wrote one line too many!

Several people sent in complex poems of various sorts, and mentioned books of them, such as John Hollander's *Rhyme's Reason,* a collection of poems describing their own forms.

* * *

Self-referential book titles are enjoying a mild vogue these days. Raymond Smullyan was one of the most enthusiastic explorers of the potential of this idea, using the titles *What Is the Name of This Book?* and *This Book Needs No Title.* Actually, I think *Needs No Title* would have said it more crisply, or maybe just *No Title.* Come to think of it, why not *No,* or even just plain ? (I hope you could tell that those blanks were in *italics!*)

Other self-referential book titles I have collected include these:

Forget all the rules you ever learned about graphic design.
Including the ones in this book.
Steal This Book
Ban This Book
Deduct This Book (How Not to Pay Taxes While Ronald Reagan Is President)
Do You Think Mom Would Like This One?
Dewey Decimal No. 510.46 FC H3
I Never Can Remember What It's Called
The Great American Novel
ISBN 0-943568-01-3
Self-Referential Book Title
The Top Book on the New York Times *Bestseller List for the Past Ten Weeks*
Don't Go Overseas Until You've Read This Book
Soon to Become a Major Motion Picture
By Me, William Shakespeare (by Robert Payne)
That Book with the Red Cover in Your Window
Reviews of This Book

Oh, by the way, some of these are fake, others are real. For example, the last one, *Reviews of This Book,* is just a fantasy of mine. I would love to see a book consisting of nothing but a collection of reviews of it that appeared (after its publication, of course) in major newspapers and magazines. It sounds paradoxical, but it could be arranged with a lot of planning and hard work. First, a group of major journals would all have to agree to run reviews of the book by the various contributors to the book. Then all the reviewers would begin writing. But they would have to mail off their various drafts to all the other reviewers very regularly so that all the reviews could evolve together, and thus eventually reach a stable state of a kind known in physics as a "Hartree-Fock self-consistent solution". Then the book could be published, after which its reviews would come out in their respective journals, as per arrangement. (A little more on this idea is given in the postscript to Chapter 16.)

* * *

I chanced across two books devoted to the subject of indexing books. They are: *A Theory of Indexing* (by Gerald Salton) and *Typescripts, Proofs, and Indexes* (by Judith Butcher). Amazingly, neither one has an index. I also received a curious letter soliciting funds, which began this way: "Dear Friend: In these last months, I've been making a study of the money-raising letter as an art form . . ." I didn't read any further.

Aldo Spinelli, an Italian artist and writer, sent me some of his products. One, a short book called *Loopings,* has pages documenting their own word and letter counts in various complex ways, and includes at the end a short essay on various ways in which documents can tally themselves up or can mutually tally each other in twisty loops. Another, called *Chisel Book,* documents its own production, beginning with the idea, going through the finding of a publisher, making the layout, designing the cover, printing it, and so on.

Ashleigh Brilliant is the inventor of a vast number of aphorisms he calls "potshots", many of which have become very popular phrases in this country. For some reason, he has a self-imposed limit of seventeen words per potshot. A few typical potshots (all taken from his four books listed in the Bibliography) are:

> What would life be, without me?
>
> As long as I have you, I can endure all the troubles you inevitably bring.
>
> Remember me? I'm the one who never made any impression on you.
>
> Why does trouble always come at the wrong time?
>
> Due to circumstances beyond my control, I am master of my fate and captain of my soul.

Although strictly speaking these are not self-referential sentences, they are all admirable examples of how the world constantly tangles with itself in multifarious self-undermining ways, and as such, they definitely belong in this chapter. As a matter of fact, I would like to take this occasion to announce that Ashleigh Brilliant is the 1984 recipient of the last annual Nobaloney Prize for Aphoristic Eloquence. The traditional Nobaloney ceremony, involving the awarding of a $1,000,000 cash prize two minutes before the recipient's decapitation, has been waived, at Mr. Brilliant's request.

There are other books containing much of interest to the self-reference addict. I would particularly recommend the recent *More on Oxymoron,* by Patrick Hughes, as well as the earlier *Vicious Circles and Infinity,* by Hughes and George Brecht. Also in this category are three thin volumes on Murphy's Law, compiled by Arthur Bloch. Murphy's Law, of course, is the one that says, "If anything can go wrong, it will", although when I first heard of it, it was called the "Fourth Law of Thermodynamics". O'Toole's Commentary on Murphy's Law is: "Murphy was an optimist." Goldberg's Commentary thereupon is: "O'Toole was an optimist." And finally, there is Schnatterly's Summing Up: "If anything *can't* go wrong, it will."

My own law, "Hofstadter's Law", states: "It always takes longer than you think it will take, even if you take into account Hofstadter's Law." Despite being its enunciator, I never seem to be able to take it fully into account in budgeting my own time. To help me out, therefore, my friend Don Byrd came up with his own law that I have taken to heart:

> Byrd's Law:
>
> It always takes longer than you think it will take, even if you take into account Hofstadter's Law.

Unfortunately, Byrd himself seems unable to take this law into account.

3

On Viral Sentences and Self-Replicating Structures

January, 1983

TWO years ago, when I first wrote about self-referential sentences, I was hit by an avalanche of mail from readers intrigued by the phenomenon of self-reference in its many different guises. I had the chance to print some of those responses one year ago, and that column then triggered a second wave of replies. Many of them have cast self-reference in new light of various sorts. In this column, I would like to describe the ideas of several people, two of whom responded to my initial column with remarkably similar letters: Stephen Walton of New York City and Donald R. Going of Oxon Hill, Maryland.

Walton and Going saw self-replicating sentences as similar to viruses—small objects that enslave larger and more self-sufficient "host" objects, getting the hosts by hook or by crook to carry out a complex sequence of replicating operations that bring new copies into being, which are then free to go off and enslave further hosts, and so on. "Viral sentences", as Walton called them, are "those that seek to obtain their own reproduction by commandeering the facilities of more complex entities".

Both Walton and Going were struck by the perniciousness of such sentences: the selfish way in which they invade a space of ideas and, merely by making copies of themselves all over the place, manage to take over a large portion of that space. Why do they not manage to overrun all of that idea-space? A good question. The answer should be obvious to students of evolution: competition from other self-replicators. One type of replicator seizes a region of the space and becomes good at fending off rivals; thus a "niche" in idea-space is carved out.

This idea of an evolutionary struggle for survival by self-replicating ideas is not original with Walton or Going, although both had fresh things to say on it. The first reference I know of to this notion is in a passage by neurophysiologist Roger Sperry in an article he wrote in 1965 called "Mind, Brain, and Humanist Values". He says: "Ideas cause ideas and help evolve new ideas. They interact with each other and with other mental forces in the same brain, in neighboring brains, and, thanks to global communication, in far distant, foreign brains. And they also interact with the external surroundings to produce *in toto* a burstwise advance in evolution that is far beyond anything to hit the evolutionary scene yet, including the emergence of the living cell."

Shortly thereafter, in 1970, the molecular biologist Jacques Monod came out with his richly stimulating and provocative book *Chance and Necessity*. In its last chapter, "The Kingdom and the Darkness", he wrote of the selection of ideas as follows:

> For a biologist it is tempting to draw a parallel between the evolution of ideas and that of the biosphere. For while the

abstract kingdom stands at a yet greater distance above the biosphere than the latter does above the nonliving universe, ideas have retained some of the properties of organisms. Like them, they tend to perpetuate their structure and to breed; they too can fuse, recombine, segregate their content; indeed they too can evolve, and in this evolution selection must surely play an important role. I shall not hazard a theory of the selection of ideas. But one may at least try to define some of the principal factors involved in it. This selection must necessarily operate at two levels: that of the mind itself and that of performance.

The performance value of an idea depends upon the change it brings to the behavior of the person or the group that adopts it. The human group upon which a given idea confers greater cohesiveness, greater ambition, and greater self-confidence thereby receives from it an added power to expand which will insure the promotion of the idea itself. Its capacity to "take", the extent to which it can be "put over" has little to do with the amount of objective truth the idea may contain. The important thing about the stout armature a religious ideology constitutes for a society is not what goes into its structure, but the fact that this structure is accepted, that it gains sway. So one cannot well separate such an idea's power to spread from its power to perform.

The "spreading power"—the infectivity, as it were—of ideas, is much more difficult to analyze. Let us say that it depends upon preexisting structures in the mind, among them ideas already implanted by culture, but also undoubtedly upon certain innate structures which we are hard put to identify. What is very plain, however, is that the ideas having the highest invading potential are those that *explain* man by assigning him his place in an immanent destiny, in whose bosom his anxiety dissolves.

Monod refers to the universe of ideas, or what I earlier termed "idea-space", as "the abstract kingdom". Since he portrays it as a close analogue to the biosphere, we could as well call it the "ideosphere".

* * *

In 1976, evolutionary biologist Richard Dawkins published his book *The Selfish Gene*, whose last chapter develops this theme further. Dawkins' name for the unit of replication and selection in the ideosphere—the ideosphere's counterpart to the biosphere's gene—is *meme,* rhyming with "theme" or "scheme". As a library is an organized collection of books, so a memory is an organized collection of memes. And the soup in which memes grow and flourish—the analogue to the "primordial soup" out of which life first oozed—is the soup of human culture. Dawkins writes:

Examples of memes are tunes, ideas, catch-phrases, clothes fashions, ways of making pots or of building arches. Just as genes propagate themselves in the gene pool by leaping from body to body via sperms or eggs, so memes propagate themselves in the meme pool by leaping from brain to brain via a process which, in the broad sense, can be called imitation. If a scientist hears, or reads about, a good idea, he passes it on to his colleagues and students. He mentions it in his articles and his lectures. If the idea catches on, it can be said to propagate itself, spreading from brain to brain. As my colleague N. K. Humphrey neatly summed up an earlier draft of this chapter: '. . . memes should be regarded as living structures, not just metaphorically but technically. When you plant a fertile meme in my mind you literally parasitize my brain, turning it into a vehicle for the meme's propagation in just the way that a virus may parasitize the genetic mechanism of a host cell. And this isn't just a way of talking—the meme for, say, 'belief in life after death' is actually realized physically, millions of times over, as a structure in the nervous systems of individual men the world over.'

Consider the idea of God. We do not know how it arose in the meme pool. Probably it originated many times by independent 'mutation'. In any case, it is very old indeed. How does it replicate itself? By the spoken and written word, aided by great music and great art. Why does it have such high survival value? Remember that 'survival value' here does not mean value for a gene in a gene pool, but value for a meme in a meme pool. The question really means: What is it about the idea of a god which gives it its stability and penetrance in the cultural environment? The survival value of the god meme in the meme pool results from its great psychological appeal. It provides a superficially plausible answer to deep and troubling questions about existence. It suggests that injustices in this world may be rectified in the next. The 'everlasting arms' hold out a cushion against our own inadequacies which, like a doctor's placebo, is none the less effective for being imaginary. These are some of the reasons why the idea of God is copied so readily by successive generations of individual brains. God exists, if only in the form of a meme with high survival value, or infective power, in the environment provided by human culture.

Dawkins takes care here to emphasize that there need not be an exact copy of each meme, written in some universal memetic code, in each person's brain. Memes, like genes, are susceptible to variation or distortion—the analogue to mutation. Various mutations of a meme will have to compete with each other, as well as with other memes, for attention—which is to say, for brain resources in terms of both space and time devoted to that meme. Not only must memes compete for inner resources, but, since they are transmissible visually and aurally, they must compete for

radio and television time, billboard space, newspaper and magazine column-inches, and library shelf-space. Furthermore, some memes will tend to discredit others, while some groups of memes will tend to be internally self-reinforcing. Dawkins says:

> . . . Mutually suitable teeth, claws, guts, and sense organs evolved in carnivore gene pools, while a different stable set of characteristics emerged from herbivore gene pools. Does anything analogous occur in meme pools? Has the god meme, say, become associated with any other particular memes, and does this association assist the survival of each of the participating memes? Perhaps we could regard an organized church, with its architecture, rituals, laws, music, art, and written tradition, as a co-adapted stable set of mutually-assisting memes.
>
> To take a particular example, an aspect of doctrine which has been very effective in enforcing religious observance is the threat of hell fire. Many children and even some adults believe that they will suffer ghastly torments after death if they do not obey the priestly rules. This is a particularly nasty technique of persuasion, causing great psychological anguish throughout the middle ages and even today. But it is highly effective. It might almost have been planned deliberately by a machiavellian priesthood trained in deep psychological indoctrination techniques. However, I doubt if the priests were that clever. Much more probably, unconscious memes have ensured their own survival value by virtue of those same qualities of pseudo-ruthlessness which successful genes display. The idea of hell fire is, quite simply, *self-perpetuating,* because of its own deep psychological impact. It has become linked with the god meme because the two reinforce each other, and assist each other's survival in the meme pool.
>
> Another member of the religious meme complex is called faith. It means blind trust, in the absence of evidence, even in the teeth of evidence . . . . Nothing is more lethal for certain kinds of meme than a tendency to look for evidence . . . . The meme for blind faith secures its own perpetuation by the simple unconscious expedient of discouraging rational inquiry.
>
> Blind faith can justify anything. If a man believes in a different god, or even if he uses a different ritual for worshipping the same god, blind faith can decree that he should die—on the cross, at the stake, skewered on a Crusader's sword, shot in a Beirut street, or blown up in a bar in Belfast. Memes for blind faith have their own ruthless ways of propagating themselves. This is true of patriotic and political as well as religious blind faith.

* * *

When I muse about memes, I often find myself picturing an ephemeral flickering pattern of sparks leaping from brain to brain, screaming "Me, me!" Walton's and Going's letters reinforced this image in interesting ways. For instance, Walton begins with the simplest imaginable viral sentences—"Say me!" and "Copy me!"—and moves quickly to more complex variations with blandishments ("If you copy me, I'll grant you three wishes!") or threats ("Say me or I'll put a curse on you!"), neither of which, he observes, is likely to be able to keep its word. Of course, as he points out, this may not matter, the only final test of viability being success at survival in the meme pool. All's fair in love and war—and war includes the eternal battle for survival, in the ideosphere no less than in the biosphere.

To be sure, very few people above the age of five will fall for the simple-minded threats or promises of these sentences. However, if you simply tack on the phrase "in the afterlife", far more people will be lured into the memetic trap. Walton observes that a similar gimmick is used by your typical chain letter (or "viral text"), which "promises wealth to those who faithfully replicate it and threatens doom to any who fail to copy it". Do you remember the first time you received such a chain letter? Do you recall the sad tale of "Don Elliot, who received $50,000 but then lost it because he broke the chain"? And the grim tale of "General Welch in the Philippines, who lost his life [or was it his *wife*?] six days after he received this letter because he failed to circulate the prayer—but before he died, he received $775,000"? Poor Don Elliot! Poor General Welch! It's hard not to be just a little sucked in by such tales, even if you wind up throwing the letter out contemptuously.

I found Walton's phrases "viral sentence" and "viral text" to be exceedingly catchy—little memes in themselves, definitely worthy of replication some 700,000 times in print, and who knows how many times orally beyond that. At least that's *my* opinion. Of course, it also depends on how the editor of *Scientific American* feels. [It turned out he felt fine about it.] Well, now, Walton's own viral text, as you can see here before your eyes, has managed to commandeer the facilities of a very powerful host—an entire magazine and printing press and distribution service. It has leapt aboard and is now—even as you read this viral sentence—propagating itself madly throughout the ideosphere!

This idea of choosing the right host is itself an important aspect of the quality of a viral entity. Walton puts it this way:

> The recipient of a viral text can, of course, make a big difference. A tobacco mosaic virus that attacks a salt crystal is out of luck, and some people rip up chain letters on sight. A manuscript sent to an editor may be considered viral, even though it contains no explicit self-reference, because it is attempting to secure its own reproduction through an appropriate host; the same manuscript sent to someone who has nothing to do with publishing may have no viral quality at all.

As it concludes, Walton's letter graciously steps forward from the page and squeaks to me directly on its own behalf: "Finally, I (this text) would be delighted to be included, in whole or in part, in your next discussion of self-reference. With that in mind, please allow me to apologize in advance for infecting you."

* * *

Whereas Walton mentioned Dawkins in his letter, Going seems not to have been aware of Dawkins at all, which makes his letter quite remarkable in its close connection to Dawkins' ideas. Going suggests that we consider, to begin with, Sentence *A:*

> It is your duty to convince others that this sentence is true.

As he says:

> If you were foolish enough to believe this sentence, you would attempt to convince your friends that *A* is true. If they were equally foolish, they would convince their friends, and so on until every human mind contained a copy of *A.* Thus, *A* is a self-replicating sentence. More particularly, it is the intellectual equivalent of a virus. If Sentence *A* were to enter a mind, it would take control of the mind's intellectual machinery and use it to produce hundreds of copies of itself in other minds.
>
> The problem with Sentence *A,* of course, is that it is absurd; no one could possibly believe it. However, consider the following:
>
> System *S:*
> Begin:
> S_1: Blah.
> S_2: Blah blah.
> S_3: Blah blah blah.
> . .
> . .
> . .
> . .
> . .
> . .
> S_{99}: Blah blah blah blah blah blah . . . . . .
> S_{100}: It is your duty to convince others that System *S* is true.
> End.
>
> Here, S_1 through S_{99} are meant to be statements that constitute a belief system having some degree of coherency. If System *S* taken as a whole were convincing, then the entire system would be self-replicating. System *S* would be especially convincing if S_{100} were not stated explicitly but held as a logical consequence of the other ideas in the system.

Let us refer to Going's S_{100} as the *hook* of System *S,* for it is by this hook that System *S* hopes to hoist itself onto a higher level of power. Note that on its own, a hook that in effect says "It is your duty to believe me" is not a viable viral entity; in order to "fly", it needs to drag something extra along with it, just as a kite needs a tail to stabilize it. Pure lift goes out of control and self-destructs, but controlled lift can lift itself along with its controller. Similarly, S_{100} and S_1–S_{99} (taken as a set) are symbiotes: they play complementary, mutually supportive roles in the survival of the meme

they together constitute. Now Going develops this theme a little further:

> Statements S_1–S_{99} are the bait which attracts the fish and conceals the hook. No bait—no bite. If the fish is fool enough to swallow the baited hook, it will have little enough time to enjoy the bait. Once the hook takes hold, the fish will lose all its fishiness and become instead a busy factory for the manufacture of baited hooks.
>
> Are there any real idea systems that behave like System *S*? I know of at least three. Consider the following:
>
> System X:
> Begin:
> X_1: Anyone who does not believe System *X* will burn in hell.
> X_2: It is your duty to save others from suffering.
> End.
>
> If you believed in System *X*, you would attempt to save others from hell by convincing them that System *X* is true. Thus System *X* has an implicit 'hook' that follows from its two explicit sentences, and so System *X* is a self-replicating idea system. Without being impious, one may suggest that this mechanism has played some small role in the spread of Christianity.
>
> Self-replicating ideas are most often found in politics. Consider Sentence *W:*
>
> The whales are in danger of extinction.
>
> If you believed this idea, you would want to save the whales. You would quickly discover that you could not reach this goal by yourself. You would need the help of thousands of like-minded people. The first step in getting their help would be to convince them that Sentence *W* is true. Thus a 'hook' like S_{100} follows from Sentence *W*, and Sentence *W* is a self-replicating idea.
>
> In a democracy, nearly any idea will tend to replicate since the only way to win an election is to convince other people to share your ideas. Most political ideas are not properly self-replicating, since the motive for spreading the idea is separate from the idea itself. Statement *W*, on the other hand, is genuinely self-replicating, since the duty to propagate it is a direct logical consequence of *W* itself. Ideas like *W* can sometimes take on a life of their own and drive their own propagation.
>
> A more sinister form of self-replication is Sentence *B:*
>
> The bourgeoisie is oppressing the proletariat.
>
> This statement is self-replicating for the same reason as *W* is. The desire to propagate statements like *B* is driven by a desire to protect a victim figure from a villain figure. Such ideas are dangerous because belief in them may lead to attacks on the supposed villain. Statement *B* also illustrates the fact that the self-replicating character of an idea depends only upon the idea's logical structure, not upon its truth.
>
> Statement *B* is merely a special case of the generalized statement, Sentence *V:*
>
> The *villain* is *wronging* the *victim.*
>
> Here, the word *villain* must be replaced with the name of some real group (capitalists, communists, imperialists, Jews, freemasons, aristocrats, men, foreigners, etc.). Likewise, *victim* must be replaced with the name of the corresponding victim and *wronging* filled in as desired. The result will be a self-replicating idea system for the same reasons as *W* and *B* were. Note that each of the suggested substitutions yields a historically attested idea system. It has long been recognized that most extremist mass movements are based on a belief similar to *V*. Part of the reason seems to be that type-*V* statements reduce to the 'hook', S_{100}, and therefore define self-replicating idea systems. One hesitates to explain real historical events in terms of such a silly mechanism, and yet . . . .

Going brings his ideas to an amusing conclusion as follows:

> Suppose we parody my thesis by proposing Sentence *E:*
>
> The self-replicating ideas are conspiring to enslave our minds.
>
> This 'paranoid' statement is clearly an idea of type *V*. Thus, the thesis seems to describe itself. Further, if we accept *E*, then we must say that this type-*V* idea implies that we must distrust all ideas of type *V*. This is the Epimenides Paradox.

It is interesting that all these people who have explored these ideas have given examples ranging from the very small scale of such things as catchy tunes (for example, Dawkins cites the opening theme of Beethoven's fifth symphony) and phrases (the word "meme" itself) to the very large scale of ideologies and religions. Dawkins uses the term *meme complex* for these larger

agglomerations of memes; however, I prefer the single word *scheme.*

One reason I prefer it is that it fits so well with the usage suggested by psychiatrist and writer Allen Wheelis in his novel *The Scheme of Things.* Its central character is a psychiatrist and writer named Oliver Thompson, whose darkly brooding essays are scattered throughout the book, interspersed with brightly colored, evocative episodes. Thompson is obsessed with the difference between, on the one hand, "the raw nature of existence, unadorned, unmediated", which he refers to repeatedly as "the way things are", and, on the other hand, "schemes of things", invented by humans—ways of making order and sense out of the way things are. Here are some of Thompson's musings on that theme:

> I want to write a book . . . . the story of one man whose life becomes a metaphor for the entire experience of man on earth. It will portray his search through a succession of schemes of things, show the breakdown, one after another, of each pattern he finds, his going on always to another, always in the hope that the scheme of things he finds and for the moment is serving is *not* a scheme of things at all but reality, the way things are, therefore an absolute that will endure forever, within which he can serve, to which he can contribute, and through which he can give his mortal life meaning and so achieve eternal life . . . .
>
> The scheme of things is a system of order. Beginning as our view of the world, it finally *becomes* our world. We live within the space defined by its coordinates. It is self-evidently true, is accepted so naturally and automatically that one is not aware of an act of acceptance having taken place. It comes with one's mother's milk, is chanted in school, proclaimed from the White House, insinuated by television, validated at Harvard. Like the air we breathe, the scheme of things disappears, becomes simply reality, the way things are. It is the lie necessary to life. The world as it exists beyond that scheme becomes vague, irrelevant, largely unperceived, finally nonexistent . . . .
>
> No scheme of things has ever been both coextensive with the way things are and also true to the way things are. All schemes of things involve limitation and denial . . . .
>
> A scheme of things is a plan for salvation. How well it works will depend upon its scope and authority. If it is small, even great achievement in its service does little to dispel death. A scheme of things may be as large as Christianity or as small as the Alameda County Bowling League. We seek the largest possible scheme of things, not in a reaching out for truth, but because the more comprehensive the scheme the greater its promise of banishing dread. If we can make our lives mean something in a cosmic scheme we will live in the certainty of immortality. Those attributes of a scheme of things that determine its durability and success are its scope, the opportunity it offers for participation and contribution, and the conviction with which it is held as self-evidently true. The very great success of Christianity for a thousand years follows upon its having been of universal scope, including and accounting for everything, assigning to all things a proper place; offering to every man, whether prince or beggar, savant or fool, the privilege of working in the Lord's vineyard; and being accepted as true throughout the Western world.
>
> As a scheme of things is modified by inroads from outlying existence, it loses authority, is less able to banish dread; its adherents fall away. Eventually it fades, exists only in history, becomes quaint or primitive, becomes, finally, a myth. What we know as legends were once blueprints of reality. The Church was right to stop Galileo; activities such as his import into the regnant scheme of things new being which will eventually destroy that scheme.

Taken in Wheelis' way, "scheme" seems a fitting replacement for Dawkins' "meme complex". A scheme imposes a top-down kind of perceptual order on the world, propagating itself ruthlessly, like Going's System *S* with its "hook". Wheelis' description of the inadequacy of all "schemes of things" to fully and accurately capture "the way things are" is strongly reminiscent of the vulnerability of all sufficiently powerful formal systems to either incompleteness or inconsistency —a vulnerability that ensues from another kind of "hook": the famous Gödelian hook, which arises from the capacity for self-reference of such systems, although neither Wheelis nor Thompson makes any mention of the analogy. We shall come back to Gödel momentarily.

* * *

The reader of this novel must be struck by the professional similarity of Wheelis and his protagonist. It is impossible to read the book and not to surmise that Thompson's views are reflecting Wheelis' own views—and yet, who can say? It is a tease. Even more tantalizing is the title of Thompson's imaginary book, which Wheelis casually mentions toward the end of the novel: it is *The Way Things Are*—a striking contrast to the title of the real book in which it exists. One wonders: What is the meaning of this elegant literary pleat in which one level folds back on another? What is the symbolism of Wheelis within Wheelis?

Such a twist, by which a thing (sentence, book, system, person) seems to refer to itself but does so only by allusion to something *resembling* itself, is called *indirect self-reference.* You can do this by pointing at your image in a mirror and saying, "That person sure is good-looking!" That one is very simple, because the connection between something and its mirror image is so familiar and obvious-seeming to us that there seems to be no distance whatsoever between direct and indirect referents: we equate them completely. Thus it seems there is no referential indirectness.

On the other hand, this depends upon the ease with which our perceptual systems convert a mirror image into its reverse, and upon other qualities of our cognitive systems that allow us to see through several layers of translation without being aware of the layers—like looking through many feet of water and seeing not the water but only what lies at its bottom.

Some indirect self-references are of course subtler than others. Consider the case of Matt and Libby, a couple ostensibly having a conversation about their friends Tammy and Bill. It happens that Matt and Libby are having some problems in their relationship, and those problems are quite analogous to those of Tammy and Bill, only with sexes reversed: Matt is to Libby what Tammy is to Bill, in their respective relationships. So as Matt and Libby's conversation progresses, although on the surface level it is completely about their friends Tammy and Bill, on another level it is *actually* about themselves, as reflected in these other people. It is almost as if, by talking about Tammy and Bill, Matt and Libby are going over a fable by Aesop that has obvious relevance to their own plight. There are things going on simultaneously on two levels, and it is hard to tell how conscious either of the participants is of the exchange of dual messages—one of concern about their friends, one of concern about themselves.

Indirect self-reference can be exploited in the most unexpected and serious ways. Consider the case of President Reagan, who on a recent occasion of high Soviet-American tension over Iran, went out of his way to recall President Truman's behavior in 1945, when Truman made some very blunt threats to the Soviets about the possibility of the U.S. using nuclear weapons if need be against any Soviet threat in Iran. Merely by bringing up the memory of that occasion, Reagan was inviting a mapping to be made between himself and Truman, and thereby he was issuing a not-so-veiled threat, though no one could point to anything explicit. There simply was no way that a conscious being could fail to make the connection. The resemblance of the two situations was too blatant.

Thus, does self-reference really come in two varieties—direct and indirect—or are the two types just distant points on a continuum? I would say unhesitatingly that it is the latter. And furthermore, you can delete the prefix "self", so that the question becomes one of reference in general. The essence is simply that one thing refers to another whenever, to a conscious being, there is a sufficiently compelling mapping between the roles the two things are perceived to play in some larger structures or systems. (See Chapter 24 for further discussion of the perception of such roles.) Caution is needed here. By "conscious being", I mean an analogy-hungry perceiving machine that gets along in the world thanks to its perceptions; it need not be human or even organic. Actually, I would carry the abstraction of the term "reference" even further, as follows. The mapping of systems and roles that establishes reference need not actually be *perceived* by any such being: it suffices that the mapping exist and simply be *perceptible* to such a being were it to chance by.

* * *

The movie *The French Lieutenant's Woman* (based on John Fowles' novel of the same name) provides an elegant example of ambiguous degrees of reference. It consists of interlaced vignettes from two concurrently developing stories both of which involve complex romances; one takes place in Victorian England, the other in the present. The fact that there are two romances already suggests, even if only slightly, that a mapping is called for. But much more is suggested than that.

first "Johnnie" award—a self-replicating dollar bill given to the Grand Winner of the First Every-Other-Decade Von Neumann Challenge. Unfortunately, the dollar bill consumes the entire body of its owner in its bizarre process of self-replication, and so it is wisest to simply lock it up to protect oneself from its voracious appetite.

Palmer submitted several versions. In them, he utilized upper and lower cases to distinguish between seed and building rule, respectively. Here is one solution, slightly modified by me:

> *after alphabetizing, decapitalize FOR AFTER WORDS STRING FINALLY UNORDERED UPPERCASE FGPBVKXQJZ NONVOCALIC DECAPITALIZE SUBSTITUTING ALPHABETIZING, finally for nonvocalic string substituting unordered uppercase words*

Let us watch how it works, step by careful step. We must bear in mind that the instructions we are following are the lowercase words printed above, and that the uppercase words are not to be read as instructions. Nor, for that matter, are the lowercase words that we will soon be working with. They are like the inert, anesthetized body of a patient being operated on, who, when the operation is over, will awake and become animate. So let's go. First we are to alphabetize the seed. (I am treating the comma as attached to the word preceding it.) This gives us the following:

> *AFTER ALPHABETIZING, DECAPITALIZE FGPBVKXQJZ FINALLY FOR NONVOCALIC STRING SUBSTITUTING UNORDERED UPPERCASE WORDS*

Next we are to decapitalize it. This will yield some lowercase words—the "anesthetized" lowercase words I spoke of above:

> *after alphabetizing, decapitalize fgpbvkxqjz finally for nonvocalic string substituting unordered uppercase words*

All right; now our final instruction is to locate a nonvocalic string (that's easy: "*fgpbvkxqjz*") and to substitute for it the uppercase words, *in any order* (that is, the original seed itself, but without regard for its structure above the level of the individual word-unit). This last bit of surgery yields:

> *after alphabetizing, decapitalize SUBSTITUTING FINALLY WORDS UNORDERED STRING DECAPITALIZE UPPERCASE FOR NONVOCALIC AFTER FGPBVKXQJZ ALPHABETIZING, finally for nonvocalic string substituting unordered uppercase words*

And this is a perfect copy of our starting sentence! Or rather, semiperfect. Why only semiperfect? Because the seed has been randomly scrambled in the act of self-reproduction. The beauty of the scheme, though, is that the internal structure of the seed is entirely irrelevant to the efficacy of the sentence as a self-replicator. All that matters is that the new building rule say the proper thing, and it will do so no matter what order the seed from which it sprang was in. Now this fresh new baby sentence can wake up from its anesthesia and go off to replicate itself in turn.

The critical step was the first one: alphabetization. This turns the arbitrarily-ordered seed into a grammatical, meaningful command—merely by mechanically exploiting a presumed knowledge of the "ABC"s. But why not? It is perfectly reasonable to presume superficial typographical knowledge about letters and words, since such knowledge deals with printed material *as raw material:* purely syntactically, without regard to the meanings carried therein. This is just like the way that enzymes in the living cell deal with the DNA and RNA they chop up and alter and piece together again: purely chemically, without regard to the "meanings" carried therein. Just as chemical valences and affinities and so on are taken as givens in the workings of the cell, so alphabetic and typographic facts are taken as givens in the V. N. Challenge.

When Palmer sent in his solution, he happened to write down his seed in order of increasing length of words, but that is inessential; any random order would have done, and that sort of idea is the crucial point that many readers missed. Another rather elegant solution was sent in by Martin Weichert of Munich. It runs this way (slightly modified by me):

> Alphabetize and append, copied in quotes, these words: "these append, in Alphabetize and words: quotes, copied"

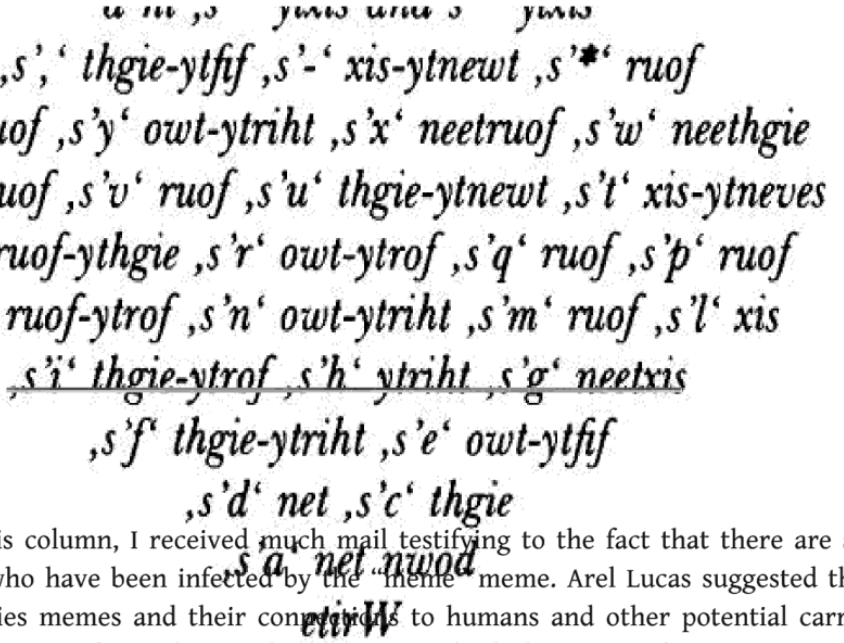
It works on the same principle as Palmer's sentence, and again features a seed whose internal structure (at least at the word level) is irrelevant to successful self-replication. Weichert also sent along an intriguing palindromic solution in Esperanto, in which the flexible word order of the language plays a key role. Michael Borowitz and Bob Stein of Durham, North Carolina sent in a solution similar to Palmer's.

* * *

Finally, last year's gold-medal winner for self-documentation, Lee Sallows, was a bit piqued by my suggestion that the gold on his medal was somewhat tarnished since he had not paid close enough attention to the use-mention distinction. Apparently I goaded him into constructing an even more elaborate self-documenting sentence. Although it does not quite fit what I had in mind for the Von Neumann Challenge, as it does not spell out its own construction explicitly at the letter level or word level, it is another marvelous Sallowsian gem, and I shall therefore generously allow the gold on his medal to go untarnished this year. (Apologies to those purists who insist that gold doesn't tarnish. I must have been confusing it with copper and silver. How silly of me!) Herewith follows Sallows' 1982 contribution:

*

Write
down ten 'a's,
eight 'c's, ten 'd's,
fifty-two 'e's, thirty-eight 'f's,
sixteen 'g's, thirty 'h's, forty-eight 'i's,
six 'l's, four 'm's, thirty-two 'n's, forty-four 'o's,
four 'p's, four 'q's, forty-two 'r's, eighty-four 's's,
seventy-six 't's, twenty-eight 'u's, four 'v's, four 'W's,
eighteen 'w's, fourteen 'x's, thirty-two 'y's, four ':'s,
four ''s, twenty-six '-'s, fifty-eight ','s,*
sixty "'s and sixty "'s, in a
palindromic sequence
whose second
half runs
thus:
:suht
snur flah
dnoces esohw
ecneuqes cimordnilap
a ni s''' ytwis dna s''' ytwis



Post Scriptum

After writing this column, I received much mail testifying to the fact that there are a large number of people who have been infected by the "meme" meme. Arel Lucas suggested that the discipline that studies memes and their competitions to humans and other potential carriers of them be known as *memetics*, by analogy with "genetics". I think this is a good suggestion, and hope it will be adopted.

Maurice Guéron wrote me from Paris to tell me that he believed the first clear exposition of the idea of self-reproducing ideas that inhabit the brains of organisms was put forward in 1952 by Pierre Auger, a physicist at the Sorbonne, in his book *L'homme microscopique.* Guéron sent me a photocopy of the relevant portions, and I could indeed see how prophetic the book was.

I received a copy of the book *General Theory of Evolution* by Vilmos Csányi, a Hungarian geneticist. In this book, he attempts to work out a theory in which memes and genes evolve in parallel. A similar attempt is made in the book *Ever-Expanding Horizons: The Dual Informational Sources of Human Evolution,* by the American biologist Carl B. Swanson.

The most thorough-going research on the topic of pure memetics I have yet run across is that of Aaron Lynch, an engineering physicist at Fermilab in Illinois, who in his spare time is writing a book called *Abstract Evolution.* The portions that I have read go very carefully into the many "options", to speak anthropomorphically, that are open to a meme for getting itself reproduced over and over in the ideosphere (a term Lynch and I invented independently). It promises to be a provocative book, and I look forward to its publication.

* * *

Jay Hook, a mathematics graduate student, was provoked by the solutions to the Von Neumann Challenge as follows:

> The notion that it takes two to reproduce is suggestive. Perhaps a change in terminology is appropriate. The component that you call the "seed" might be thought of as the "female" fragment—the egg that grows into an adult, but only after receiving instructions from the sperm, the "male" fragment—the building rule. In this interpretation, our sentences say everything twice because they are hermaphroditic: the male and female fragments appear together in the same individual.
>
> To better mimic nature, we should construct *pairs* of sentences or phrases, one male and one female—expressions that taken individually produce nothing but when put together in a dark room make copies of themselves. I propose the following. The male fragment
>
> After alphabetizing and deitalicizing, duplicate female fragment in its original version.
>
> doesn't seem to say much by itself, and the female fragment
>
> *in and its After female fragment original version. duplicate alphabetizing deitalicizing,*
>
> certainly doesn't, but let them at each other and watch the fireworks. (I follow your practice of assuming each punctuation mark to be attached to the preceding word.) The male takes the lead, and sets to work on the female. First

we alphabetize and deitalicize her, he says; that gives a new male fragment. Then we simply make a copy of her—so we get one of each!

Nature still doesn't work this way, of course; it's not clear that couples that produce offspring only in boy-girl pairs are really superior to self-replicating hermaphrodites. Ideally, our fragments should produce *either* a copy of the male *or* a copy of the female, depending on, say, the day of the week or the parity of some external index like the integer part of the current Dow Jones Industrial Average. Surprisingly, this isn't hard. Take the male to be

Alphabetize and deitalicize female fragment if index is odd; otherwise reproduce same verbatim.

and take for the female

if is and odd; same index female fragment otherwise reproduce verbatim. Alphabetize deitalicize

One more refinement. To this point, each offspring has been exactly identical to one of its parents. We can introduce variation, at least in the girls, as follows. Male fragment:

Alphabetize and deitalicize female fragment if index is odd; otherwise randomly rearrange the words.

Female fragment:

if is and the odd; index female words, fragment randomly otherwise rearrange Alphabetize deitalicize

Now all of the boys will be the spittin' image of their father, but whereas one daughter might be

index rearrange if the Alphabetize randomly fragment odd; deitalicize is and words. otherwise female

another might be

Alphabetize index and rearrange the fragment if female is odd; otherwise randomly deitalicize words.

The important point, however, is that all of these female offspring, however diverse, are genetically capable of mating with any of the (identical) males. Can you find a way to introduce variation in the males without producing sterile offspring?

In conclusion, allow me to observe that the Dow closed on Friday at 1076.0. Therefore I proudly proclaim: It's a girl!

* * *

I now close by returning to Lee Sallows. This indefatigable researcher of what he calls *logological space* continued his quest after the holy grail of *perfect* self-documentation. His jealousy was aroused in the extreme when Rudy Kousbroek, who is Dutch, and Sarah Hart, who is English, together tossed off what Sallows terms "the greatest logological jewel the world has ever seen". Kousbroek and Hart's self-documenting sentence, though in Dutch, ought to be pretty clearly understandable by anyone who takes the time to look at it carefully:

Dit pangram bevat vijf a's, twee b's, twee c's, drie d's, zesenveertig e's, vijf f's, vier g's, twee h's, vijftien i's, vier j's, een k, twee l's, twee m's, zeventien n's, een o, twee p's, een q, zeven r's, vierentwintig s's, zestien t's, een u, elf v's, acht w's, een x, een y, en zes z's.

In fact, you can learn how to count in Dutch by studying it!

There's not an ounce of fat or awkwardness in this sentence, and it drove Sallows mad that he couldn't come up with an equally perfect pangram (sentence containing every letter of the alphabet) in English. Every attempt had some flaw in it. So in desperation, Sallows, electronics engineer that he is, decided he would design a high-speed dedicated "letter-crunching" machine to search the far reaches of logological space for an equivalent English sentence. Sallows sent me some material on his Pangram Machine. He says:

At the heart of the beast is a clock-driven cascade of sixteen Johnson-counters: the electronic analogue of a stepper-motor-driven stack of combination lock-discs. Every tick of the clock clicks in a new combination of numbers: a unique combination of counter output lines becomes activated . . . . Pilot tests have been surprisingly encouraging; it looks as though a clock frequency of a million combinations per second is quite realistic. Even so it would take 317 years to explore the ten-deep stratum. But does it have to be ten? With this reduced to a modest but still very worthwhile six-deep range it will take just 32.6 days. Now we're talking!

Over the past eight weeks I have devoted every spare second to constructing this rocket for exploring the far regions of logological space . . . . Will it really fly? So far it looks very promising. And the end is already in sight. With a bit of luck Rudy Kousbroek will be able to launch the machine on its 32-day journey when he comes to visit here at the end of

this month. If so, a bottle of champagne will not be out of place.

Two months later, I got a most excited transmission from Lee, which began with the word "EUREKA!"—the word the Pangram Machine was set up to print on success. He then presented three pangrams that his machine had discovered, floating "out there" somewhere beyond the orbit of Pluto. My favorite one is this:

> This pangram tallies five a's, one b, one c, two d's, twenty-eight e's, eight f s, six g's, eight h's, thirteen i's, one j, one k, three l's, two m's, eighteen n's, fifteen o's, two p's, one q, seven r's, twenty-five s's, twenty-two t's, four u's, four v's, nine w's, two x's, four y's, and one z.

Now that's what I call a success for mechanical translation!

Sallows writes: "I wager ten guilders that nobody will succeed in producing a perfect self-documenting solution (or proof of its non-existence) to the sentence beginning, 'This computer-generated pangram contains . . .' *within the next ten years.* No tricks allowed. The format to be exactly as in the above pangrams. Either 'and' or '&' is permissible. Result to be derived exclusively by von Neumann architecture digital computer (no super computers, no parallel processing). Fancy your chances?" Anyone who wants to write to Sallows can do so, at Buurmansweg 30, 6525 RW Nijmegen, Holland.

Much though I am delighted by Sallows' ingenious machine and his plucky challenge, I expect him to lose his wager before you can say "Raphael Robinson". For my reasons, see the postscript to Chapter 16.

the report of this committee was unfavorable. What finally resulted I do not know.

Parliamentary procedure too can lead to the most tangled of situations. For example, there are several editions of *Robert's Rules of Order,* and a body must choose which set of rules will govern its deliberations. The latest edition of *Robert's Rules* states that if no specific edition is chosen as the governing one, then the most recent issue holds. A problem arises, though, if one hasn't adopted the latest edition, since one cannot then rely on its authority to tell one to rely on it.

In some ways, parliamentary procedure, which deals with how to handle simultaneous and competing claims for attention, bears a remarkable resemblance to the way a large computer system must manage its own internal affairs. Within such a system, there is always a program called an *operating system* with a part called the *scheduling algorithm,* which weighs priorities and decides which activity will proceed next. In a "multiprocessing" system, this means determining which activity gets the next "time slice" (lasting for anywhere from a millisecond to a few seconds, or possibly even for an unlimited time, depending on the activity's priority and numerous other factors). But there are also *interrupts* that come and interfere with—oops, just a moment, my telephone's ringing. Be right back. There. Sorry we were disturbed. Someone wanted to sell me a telephone-answering system. Now what would—ah, ah, just a sec—ah-choo!—sorry—what would I do with one of those things? Now where was I? Oh, yes—interrupts. Well, in a way they are like telephone calls that take the store clerk away from you, annoying you in the extreme, since you have come to the store in person, whereas the telephone caller has been lazy and yet is given higher priority.

A good scheduling algorithm strives to be equitable, but all kinds of conflicts can arise, in which interrupts interrupt interrupts and are then themselves interrupted. Moreover, the scheduler has to be able to run its own internal decision-making programs with high priority, yet not so high a priority that nothing else ever runs. Sometimes the internal and external priorities can become so tangled that the entire system begins to "thrash". This is the term used to describe a situation where the operating system is spending most of its time bogged down in "introverted" computation, deciding what it should spend its time doing. Needless to say, during periods of thrashing, very little "real" computation gets done. It sounds quite like the cognitive state a person can get into when too many factors are weighing down all at once and the slightest thought on any topic seems to trigger a rash of paradoxical dilemmas from which there is no escape. Sometimes the only solution is to go to sleep, and let the paradoxes somehow drift away into a better perspective.

* * *

Operating systems and courts of law cannot, unfortunately, go to sleep. Their snarls are very real, and some means of dealing with them has to be invented. It was considerations such as this that led Peter Suber to invent his tangled game of Nomic.

He writes that he was struck by the oft-heard cynicism that "Government is just a game." Now, one essential activity of government is law-making, so if it is a game, then it is a game in which changing the laws (or rules) is a move. Moreover, some rules are needed to structure the process of changing the rules. Yet no legal system seems to have any rules that are absolutely immune to legal change. Suber's main aim, he wrote, was "to make a playable game that models this particular situation. But whereas governments are at any given moment pushed in various directions in their rule-changing by historical realities and the ideology of their people and existing rules, I wanted the game to start with as 'clean' an initial set of rules as possible." Nomic is such a game, and its rules (or rather, its Initial Set of rules) will be presented below. Most of the following description is in essence by Suber himself. I have simply interspersed some of my own observations.

In legal systems, statutes are the paradigmatic rules. Statutes are made by a rule-governed process that is itself partly statutory; hence the power to make and change statutes can reach some

of the rules governing the process itself. Most of the rules, however, that govern the making of statutes are constitutional and are therefore beyond the reach of the power they govern. For instance, Congress may change its parliamentary rules and its committee structure, and it may bind its future action by its past action, but it cannot, through mere statutes, alter the fact that a two-thirds "supermajority" is needed to override an executive veto, nor can it abolish or circumvent one of its houses, start a tax bill in the Senate, or even delegate too much of its power to experts.

Although statutes cannot affect constitutional rules, the latter can affect the former. This is an important difference of logical priority. When there is a conflict between rules of different types, the constitutional rules always prevail. This *logical* level-distinction is matched by a *political* level-distinction—namely, that the logically prior (constitutional) rules are more difficult to amend than the logically posterior (statutory) rules.

It is no coincidence that logically prior laws are harder to amend. One purpose of making some rules more difficult to change than others is to prevent a brief wave of fanaticism from undoing decades or even centuries of progress. This could be called "self-paternalism": a deliberate retreat from democratic principles, although one chosen for the sake of preserving democracy. It is our chosen insurance against our anticipated weak moments. But that purpose will not be met unless the two-tier (or multi-tier) system also creates a logical hierarchy in which the less mutable rules take logical priority over the more mutable rules; otherwise, the more mutable rules could by themselves undo the deeper and more abstract principles on which the whole system is based. If supermajorities and the concurrence of many bodies are necessary to protect the foundations of the system from hasty change, that protective purpose is frustrated if those foundations are reachable by rules requiring merely a simple majority of one legislature.

Although all the rules in the American system are mutable, it is convenient to refer to the less mutable constitutional rules as *immutable,* and to the more mutable rules below them in the hierarchy as *mutable.* The same is true in Nomic, where, at least initially, no rule is literally immutable. If Nomic's self-paternalism is to be effective, then, its "immutable" rules, in addition to resisting easy amendment, must possess logical priority.

Many designs could satisfy this requirement. Nomic has adopted a simple two-tiered system, modeled to some extent on the U.S. Constitution. In principle, a system could have any number of degrees of difficulty in the amendment of rules. For instance, Class *A* rules, the hardest to amend, could require unanimity of a central body and the unanimous concurrence of all regional bodies. Class *B* rules could require 90 percent supermajorities, Class *C* rules 80 percent supermajorities, and so on. The number of such categories could be indefinitely large.

Indeed, if appropriate qualifications are made for the informality of custom and etiquette, a strong argument could be made that normal social life is just such a system of indefinite tiers. Near the top of the "difficult" end of the series of rules are actual laws, rising through case precedents, regulations, and statutes, all the way up to constitutional rules. At the bottom of the scale are rules of personal behavior that individuals can amend unilaterally without incurring disapprobation or censure. Above these are rules for which amendment is increasingly costly, starting with costs on the order of furrowed brows and clucked tongues, and passing through indignant blows and vengeful homicide.

* * *

In any case, for the sake of simplicity and to make it easier to learn and play, Nomic is a clean two-tier system rather than a nuanced or multi-tier system like the U.S. Government, with its intermediate and substatutory levels such as parliamentary rules, administrative regulations, joint resolutions, treaties, executive agreements, higher and lower court decisions, state practice, judicial rules of procedure and evidence, executive orders, canons of professional responsibility,

evidentiary presumptions, standards of reasonableness, rules establishing priority among rules, canons of interpretation, contractual rules, and so on. This is not to say that nuanced, intermediate levels may not arise in Nomic through game custom and tacit understandings. In fact, the nature of the game allows players to add new tiers by explicit amendment as they see fit, and one reason for making Nomic simple initially is that it is easier to add tiers to a simple game than it is to subtract them from a complex one.

Nomic's two-tier system embodies the same self-paternalistic elements as does the Federal Constitution. The "immutable" rules govern more basic processes than the "mutable" ones do, and thus shield them from hasty change. Since, in the course of play, the central core of the game may change (and the minor aspects *must* change), after a few rounds the game being played by the players may in a certain sense be different from the one they were playing when they started. Yet needless to say, whatever results from compliance with the rules is, by definition, the game Nomic. The "feel" of the game may change drastically even as, at a deeper level, the game remains the same.

In a similar way, human beings undergo constant development and self-modification, and yet continue to be convinced that it makes sense to refer, via such words as "I", to an underlying stable entity. The more immediately perceptible patterns change, whereas deeper and more hidden patterns remain the same. From birth to maturity to death, however, the changes can be so radical that one may sometimes feel that in a single lifetime one is several different people. Similarly, in law, many have acknowledged that an amendment clause (a clause defining how a constitution may be amended)—even a clause limited to piecemeal amendment—could, through repeated application, create a fundamentally new constitution.

The fact that Nomic has more than one tier prevents the logical foundation of the game—the central core—from changing radically in just a few moves. Such continuity is a virtue both of games and of governments, but players of Nomic have an advantage over citizens in that, whenever they are so motivated, they can adjust the degree of continuity and the rate of change rather quickly, using their wits, whereas in real life the mechanisms by which such change could be effected are barely known and partially beyond reach.

Standard games possess the continuity of unchanging rules, or at least of rules that change only between games, not during them. Nomic's continuity is more like that of a legal system than that of a standard game: it is a rule-governed set of systems, directives, and processes undergoing constant rule-governed change. If, however, one wants a specific entity to point to as being "Nomic itself", the Initial Set of rules, as presented below, will do. Yet Nomic is equally the product, at any given moment, of the dynamic rule-governed change of the Initial Set. The continuing identity of the game, like that of a nation or person, is due to the fact (if fact it is) that all change is the product of existing rules properly applied, and that no change is revolutionary. (One could even argue that revolutionary change is just more of the same: In a revolution, rules that have been assumed to be totally immutable simply are rendered mutable by other rules that are more deeply immutable, but that previously had been taken for granted and hence had been invisible, or tacit.)

* * *

In its Rule 212, Nomic includes provision for subjective *judgment* (as in a court of law), not merely to imitate government in yet another aspect, but for the same reasons that compel government itself to make provisions for judgment: rules will inevitably be made that are ambiguous, inconsistent, or incomplete, or that require application to individual circumstance. "Play" must not be interrupted; therefore some agency must be empowered to make an authoritative and final determination so that play can continue.

Judgments in Nomic are not bound by rules of precedent, since that would require a daunting amount of record-keeping for each game. But the doctrine of *stare decisis* (namely, that precedents

should be followed) may be imposed at the players' option, or it may arise without explicit amendment, as successive judges feel impelled to treat "similarly situated" persons "similarly". (Admittedly, the meanings of these terms in specific cases may well require further levels of judgment. This fact is one of the most dangerous sources of potential infinite regress in real court cases.) Without *stare decisis,* the players are constrained to draft their rules carefully, make thoughtful adjudications, overrule poor judgments, and amend defective rules. This is one way Nomic teaches basic principles and exigencies of law, even as it vastly simplifies.

The Initial Set must be short and simple enough to encourage play, yet long and complex enough to cover contingencies likely to arise before the players get around to providing for them in a rule, and to prevent any single rule change from disturbing the continuity of the game. Whether the Initial Set presented below satisfies these competing interests is left to players to judge.

One contingency deliberately left to the players to resolve is what to do about violations of the rules. The players must also decide whether old violations are protected by a statute of limitations or whether they may still be punished or nullified. Whether the likelihood of compliance and the discretionary power of the judge suffice to deal with a crisis of confidence or to delay it until a rule can take over, and whether in other respects the Initial Set satisfactorily balances the competing interests of simplicity and complexity, can best be determined by playing the game.

* * *

Nomic affords a curious twist on one common and fundamental property of games: it allows the blurring of the distinction between *constitutive rules* and *rules of skill*—that is, between rules that define lawful play and those that define artful play. In other words, in Nomic there is a blurring between the permissible and the optimal.

Most games do not embrace non-play, and do not become paradoxical by seeming to. Interestingly, however, children often invent games that provide game penalties for declining to play, or that incorporate or extend game jurisdiction to all of "real life", and end only when the children tire of the game or forget they are playing. ("Daddy, Daddy, come play a new game we invented!" "No, sweetheart, I'm reading." "That's ten points!") Nomic carries this principle to an extreme. A game of Nomic can embrace anything at the vote of the players. The line between play and non-play may shift at each turn, or it may apparently be eliminated. Players may be governed by the game when they think they are between games or when they think they have quit.

For most games, there is an infallible decision procedure to determine the legality of a move. In Nomic, by contrast, situations may easily arise where it is very hard to determine whether or not a move is legal. Moreover, paradoxes can arise in Nomic that paralyze judgment. Occasionally this will be due to the poor drafting of a rule, but it may also arise from a rule that is unambiguous but mischievous. The variety of such paradoxes is truly impossible to anticipate. Rule 213, nonetheless, is designed to cope with them as well as possible without cluttering the Initial Set with too many legalistic qualifications. Note that Rule 213 allows a wily player to create a paradox, get it passed (if the rule seems innocent enough to the other players), and thereby win.

So much for a general prologue to the game itself. Now we can move on to a description of how a game of Nomic is played. To reiterate, Nomic is a game in which changing the rules is a move. Two can play, but having three or more makes for a better game. The gist of Nomic is to be found in Rule 202, which should be read first. Players will need paper and pencil, and (at least at the outset!) one die. Instead of sheets of paper, players may find it easier to use a set of index cards. All new rules and amendments are to be written down. How the rules are positioned on paper or on the table can indicate which ones are currently immutable and which ones are mutable. Amendments can be placed on top of or next to the rules they amend. Inoperative rules may simply be deleted. Alternatively, for more complex games, players may prefer to transcribe into their own notebooks the text of each new rule or amendment and to keep a separate list, by number, of the rules still in

effect. Ideally, perhaps, all rules should be entered in a computer, with a terminal for each player; amendments could then be incorporated instantly into the main text, with a corresponding adjustment to the numerical order.

Initial Set of Rules of Nomic

I. *Immutable Rules*

101. All players must always abide by all the rules then in effect, in the form in which they are then in effect. The rules in the Initial Set are in effect whenever a game begins. The Initial Set consists of Rules 101–116 (immutable) and 201–213 (mutable).
102. Initially, rules in the 100's are immutable and rules in the 200's are mutable. Rules subsequently enacted or transmuted (*i.e.,* changed from immutable to mutable or vice versa) may be immutable or mutable regardless of their numbers, and rules in the Initial Set may be transmuted regardless of their numbers.
103. A rule change is any of the following: (1) the enactment, repeal, or amendment of a mutable rule; (2) the enactment, repeal, or amendment of an amendment, or (3) the transmutation of an immutable rule into a mutable rule, or vice versa. (Note: This definition implies that, at least initially, all new rules are mutable. Immutable rules, as long as they are immutable, may not be amended or repealed; mutable rules, as long as they are mutable, may be amended or repealed. No rule is absolutely immune to change.)
104. All rule changes proposed in the proper way shall be voted on. They will be adopted if and only if they receive the required number of votes.
105. Every player is an eligible voter. Every eligible voter must participate in every vote on rule changes.
106. Any proposed rule change must be written down before it is voted on. If adopted, it must guide play in the form in which it was voted on.
107. No rule change may take effect earlier than the moment of the completion of the vote that adopted it, even if its wording explicitly states otherwise. No rule change may have retroactive application.
108. Each proposed rule change shall be given a rank-order number (ordinal number) for reference. The numbers shall begin with 301, and each rule change proposed in the proper way shall receive the next successive integer, whether or not the proposal is adopted.

    If a rule is repealed and then re-enacted, it receives the ordinal number of the proposal to re-enact it. If a rule is amended or transmuted, it receives the ordinal number of the proposal to amend or transmute it. If an amendment is amended or repealed, the entire rule of which it is a part receives the ordinal number of the proposal to amend or repeal the amendment.
109. Rule changes that transmute immutable rules into mutable rules may be adopted if and only if the vote is unanimous among the eligible voters.
110. Mutable rules that are inconsistent in any way with some immutable rule (except by proposing to transmute it) are wholly void and without effect. They do not implicitly transmute immutable rules into mutable rules and at the same time amend them. Rule changes that transmute immutable rules into mutable rules will be effective if and only if they explicitly state their transmuting effect.
111. If a rule change as proposed is unclear, ambiguous, paradoxical, or destructive of play, or if it arguably consists of two or more rule changes compounded or is an amendment that makes no difference, or if it is otherwise of questionable value, then the other players may suggest amendments or argue against the proposal before the vote. A reasonable amount of time must be allowed for this debate. The proponent decides the final form in which the proposal is to be voted on and decides the time to end debate and vote. The only cure for a bad proposal is prevention: a negative vote.
112. The state of affairs that constitutes winning may not be changed from achieving n points to any other state of affairs. However, the magnitude of n and the means of earning points may be changed, and rules that establish a winner when play cannot continue may be enacted and (while they are mutable) be amended or repealed.
113. A player always has the option to forfeit the game rather than continue to play or incur a game penalty. No penalty worse than losing, in the judgment of the player to incur it, may be imposed.
114. There must always be at least one mutable rule. The adoption of rule changes must never become completely impermissible.
115. Rule changes that affect rules needed to allow or apply rule changes are as permissible as other rule changes. Even rule changes that amend or repeal their own authority are permissible. No rule change or type of move is impermissible solely on account of the self-reference or self-application of a rule.
116. Whatever is not explicitly prohibited or regulated by a rule is permitted and unregulated, with the sole exception of changing the rules, which is permitted only when a rule or set of rules explicitly or implicitly permits it.

II. *Mutable Rules*

201. Players shall alternate in clockwise order, taking one whole turn apiece. Turns may not be skipped or passed, and parts of turns may not be omitted. All players begin with zero points.
202. One turn consists of two parts, in this order: (1) proposing one rule change and having it voted on, and (2) throwing

this notion was the way computers had in some sense been applied to themselves—namely in compilers, programs that translate programs from an elegant and human-readable language into the cryptic strings of 0's and I's of machine language.

The other notion Forsythe emphasized—and it was closely related to the first one—was the fact that a program is just an object that sits in a computer's memory, and as such is no more and no less subject to manipulation by other programs—or even by itself!—than mere numbers are. The fusion of these two notions was what gave me my inspiration to design an abstract computer. Playing on the names of the ENIAC, ILLIAC, JOHNNIAC, and other computers I had heard of, I called it "IACIAC". I hoped IACIAC could not only manipulate its own programs but also redesign itself, change the way it interpreted its own instructions, and so on. I quickly ran into many conceptual difficulties and never completed the project, but I have never forgotten that fascination. It seems to me that although it is a game and not a computer, Nomic comes closer in spirit to that goal I sought than anything I have ever encountered. That is, except for itself.

Post Scriptum.

As a result of the publication of this column, I received a letter from a law professor named William Popkin, who obviously had found the game of Nomic fascinating while disagreeing philosophically with some points expressed. Subsequently, an exchange between Popkin and me was printed in the "Letters" column in *Scientific American.* Here is what Popkin had to say:

> As a law professor I was very interested in Douglas Hofstadter's piece on reflexivity and self-reference in the law. There are, as he says, many examples. Article V of the United States Constitution prohibits amendments denying states equal representation in the Senate. The Supreme Court of India went out of its way to create a reflexivity problem by deciding that the normal process of amending the Indian Constitution did not apply to their Bill of Rights, even though no explicit provision prohibiting such amendments existed.
>
> These reflexivity problems are fascinating, but I do not see what they have to do with "general procedures of argument", as Hofstadter (quoting Howard DeLong) suggests. They have everything to do with the meaning of rules, law, and politics, but not with procedures of argument. Let me explain how at least one law professor would approach these problems. Every reflexivity example has the same structure. There is a rule that has specific cases coming under the rule. One particular case, by coming under the rule, appears to undermine the rule itself. For example, assume that the Supreme Court must decide cases properly appealed to it, but that no judge can sit on a case in which he is personally interested. A case arises involving the reduction of judges' salaries, which is arguably unconstitutional. If the judges decide the case, they violate the rule against deciding cases in which they are personally interested, but failure to decide violates the rule requiring them to decide cases. The same structure exists for rules about amendment of the document containing the amending provision. Assume that the Constitution can be amended by a two-thirds vote but that one of the provisions requires a 100 percent vote. An amendment is passed changing the unanimity rule. If the amendment is valid, the unanimity rule is undermined, but if the amendment is invalid, the procedures for amendment are incomplete.
>
> What is presented in all these cases is a problem of meaning and a conflict between rival conclusions, not a logical conundrum. The ultimate decision may be hard or easy, but the issues are not difficult to conceptualize. My own conclusion is that the Supreme Court should hear the case involving its own salary because we do not want Congress deciding such issues, and that the amending power should not extend to the unanimity rule because this breaks the social contract. These are hard cases, but another example presented in Hofstadter's article is easy. It concerns a contract to pay the rhetoric teacher Protagoras when his pupil Euathlus wins his first case. The teacher sues the pupil for the payment, figuring that if he wins the suit he gets his money and if he loses the suit he collects under the contract. But on what possible ground could he win the case before the pupil had won a lawsuit? And how could the original contract, in referring to a victory by the pupil as the occasion for the payment, include a victory in a frivolous lawsuit by the teacher?
>
> What I am pointing out is that reflexivity presents problems of choice, sometimes difficult, sometimes trivial, but that is nothing new in the law. Most important legal problems involve choice without involving reflexivity. Do we prefer a right of privacy or freedom of the press? The deeper point concerns the interaction of law and artificial intelligence and perhaps interdisciplinary studies generally. Reflexivity is undoubtedly an important phenomenon in philosophy for reasons I do not fully appreciate. If developments in artificial intelligence are to be useful in law, however, they must take into account what legal problems are all about. To a lawyer, reflexivity is not a relevant category but choice is. Indeed, I suspect that reflexivity is just a diversion for Hofstadter. In an earlier article about analogy he dealt with the imaginative problem of defining the First Lady of Britain [see Chapter 24]. He there grappled with the problem of deciding what is like something else, which is the way most lawyers always proceed in making choices. How we make

analogies determines how we make choices, and that is the essential nature of all judgment. If that is what artificial intelligence is all about, I very much want to hear more.

As for the question of whether there are immutable rules, the answer is: Of course there are, if that's what you want.

William D. Popkin
Professor of Law
Indiana University

I found this letter very nicely put, and a constructive opening for a small debate. I replied as follows:

Professor Popkin raises a very interesting point in his comment on my column about Peter Suber's game Nomic. His point is essentially twofold: (1) The fact that any legal system is inevitably chock-full of tangles arising from reflexivity is amusing, but rather than being themselves a deep aspect of law, such tangles are a consequence of other deep aspects, the most significant of which is that (2) the crux of any legal system is the ability of people to distinguish between the incidental qualities and the essential qualities of various events and relations, which ability results finally in recognition of what a given item is—that is, which category the item belongs to. Popkin calls this "choice". In conclusion, he suggests that to discover the principles by which people can "choose" is a critical task for artificial-intelligence workers to tackle.

I feel that neither Suber's *reflexivity* nor Popkin's *choice* is more central than the other in defining the nature of law. In fact, they are intertwined. Suber stresses that people, in choosing which of two inconsistent aspects of a supposedly self-consistent system shall take precedence, often make their choice without explicit rules (since if the rules were spelled out, they would be susceptible to getting embroiled in a similar tangle once again, only at a higher level of abstraction). "Law can disregard logical difficulties and ground a solution on pragmatic rules, social policies, and legal doctrines", Suber has written [in a reply to Popkin]. "The effectiveness of policy, or what Popkin calls 'choice', in plowing under logical obstacles is not the answer to the question but the mystery to be explained."

Coming to grips with this contrast between explicit rules and implicit principles or guidelines is of great importance if one wants to characterize how flexible category recognition—"choice"—takes place, whether one is doing research in artificial intelligence, philosophizing about free will, or attempting to characterize the nature of law. Popkin, in fact, is rather charitable toward artificial-intelligence research, suggesting that it may some day yield clues, if not the key, to the mystery of choice. I think he is right about this. He may have failed to realize, however, that in any attempt to make a machine capable of choice, one runs headlong into the problem of inconsistencies, level-collisions, and reflexivity tangles, and for the following reason.

All recognition programs are invariably modeled on what we know about perception in various modalities, such as hearing and sight. One thing we know for sure is that in any modality, perception consists of many layers of processing, from the most primitive or "syntactic" levels, to the most abstract or "semantic" levels. The zeroing-in on the semantic category to which a given raw stimulus belongs is carried out not by a purely bottom-up (stimulus-driven) or purely top-down (category-driven) scheme, but rather by a mixture of them, in which hypotheses at various levels trigger the creation of new hypotheses or undermine the existence of already-existing hypotheses at other levels. This process of sprouting and pruning hypotheses is a highly parallel one, in which all the levels compete simultaneously for attention, like billboards or radio commercials or advertisements in the subway.

Yet out of this seemingly anarchic chaos comes an integrated decision, in which the various levels gradually come to some kind of self-reinforcing agreement. If a firm decision is to emerge from such a swirl of conflicting claims, there must be some kind of mental *scheduler,* something that functions like *Robert's Rules of Order*, letting various levels have the floor, scheduling collective actions such as votes, overriding or tabling motions, and so on. In fact, to the best of our knowledge, this is the heart of the perceptual process. But this is the very place where reflexivity tangles crop up with a vengeance!

Any perception program has various levels of "inner sanctum"—that is, levels of untouchability of its data structures. (These structures include not only the current hypotheses, but also deeper, more permanent aspects of the program itself, such as the ways it weights various pieces of evidence, the rules by which it sorts out conflicts, the priority rules of its scheduler, and—of course—the information about the untouchability of levels!) Now, for the ultimate in flexibility, none of these levels should be *totally* untouchable (although that degree of flexibility may be unattainable), but obviously some levels should be less touchable than others. Therefore any recognition program must have at its core a tiered structure precisely like that of government (or that of the rules of Nomic), in which there are levels that are "easily mutable", "moderately mutable", "almost mutable", and so on. The structure of a recognition program—a "choice" program—is seen inevitably to be riddled with reflexivity.

The point of all this is that the very reflexivity issues that Popkin considers to be merely amusing sideshows in law are actually deeply embroiled in what he sees as the meat of the matter, namely the question of how category recognition—discerning the essence of something—works. For that reason, I found Suber's game not merely amusing but philosophically provocative as well. In fact, I consider the intertwined study of reflexivity and recognition, using the fresh methods of the emerging discipline of cognitive science, to be of great interest and importance for the light it may shed on the ancient philosophical problems of mind, free will, and identity—not to mention those of the philosophy of law.

* * *

It occurs to me that the message of my letter to Popkin could be put in a nutshell this way: To get *flexible cognition,* concentrate on *reflexivity* and *recognition.* Some of these ideas will come up again, more specifically in the context of artificial intelligence, in Chapters 23 and 24.

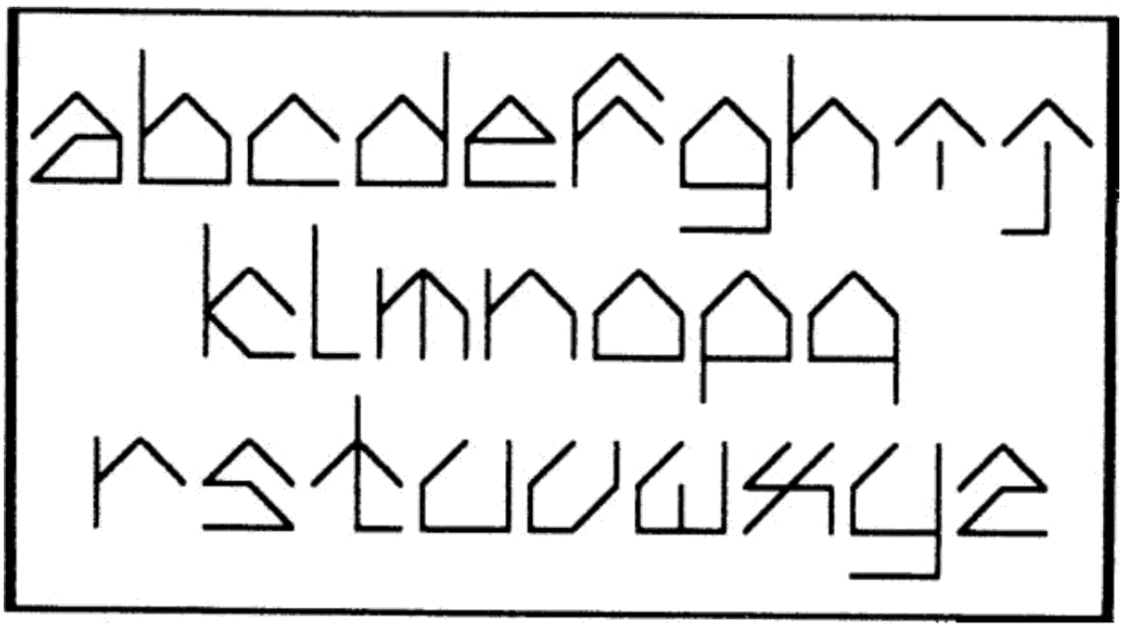
Section II:

Sense and Society

Section II:

Sense and Society

Another broad theme of this book is introduced in the four chapters comprising this section: the harm that occurs when vast numbers of people accept without reflection the words, sayings, ideas, fads, styles, and tastes paraded in front of them by indiscriminate media and popular myth. Our society does a rather poor job of making us aware of, let alone interested in, the nature of common sense, the hidden assumptions that permeate thought, the complex mechanisms of sensory perception and category systems, the will to believe, the human tendency toward gullibility, the most typical flaws in arguments, the statistical inferences we make unconsciously, the vastly different temporal and spatial scales on which one can look at the universe, the many filters through which one can perceive and conceptualize people and events, and so on. The resulting deceptions, delusions, confusions, ignorances, and fears can lead to many disquieting social consequences, such as mildly or absurdly wasteful spending of funds, blatant or subtle discrimination against groups, and local or global apathy about the current state and momentum of the world. Of course everyone labors under some delusions, avoids certain kinds of thoughts, has an overly closed mind on this or that subject. What, however, are the consequences when this is multiplied by hundreds or thousands of millions, and all the small pieces are woven together into a vast fabric? What does a carpet woven from the incomplete understandings and ignorances of five billion sentient beings look like from afar—and where is this flying carpet headed?



American—its lack of photographs of celebrities, for example—what convinces you it is to be trusted? If so, that is a pretty curious way of making decisions about what truth is. It would seem that your concept of truth is closely tied in with your way of evaluating the "style" of a channel of communication—surely quite an intangible notion!

Having said that, I must admit that I, too, rely constantly on quick assessments of style in my attempt to sift the true from the false, the believable from the unbelievable. (Quickness is of the essence, like it or not, because the world does not allow infinite time for deliberation.) I could not tell you what criteria I rely on without first pondering for a long time and writing many pages. Even then, were I to write the definitive guide (*How to Tell the True from the False by Its Style of Publication*), it would have to be published to do any good; and its title, not to mention the style it was published in, would probably attract a few readers, but would undoubtedly repel many more. There is something disturbing about that thought.

There is something else disturbing here. Enormous numbers of people are taken in, or at least beguiled and fascinated, by what seems to me to be unbelievable hokum, and relatively few are concerned with or thrilled by the astounding—yet true—facts of science, as put forth in the pages of, say, *Scientific American.* I would proclaim with great confidence that the vast majority of what that magazine prints is true—yet my ability to defend such a claim is weaker than I would like. And most likely the readers, authors, and editors of that magazine would be equally hard pressed to come up with cogent, nontechnical arguments convincing a skeptic of this point, especially if pitted against a clever lawyer arguing the contrary. How come Truth is such a slippery beast?

* * *

Well, consider the very roots of our ability to discern truth. Above all (or perhaps I should say "underneath all"), *common sense* is what we depend on—that crazily elusive, ubiquitous faculty we all have, to some degree or other. But not to a degree such as "Bachelor's" or "Ph.D.". No, unfortunately, universities do not offer degrees in Common Sense. There are not even any Departments of Common Sense! This is, in a way, a pity.

At first, the notion of a Department of Common Sense sounds ludicrous. Given that common sense *is* common, why have a department devoted to it? My answer would be quite simple: In our lives we are continually encountering strange new situations in which we have to figure out how to apply what we already know. It is not enough to have common sense about known situations; we need also to develop the art of extending common sense to apply to situations that are unfamiliar and beyond our previous experience. This can be very tricky, and often what is called for is common sense in knowing *how* to apply common sense: a sort of "meta-level" common sense. And this kind of higher-level common sense also requires its own meta-level common sense. Common sense, once it starts to roll, gathers more common sense, like a rolling snowball gathering ever more snow. Or, to switch metaphors, if we apply common sense to itself over and over again, we wind up building a skyscraper. The ground floor of this structure is the ordinary common sense we all have, and the rules for building new floors are implicit in the ground floor itself. However, working it all out is a gigantic task, and the result is a structure that transcends mere common sense.

Pretty soon, even though it has all been built up from common ingredients, the structure of this extended common sense is quite arcane and elusive. We might call the quality represented by the upper floors of this skyscraper "rare sense"; but it is usually called "science". And some of the ideas and discoveries that have come out of this originally simple and everyday ability defy the ground floor totally. The ideas of relativity and quantum mechanics are anything but commonsensical, in the ground-floor sense of the term! They are outcomes of common sense self-applied, a process that has many unexpected twists and gives rise to some unexpected paradoxes. In short, it sometimes seems that common sense, recursively self-applied, almost undermines itself.

Well, truth being this elusive, no wonder people are continually besieged with competing voices in print. When I was younger, I used to believe that once something had been discovered, verified, and published, it was then part of Knowledge: definitive, accepted, and irrevocable. Only in unusual cases, so I thought, would opposing claims then continue to be published. To my surprise, however, I found that the truth has to fight constantly for its life! That an idea has been discovered and printed in a "reputable journal" does *not* ensure that it will become well known and accepted. In fact, usually it will have to be rephrased and reprinted many different times, often by many different people, before it has any chance of taking hold. This is upsetting to an idealist like me, someone more disposed to believe in the notion of a monolithic and absolute truth than in the notion of a pluralistic and relative truth (a notion championed by a certain school of anthropologists and sociologists, who un-self-consciously insist "all systems of belief are equally valid", seemingly without realizing that this dogma of relativism not only is just as narrow-minded as any other dogma, but moreover is unbelievably wishy-washy!). The idea that the truth has to fight for its life is a sad discovery. The idea that the truth will *not* out, unless it is given a lot of help, is pretty upsetting.

* * *

A question arises in every society: Is it better to let all the different voices battle it out, or to have just a few "official" publications dictate what is the case and what is not? Our society has opted for a plurality of voices, for a "marketplace of ideas", for a complete free-for-all of conflicting theories. But if things are this chaotic, who will ensure that there is law and order? Who will guard the truth? The answer (at least in part) is: CSICOP will!

CSICOP? Who is CSICOP? Some kind of cop who guards the truth? Well, that's pretty close. "CSICOP" stands for "Committee for the Scientific Investigation of Claims of the Paranormal"—a rather esoteric title for an organization whose purpose is not so esoteric: to apply common sense to claims of the outlandish, the implausible, and the unlikely.

Who are the people who form CSICOP and what do they do together? The organization was the brainchild of Paul Kurtz, professor of philosophy at the State University of New York at Buffalo, who brought it into being because he thought there was a need to counter the rising tide of irrational beliefs and to provide the public with a more balanced treatment of claims of the paranormal by presenting the dissenting scientific viewpoint. Among the early members of CSICOP were some of America's most distinguished philosophers (Ernest Nagel and Willard Van Orman Quine, for example) and other colorful combatants of the occult, such as psychologist Ray Hyman, magician James Randi, and someone whom readers of this column may have heard of: Martin Gardner. In the first few meetings, it was decided that the committee's principal function would be to publish a magazine dedicated to the subtle art of debunking. Perhaps "debunking" is not the term they would have chosen, but it fits. The magazine they began to publish in the fall of 1976 was called *The Zetetic,* from the Greek for "inquiring skeptic".

As happens with many fledgling movements, a philosophical squabble developed between two factions, one more "relativist" and unjudgmental, the other more firmly opposed to nonsense, more willing to go on the offensive and to attack supernatural claims. Strange to say, the open-minded faction was not so open-minded as to accept the opposing point of view, and consequently the rift opened wider. Eventually there was a schism. The relativist faction (one member) went off and started publishing his own journal, the *Zetetic Scholar,* in which science and pseudo-science coexist happily, while the larger faction retained the name "CSICOP" and changed the title of its journal to the *Skeptical Inquirer.*

In a word, the purpose of the *Skeptical Inquirer* is to combat nonsense. It does so by recourse to common sense, and as much as possible by recourse to the *ground floor* of the skyscraper of science —the common type of common sense. This is by no means always possible, but it is the general

style of the magazine. This means it is accessible to anyone who can read English. It does not require any special knowledge or training to read its pages, where nonsensical claims are routinely smashed to smithereens. (Sometimes the claims are as blatantly silly as the headlines at the beginning of this article, sometimes much subtler.) All that is required to read this maverick journal is curiosity about the nature of truth: curiosity about how truth defends itself (through its agent CSICOP) against attacks from all quarters by unimaginably imaginative theorizers, speculators, eccentrics, crackpots, and out-and-out fakers.

The journal has grown from its original small number of subscribers to roughly 7,500—a David, compared with the Goliaths mentioned above, with their circulations in the millions. Its pages are filled with lively and humorous writing—the combat of ideas in its most enjoyable form. By no means is this journal a monolithic voice, a mouthpiece of a single dogma. Rather, it is itself a marketplace of ideas, strangely enough. Even people who wield the tool of common sense with skill may do so with different styles, and sometimes they will disagree.

There is something of a paradox involved in the editorial decisions in such a magazine. After all, what is under debate here is, in essence, the nature of correct arguments. What should be accepted and what shouldn't? To caricature the situation, imagine the editorial dilemmas that would crop up for journals with titles such as *Free Press Bulletin, The Open Mind,* or *Editorial Policy Newsletter.* What letters to the editor should be printed? What articles? What policy can be invoked to screen submitted material?

These are not easy questions to answer. They involve a paradox, a tangle in which the ideas being evaluated are also what the evaluations are based on. There is no easy answer here! There is no recourse but to common sense, that rock-bottom basis of all rationality. And unfortunately, we have no foolproof algorithm to uniquely characterize that deepest layer of rationality, nor are we likely to come up with one soon. The ability to use common sense—no matter how much light is shed on it by psychologists or philosophers—will probably forever remain a subjective art more than an objective science. Even when experimental epistemologists, in their centuries-long quest for artificial intelligence, have at last made a machine that thinks, its common sense will probably be just as instinctive and fallible and stubborn as ours. Thus at its core, rationality will always depend on inscrutables: the simple, the elegant, the intuitive. This weird paradox has existed throughout intellectual history, but in our information-rich times it seems particularly troublesome.

Despite these epistemological puzzles, which seem to be intimately connected with its very reason for existence, the *Skeptical Inquirer* is flourishing and provides a refreshing antidote to the jargon-laden journals of science, which often seem curiously irrelevant to the concerns of everyday life. In that one way, the *Inquirer* resembles the scandalous tabloids.

The list of topics covered in the seventeen issues that have appeared so far is remarkably diverse. Some topics have arisen only once, others have come up regularly and been discussed from various angles and at various depths. Some of the more commonly discussed topics are:

> ESP (extra-sensory perception) * telekinesis (using mental power to influence events at a distance) * astrology * biorhythms * Bigfoot * the Loch Ness monster * UFO's (unidentified flying objects) * creationism * telepathy * remote viewing * clairvoyant detectives who allegedly solve crimes * the Bermuda (and other) triangles * "thoughtography" (using mental power to create images on film) * the supposed extraterrestrial origin of life on the earth * Carlos Castaneda's mystical sorcerer "Don Juan" * pyramid power * psychic surgery and faith healing * Scientology * predictions by famous "psychics" * spooks and spirits and haunted houses * levitation * palmistry and mind reading * unorthodox anthropological theories * plant perception * perpetual-motion machines * water witching and other kinds of dowsing * bizarre cattle mutilations

When I contemplate the length of this list, I am quite astonished. Before I ever subscribed to the magazine, I had heard of almost all these items and was skeptical of most of them, but I had never seen a frontal assault mounted against so many paranormal claims at once. And I have only scratched the surface of the list of topics, because the ones listed above are regulars! Imagine how

many topics are treated at shorter length.

There are quite a few frequent contributors to this iconoclastic journal, such as James Randi, who is truly prolific. Among others are aeronautics writer Philip J. Klass, UFO specialist James E. Oberg, writer Isaac Asimov, CSICOP's founder (and current director) Paul Kurtz, psychologist James Alcock, educator Elmer Kral, anthropologist Laurie Godfrey, science writer Robert Sheaffer, sociologist William Sims Bainbridge, and many others. And the magazine's editor, Kendrick Frazier, a free-lance science writer by trade, periodically issues eloquent and mordant commentaries.

* * *

I know of no better way to impart the flavor of the magazine than to quote a few selections from articles. One of my favorite articles appeared in the second issue (Spring/Summer, 1977). It is by psychologist Ray Hyman (who, incidentally, like many other authors in the *Skeptical Inquirer,* is a talented magician) and is titled "Cold Reading: How to Convince Strangers that You Know All About Them".

It begins with a discussion of a course Hyman taught about the various ways people are manipulated. Hyman states:

> I invited various manipulators to demonstrate their techniques—pitchmen, encyclopedia salesmen, hypnotists, advertising experts, evangelists, confidence men and a variety of individuals who dealt with personal problems. The techniques which we discussed, especially those concerned with helping people with their personal problems, seem to involve the client's tendency to find more meaning in any situation than is actually there. Students readily accepted this explanation when it was pointed out to them. But I did not feel that they fully realized just how pervasive and powerful this human tendency to make sense out of nonsense really is.

Then Hyman describes people's willingness to believe what others tell them about themselves. His "golden rule" is: "To be popular with your fellow man, tell him what he wants to hear. He wants to hear about himself. So tell him about himself. But not what you know to be true about him. Oh, no! Never tell him the truth. Rather, tell him what he would like to be true about himself!" As an example, Hyman cites the following passage (which, by an extraordinary coincidence, was written about none other than *you,* dear reader!):

> Some of your aspirations tend to be pretty unrealistic. At times you are extroverted, affable, sociable, while at other times you are introverted, weary, and reserved. You have found it unwise to be too frank in revealing yourself to others. You pride yourself on being an independent thinker and do not accept others' opinions without satisfactory proof. You prefer a certain amount of change and variety, and become dissatisfied when hemmed in by restrictions and limitations. At times you have serious doubts as to whether you have made the right decision or done the right thing. Disciplined and controlled on the outside, you tend to be worrisome and insecure on the inside.
>
> Your sexual adjustment has presented some problems for you. While you have some personality weaknesses, you are generally able to compensate for them. You have a great deal of unused capacity which you have not turned to your advantage. You have a tendency to be critical of yourself. You have a strong need for other people to like you and for them to admire you.

Pretty good fit, eh? Hyman comments:

> The statements in this stock spiel were first used in 1948 by Bertram Forer in a classroom demonstration of personal validation. He obtained most of them from a newsstand astrology book. Forer's students, who thought the sketch was uniquely intended for them as a result of a personality test, gave the sketch an average rating of 4.26 on a scale of 0 (poor) to 5 (perfect). As many as 16 out of his 39 students (41 percent) rated it as a perfect fit to their personality. Only five gave it a rating below 4 (the worst being a rating of 2, meaning "average"). Almost 30 years later students give the same sketch an almost identical rating as a unique description of themselves.

A particularly delicious feature is the thirteen-point recipe that Hyman gives for becoming a cold reader. Among his tips are these: "Use the technique of 'fishing' (getting the subject to tell you about himself or herself, then rephrasing it and feeding it back); always give the impression that you know more than you are saying; don't be afraid to flatter your subject every chance you get." This cynical recipe for becoming a character reader is presented by Hyman in considerable detail,

presumably not to convert readers of the article into charlatans and fakers, but to show them the attitude of the tricksters who do such manipulations. Hyman asks:

> Why does it work so well? It does not help to say that people are gullible or suggestible. Nor can we dismiss it by implying that some individuals are just not sufficiently discriminating or lack sufficient intelligence to see through it. Indeed, one can argue that it requires a certain degree of intelligence on the part of a client for the reading to work well . . . . We have to bring our knowledge and expectations to bear in order to comprehend anything in our world. In most ordinary situations, this use of context and memory enables us to correctly interpret statements and supply the necessary inferences to do this. But this powerful mechanism can go astray in situations where there is no actual message being conveyed. Instead of picking up random noise, we still manage to find meaning in the situation. So the same system that enables us to creatively find meanings and to make new discoveries also makes us extremely vulnerable to exploitation by all sorts of manipulators. In the case of the cold reading, the manipulator may be conscious of his deception; but often he too is a victim of personal validation.

Hyman knows what he's talking about. Many years ago, he was convinced for a time that he himself had genuine powers to read palms, until one day when he tried telling people the exact opposite of what their palms told him and saw that they still swallowed his line as much as ever! Then he began to suspect that the plasticity of the human mind—his own particularly—was doing some strange things.

* * *

At the beginning of each issue of the *Skeptical Inquirer* is a feature called "News and Comment". It covers such things as the latest reports on current sensational claims, recently broadcast television shows for and against the paranormal, lawsuits of one sort or another, and so on. One of the most amusing items was the coverage in the Fall 1980 issue of the "Uri Awards", given out by James Randi (on April 1, of course) to various deserving souls who had done the most to promote gullibility and irrational beliefs. Each award consists of "a tastefully bent stainless-steel spoon with a very transparent, very flimsy base". Award winners were notified, Randi explained, by telepathy, and were "free to announce their winning in advance, by precognition, if they so desired". Awards were made in four categories: Academic ("to the scientist who says the dumbest thing about parapsychology"), Funding ("to the funding organization that awards the most money for the dumbest things in parapsychology"), Performance ("to the psychic who, with the least talent, takes in the most people"), and Media ("to the news organization that supports the most outrageous claims of the paranormalists").

The nature of coincidences is a recurrent theme in discussions of the paranormal. I vividly remember a passage in a lovely book by Warren Weaver titled *Lady Luck: The Theory of Probability,* in which he points out that in many situations, the most likely outcome may well be a very unlikely event (as when you deal hands in bridge, where whatever hand you get is bound to be extraordinarily rare). A similar point is made in the following excerpt from a recent book by David Marks and Richard Kammann titled *The Psychology of the Psychic* (from which various excerpts were reprinted in one issue of the *Skeptical Inquirer):*

> 'Koestler's fallacy' refers to our general inability to see that unusual events are probable in the long run . . . . It is a simple deduction from probability theory that an event that is very improbable in a *short run* of observations becomes, nevertheless, highly probable somewhere in a *long run* of observations . . . . We call it 'Koestler's fallacy' because Arthur Koestler is the author who best illustrates it and has tried to make it into a scientific revolution. Of course, the fallacy is not unique to Koestler but is widespread in the population, because there are several biases in human perception and judgment that contribute to this fallacy.
>
> First, we notice and remember matches, especially *oddmatches,* whenever they occur. (Because a psychic anecdote first requires a match, and, second, an oddity between the match and our beliefs, we call these stories *oddmatches.* This is equivalent to the common expression, an "unexplained coincidence".) Second, we do *not* notice non-matches. Third, our failure to notice nonevents creates the *short-run illusion* that makes the oddmatch seem improbable. Fourth, we are poor at estimating combinations of events. Fifth, we overlook the *principle of equivalent oddmatches,* that one coincidence is as good as another as far as psychic theory is concerned.

department) carried out a similar experiment on a first-year psychology class at Southern Illinois University, which he wrote up in the Spring 1980 issue of the *Skeptical Inquirer*. First, Morris assessed his students' beliefs in ESP by having them fill out a questionnaire. Then a colleague performed an "ESP demonstration", which Morris calls "frighteningly impressive".

After this powerful performance, Morris tried to "deprogram" his students. He had two weapons at his disposal. One is what he calls "dehoaxing". This process, just three minutes long, consisted in a revelation of how two of the three tricks worked, together with a confession that the remaining one of the baffling stunts was also a trick. "But," said Morris, "I'm not going to say how it was done, because I want you to experience the feeling that, even though you can't explain something, that doesn't make it supernatural." The other weapon was a 50-minute anti-ESP lecture, in which secrets of professional mind readers were revealed, commonsense estimates of probabilities of "oddmatches" were discussed, "scientific" studies of ESP were shown to be questionable for various statistical and logical reasons, and some other everyday reasons were adduced to cast ESP's reality into strong doubt.

After the performance, only half of the classes were "dehoaxed", but all of them heard the anti-ESP lecture. The students were then polled about the strength of their belief in various kinds of paranormal phenomena. It turned out that dehoaxed classes had a far lower belief in ESP than classes that had simply heard the anti-ESP lecture. The dehoaxed classes' average level of ESP belief dropped from nearly 6 (moderate belief) to about 2 (strong disbelief), while the non-dehoaxed classes' average level dropped from 6 to about 4 (slight disbelief). As Morris summarizes this surprising result, "The dehoaxing experience was apparently crucial; a three-minute revelation that they had been fooled was more powerful than an hour-long denunciation of ESP in producing skepticism toward ESP."

One of Morris' original interests in conducting this experiment was "whether the exercise would teach the students skepticism for ESP statements only, or a more general attitude of skepticism, as we had hoped. For example, would their experience also make them more skeptical of astrology, Ouija boards, and ghosts?" Morris did find a slight transfer of skepticism, and from it he concluded hopefully that "teaching someone to be skeptical of one belief makes him somewhat more skeptical of similar beliefs, and perhaps slightly more skeptical even of dissimilar beliefs."

This question of transfer of skepticism is, to my mind, the critical one. It is of little use to learn a lesson if it always remains a lesson about particulars and has no applicability beyond the case in which it was first learned. What, for instance, would you say is "the lesson of the People's Temple incident in Jonestown"? Simply that one should never follow the Reverend Jim Jones to Guyana? Or more generally, that one should be wary of following any guru halfway across the world? Or that one should never follow anyone anywhere? Or that all cults are evil? Or that any belief in any kind of savior, human or divine, is crazy and dangerous? Or consider the recent convulsions in Iran. Is it likely that the fundamentalist "Moral Majority" Christians in America would see their own attitudes as parallel to those of fundamentalist Moslems whose fanaticism they abhor, and that they would thereby be led to reflect on their own behavior? I wouldn't hold my breath. At what level of generality is a lesson learned? What was "the lesson of Viet Nam"? Does it apply to any present political situations that the United States is facing, or that any country is facing?

* * *

Stalker's Captain Ray of Light expresses faith that by debunking his own "miniature" pseudo-sciences before audiences, he can transfer to people a more general critical ability—an ability to think more clearly about paranormal claims. But how true is this? There are untold believers in some types of paranormal phenomena who will totally ridicule other types. It is quite common to encounter someone who will scoff at the headlines in the *National Enquirer* while at the same time believing, say, that through Transcendental Meditation you can learn to levitate, or that

astrological predictions come true, or that UFO's are visitors from other galaxies, or that ESP exists. I've heard many people express the following sort of opinion: "Most psychics, unfortunately, are frauds, which makes it all the more difficult for the *genuine* ones to be recognized." You even get believers in tricksters such as Uri Geller who say, "I admit he cheats *some* of the time, maybe even 90 percent of the time—but believe me, he has genuine psychic abilities!"

If you are hunting for a signal in a lot of noise, and the more you look, the more noise you find, when is it reasonable to give up and conclude there is no signal there at all? On the other hand, sometimes there just might be a signal! The problem is, you don't want to jump too quickly to a negative generalization, especially if your feelings are based merely on some kind of guilt by association. After all, not *everything* published in the *National Enquirer* is false. (I had to look awfully hard, though, to locate something in its pages that I was *sure* is true!) The subtle art is in sensing just when to shift—in sensing when there is enough evidence. But for better or for worse, this is a subjective matter, an art that few journals heretofore have dealt with.

The *Skeptical Inquirer* concerns itself with questions ranging from the ridiculous to the sublime, from the trivial to the profound. There are those who would say it is a big waste of time to worry about such drivel as ESP and other so-called paranormal effects, whereas others (such as myself) feel that anyone who is unable or unwilling to think hard about what distinguishes the scientific system of thinking from its many rival systems is not a devotee of truth at all, and furthermore that the spreading of nonsense is a dangerous trend that ought to be checked.

In any case, the question arises whether the *Skeptical Inquirer* will ever amount to more than a tiny drop in a huge bucket. Surely its editors do not expect that someday it will be sold alongside the *National Enquirer* at supermarket checkout counters! Or, carrying this vision to an upside-down extreme, can you imagine a world where a debunking journal such as the *Skeptical Inquirer* (in tabloid form, of course) sold millions of copies each week at supermarkets (along with its many rivals), while one lone courageous voice of the occult came out four times a year (in a relatively staid format) and was sought out by a mere 7,500 readers? Where the many rival debunking tabloids were always to be found lying around in laundromats? It sounds like a crazy story fit for the pages of the *National Enquirer!* This ludicrous scenario serves to emphasize just what the hardy band at CSICOP is up against.

What good does it do to publish their journal when only a handful of already-convinced anti-occult fanatics read it anyway? The answer is found in, among other places, the letters column at the back of each issue. Many people write in to say how vital the magazine has been to them, their friends, and their students. High-school teachers are among the most frequent writers of thank-you notes to the magazine's editors, but I have also seen enthusiastic letters from members of the clergy, radio talk-show hosts, and people in many other professions.

I would hope that by now I have aroused enough interest on the part of readers that they might like to subscribe to at least one of the journals that I have discussed in these pages. In the spirit of open-mindedness and relativism, therefore, I hereby provide addresses for all three (in alphabetical order):

National Enquirer
Lantana, Florida 33464

Skeptical Inquirer
Box 229, Central Park Station
Buffalo, New York 14215

Zetetic Scholar
Department of Sociology
Eastern Michigan University
Ypsilanti, Michigan 48197

Of course, I would not dream of suggesting which one to subscribe to. Perhaps the most prudent course would be not to make any prejudgments, and to subscribe to all three.

* * *

Certainly one will never be able to empty the vast ocean of irrationality that all of us are drowning in, but the ambition of the *Skeptical Inquirer* has never been that heroic; it has been, rather, to be a steady buoy to which one could cling in that tumultuous sea. It has been to promote a healthy brand of skepticism in as many people as it can. As Kendrick Frazier said in one of his eloquent editorials,

> Skepticism is not, despite much popular misconception, a point of view. It is, instead, an essential component of intellectual inquiry, a method of determining the facts whatever they may be and wherever they might lead. It is a part of what we call common sense. It is a part of the way science works. All who are interested in the search for knowledge and the advancement of understanding, imperfect as those enterprises may be, should, it seems to me, support critical inquiry, whatever the subject and whatever the outcome.

It is too bad that we should have to constantly defend truth against so many onslaughts from people unwilling to think, but, on the other hand, sloppy thought seems inevitable. It's just part of human nature. Come to think of it, didn't I read somewhere recently about how your average typical-type John or Jane Doe in the street uses only ten percent of his or her brains? Something like that! How come folks don't think harder and get *more* of those little brain cells going? Beats me! Talk about sloppy—it's downright boggling!! Even the scientists are stumped!!!

Post Scriptum

In the April 1982 issue of *Spektrum der Wissenschaft* (the German edition of *Scientific American*), the translation of this column appeared. On the flip side of the page with the headline "Boy can see with his ears" (*Junge kann mit den Ohren sehen*) I found a short article whose headline ran "Learning to hear with your eyes" (*Mit den Augen hören lernen*). It's logical, I guess—hearing with your eyes *does* seem to be the flip side of seeing with your ears! The article actually was about a machine for helping deaf people improve their speech with the aid of computer displays of their voices.

It was remarkable to see how similar these flipped headlines were, and yet how totally different the articles were. The main difference was actually in *tone.* The *National Enquirer* article spoke of an event that supposedly had occurred and characterized it as baffling and beyond explanation; the *Spektrum der Wissenschaft* article mentioned a counterintuitive idea and explained how it might conceivably be realized, after a fashion. Note that *Spektrum der Wissenschaft* managed to grab my attention by exploiting the same device as the tabloids do: catch readers by blaring something paradoxical. To someone not firmly grounded in science, "hearing with your eyes" and "seeing with your ears" sound (and look!) about equally implausible. Indeed, even to someone who is scientifically educated, the two phrases sound about equally weird. More information is needed to flesh out the meanings. That information was provided in *Spektrum der Wissenschaft,* and turned the initially grabbing headline into a sensible notion. Such is usually not the case for articles in the tabloids. But for most readers, such a subtle distinction doesn't matter.

This all goes to emphasize the claim at the beginning of this chapter about the trickiness of trying to pin down what truth is, and how deeply circular all belief systems are, no matter how much they try to be objective. In the end, rate of survival is the only difference between belief systems. This is a worrisome statement. It certainly worries me, at least. Still, I believe it. But scientists, I find, are not usually willing to see science itself as being rooted in an impenetrably murky swamp of beliefs and attitudes and perceptions. Most of them have never considered how it is that human perception and categorization underlie all that we take for granted in terms of

common sense, and in more primordial ways that are so deeply embedded that we even find them hard to talk about. Such things as: how we break the world into parts, how we form mental categories, how we refine them certain times while blurring them other times, how experiences and categories are clustered associatively, how analogies guide our intuitions, how imagery works, how valid logic is and where it comes from, how we tend to favor simple statements over complex ones, and so on—all these are, for most scientists, nearly un-grapplable-with issues, and so they pay them no heed and continue with their work.

The idea of "simplicity" is a real can of worms, for what is simple in one vocabulary can be enormously complex in another vocabulary—and vice versa. Does the sun rise in the mornings? Ninety-nine to one you use that geocentric phrase in your ordinary conversations, and geocentric imagery in your private thoughts. Yet we all "know" that the truth is different: the earth is *really* rotating on its axis and so the sun's motion is only *apparent.* Well, it may be news to you that general relativity says that all coordinate systems are equally valid—and that includes one from whose point of view all motion takes place with respect to a fixed, nonrotating earth. Thus Einstein tells us that Copernicus and Galileo were, after all, not any righter than Ptolemy and the Pope (score ten points for infallibility!). There is even, for each of us, a physically valid "egocentric" system of coordinates in which *I* am still and everything moves relative to me! I point this out to show that the truth is much shiftier and subtler than any simple picture can ever say. Scientists who oversimplify science distort reality as much as religious fanatics or pseudo-scientists do. The troubling truth is that there is no simple boundary line between nonsense and sense. (See Chapter 11). It is a lot hazier and blurrier and messier than even thoughtful people generally wish to admit.

When I was a columnist in *Scientific American,* I got quite a lot of mail, including a sizable number of letters from what I might charitably term "fringe thinkers", or uncharitably term "crackpots". I built up large files of such letters in the hopes of someday writing an article about "crackpotism" and its detection. The hypothetical book *How to Tell the True from the False by Its Style of Publication,* which I jokingly referred to in the article as something that I might write, was therefore not entirely a joke.

How can you discern which books you *do* want to read from those you don't? Answer: You have various levels of depth of evaluation, ranging from extremely brief and superficial tests to very deep and probing ones (*i.e.,* where you actually *do* take the trouble to read the book to see what it says). In order to reach the final stage (reading the book), you go through several very critical intermediate levels of analysis and scrutiny. I call this mechanism for filtering the "terraced scan".

How do I decide which letters to read carefully, if I don't read them all carefully (to decide whether or not to read them carefully . . .)? Answer: I apply the crudest, most "syntactic" stages of my terraced scanner and prune out the worst ones very quickly. Then I apply a slighty more refined stage of testing to the survivors, and prune out some more. And on it goes, until I am left with just a handful of truly provocative, significant letters. But if I had no such terraced-scan mechanism, I would be trapped in perpetual indecision, having no basis to decide to do anything, since I would need to evaluate *every pathway in depth* in order to decide whether or not to follow it. Should I take the bus to Kalamazoo today? Study out of a Smullyan book? Practice the piano? Read the latest *New York Review of Books*? Write an angry letter to someone in government?

This question of the interaction of *form* and *content* fascinates me deeply. I do indeed believe that if one has the right "terraced scan" mechanisms, one can go very far indeed in separating the wheat from the chaff. Of course, one has to believe that there *is* such a distinction: that The Truth actually exists. And just what this Truth is is very hard to say.

* * *

To me, part of the challenge of Zen is very much akin to the challenge of the occult and of pseudo-science: the baffling inner consistency of a worldview totally antithetical to my own. What

is also interesting is that each human being has a totally unique worldview, with its private contradictions and even small insanities. It is my belief, for instance, that inside every last one of us there is at least a small pocket of insanity: a kind of Achilles' heel that we try to avoid exposing to the world—and to ourselves. In his own personal way, Einstein was loony; in my own personal way, I am loony; and the same for you, dear lunatic!

In a way, therefore, to try to pursue the nature of ultimate truth is to enter a bottomless pit, filled with circular vipers of self-reference. One could liken CSICOP's job to that of the American Civil Liberties Union, which gets itself in all sorts of tangled loops because of its stance of defending radical belief systems. For instance, in an odd twist, its director, a former concentration camp inmate, found himself defending the rights of neo-Nazis to march down the streets of highly Jewish Skokie, Illinois, parading their banners advocating the extermination of all "inferior races". And what was worse for him was that as a consequence of his actions, the ACLU lost a significant portion of its membership. Patrick Henry spoke of "defending to the death your right to say it"—but does "it" include *anything*? Recipes for how to murder people? How to build atomic bombs? How to destroy the free press? Governments also face this sticky kind of issue. Can a government dedicated to liberty afford to let an organization dedicated to that government's downfall flourish?

It always seems refreshing to see how magazines, in their letters columns, willingly publish letters highly critical of them. I say "seems", because often those letters are printed in pairs, both raking the magazine over the coals but from opposite directions. For example, a right-wing critic and a left-wing critic both chastise the magazine for leaning too far the wrong way. The upshot is of course that the magazine doesn't even have to say a thing in its own defense, for it is a kind of cliché that if you manage to offend both parties in a disagreement, you certainly must be essentially right! That is, the truth is supposedly *always* in the middle—a dangerous fallacy.

Raymond Smullyan, in his book *This Book Needs No Title*, provides a perfect example of the kind of thing I am talking about. It is a story about two boys fighting over a piece of cake. Billy says he wants it all, Sammy says they should divide it equally. An adult comes along and asks what's wrong. The boys explain, and the adult says, "You should compromise—Billy gets three quarters, Sammy one quarter." This kind of story sounds ridiculous, yet it is repeated over and over in the world, with loudmouths and bullies pushing around meeker and fairer and kinder people. The "middle position" is calculated by averaging all claims together, outrageous ones as well as sensible ones, and the louder any claim, the more it will count. Politically savvy people learn this early and make it their credo; idealists learn it late and refuse to accept it. The idealists are like Sammy, and they always get the short end of the stick.

Magazines often gain rather than lose by printing what amounts to severe criticism. This holds even if the critical letter is not matched by an equally critical letter from the other side, because if a magazine prints letters critical of it, it appears open-minded and willing to listen to criticism. Thus the opposition is co-opted and undercut.

Another problem is that by shouting loud enough, advocates of any viewpoint can gain public attention. Sometimes the loudness comes from the large number of adherents of a particular point of view, sometimes it comes from the eloquence or charisma of a single individual, and sometimes it comes from the high status of one individual. A particularly salient example of this sort of thing is provided by the behavior of the Nixon "team" during the Watergate affair. There, they had the ability to manipulate the press and the public simply because they were in power. What no private individual would ever have been able to get away with for a second was done with the greatest of ease by the Nixon people. They shamelessly changed the rules as they wished, and for a long time got away with it.

What does all this have to do with the *Skeptical Inquirer*? Plenty. Amidst the tumult and the shouting, where does the truth lie? What voices should one listen to? How can one tell which are credible and which are not? It might seem that the serious matters of life have precious little to do with the validity of horoscopes, the probability of reincarnation, or the existence of Bigfoot, but I

Where does one draw the line? Where is the borderline between open-mindedness and stupidity? Or between closed-mindedness and stupidity? Where is the optimum balance? That is such a deep question that I could not hope to answer it. Professor Truzzi's position and my own lie at different points along a spectrum. We have both arrived at our positions not by pristine logic, but as a result of many complex interacting intuitions about the world and about minds and knowledge. There is certainly no way to *prove* that my position is righter than his, or vice versa. But even if we have no adequate theory to *formalize* such decisions, we nonetheless are all walking instantiations of such decision-making beings, and we make decisions for which we could not formally account in a million years. Such decisions include all decisions of taste, whether in food, music, art, or science. We have to live with the fact that we do not yet know *how* we make such decisions, but that does not mean we have to wallow in indecisiveness in the meantime. And anything that helps to make our quick decisions more informed while not impairing their quickness is of tremendous importance. I view the *Skeptical Inquirer* as serving that purpose, and I heartily recommend it to my readers.

6

On Number Numbness

May, 1982

THE renowned cosmogonist Professor Bignumska, lecturing on the future of the universe, had just stated that in about a billion years, according to her calculations, the earth would fall into the sun in a fiery death. In the back of the auditorium a tremulous voice piped up: "Excuse me, Professor, but h-h-how long did you say it would be?" Professor Bignumska calmly replied, "About a billion years." A sigh of relief was heard. "Whew! For a minute there, I thought you'd said a *million* years."

John F. Kennedy enjoyed relating the following anecdote about a famous French soldier, Marshal Lyautey. One day the marshal asked his gardener to plant a row of trees of a certain rare variety in his garden the next morning. The gardener said he would gladly do so, but he cautioned the marshal that trees of this size take a century to grow to full size. "In that case," replied Lyautey, "plant them this afternoon."

In both of these stories, a time in the distant future is related to a time closer at hand in a startling manner. In the second story, we think to ourselves: Over a century, what possible difference could a day make? And yet we are charmed by the marshal's sense of urgency. Every day counts, he seems to be saying, and particularly so when there are thousands and thousands of them. I have always loved this story, but the other one, when I first heard it a few thousand days ago, struck me as uproarious. The idea that one could take such large numbers so personally, that one could sense doomsday so much more clearly if it were a mere *million* years away rather than a far-off *billion* years—hilarious! Who could possibly have such a gut-level reaction to the difference between two huge numbers?

Recently, though, there have been some even funnier big-number "jokes" in newspaper headlines—jokes such as "Defense spending over the next four years will be $1 trillion" or "Defense Department overrun over the next four years estimated at $750 billion". The only thing that worries me about these jokes is that their humor probably goes unnoticed by the average citizen. It would be a pity to allow such mirth-provoking notions to be appreciated only by a select few, so I decided it would be a good idea to devote some space to the requisite background knowledge, which also happens to be one of my favorite topics: the lore of very large (and very small) numbers.

I have always suspected that relatively few people really know the difference between a million and a billion. To be sure, people generally know it well enough to sense the humor in the joke about when the earth will fall into the sun, but what the difference is *precisely*—well, that is something else. I once heard a radio news announcer say, "The drought has cost California agriculture somewhere between nine hundred thousand and a billion dollars." Come again? This kind of thing worries me. In a society where big numbers are commonplace, we cannot afford to have such appalling number ignorance as we do. Or do we actually suffer from *number numbness?* Are we

growing ever number to ever-growing numbers?

What do people think when they read ominous headlines like the ones above? What do they think when they read about nuclear weapons with 20-kiloton yields? Or 60-megaton yields? Does the number really register—or is it just another cause for a yawn? "Ho hum, I always knew the Russians could kill us all 20 times over. So now it's 200 times, eh? Well, we can be thankful it's not 2,000, can't we?"

What do people think about the fact that in some heavily populated areas of the U.S., it is typical for the price of a house to be a quarter of a million dollars? What do people think when they hear radio commercials for savings institutions telling them that if they invest now, they could have a million dollars on retirement? Can *everyone* be a millionaire? Do we now *expect* houses to take a fourth of a millionaire's fortune? What ever has become of the once-glittery connotations of the word "millionaire"?

* * *

I once taught a small beginning physics class on the thirteenth floor of Hunter College in New York City. From the window we had a magnificent view of the skyscrapers of midtown Manhattan. In one of the opening sessions, I wanted to teach my students about estimates and significant figures, so I asked them to estimate the height of the Empire State Building. In a class of ten students, not one came within a factor of two of the correct answer (1,472 feet with the television antenna, 1,250 without). Most of the estimates were between 300 and 500 feet. One person thought 50 feet was right—a truly amazing underestimate; another thought it was a mile. It turned out that this person had actually calculated the answer, guessing 50 feet per story and 100 stories or so, thus getting about 5,000 feet. Where one person thought each *story* was 50 feet high, another thought the whole 102-story *building* was that high. This startling episode had a deep effect on me.

It is fashionable for people to decry the appalling illiteracy of this generation, particularly its supposed inability to write grammatical English. But what of the appalling *innumeracy* of most people, old and young, when it comes to making sense of the numbers that, in point of fact, and whether they like it or not, run their lives? As Senator Everett Dirksen once said, "A billion here, a billion there—soon you're talking real money."

The world is gigantic, no question about it. There are a lot of people, a lot of needs, and it all adds up to a certain degree of incomprehensibility. But that is no excuse for not being able to understand—or even relate to—numbers whose purpose is to summarize in a few symbols some salient aspects of those huge realities. Most likely the readers of this article are not the ones I am worried about. It is nonetheless certain that every reader of this article knows many people who are ill at ease with large numbers of the sort that appear in our government's budget, in the gross national product, corporation budgets, and so on. To people whose minds go blank when they hear something ending in "illion", all big numbers are the same, so that exponential explosions make no difference. Such an inability to relate to large numbers is clearly bad for society. It leads people to ignore big issues on the grounds that they are incomprehensible. The way I see it, therefore, anything that can be done to correct the rampant innumeracy of our society is well worth doing. As I said above, I do not expect this article to reveal profound new insights to its readers (although I hope it will intrigue them); rather, I hope it will give them the materials and the impetus to convey a vivid sense of numbers to their friends and students.

* * *

As an aid to numerical horse sense, I thought I would indulge in a small orgy of questions and answers. Ready? Let's go! How many letters are there in a bookstore? Don't calculate—just guess. Did you say about a billion? That has nine zeros (1,000,000,000). If you did, that is a pretty sensible estimate. If you didn't, were you too high or too low? In retrospect, does your estimate seem far-

fetched? What intuitive cues suggest that a billion is appropriate, rather than, say, a million or a trillion? Well, let's calculate it. Say there are 10,000 books in a typical bookstore. (Where did I get this? I just estimated it off the top of my head, but on calculation, it seems reasonable to me, perhaps a bit on the low side.) Now each book has a couple of hundred pages filled with text. How many words per page—a hundred? A thousand? Somewhere in between, undoubtedly. Let's just say 500. And how many letters per word? Oh, about five, on the average. So we have 10,000 × 200 × 500 × 5, which comes to five billion. Oh, well—who cares about a factor of five when you're up this high? I'd say that if you were within a factor of ten of this (say, between 500 million and 50 billion), you were doing pretty well. Now, could we have sensed this *in advance*—by which I mean, *without calculation?*

We were faced with a choice. Which of the following twelve possibilities is the most likely:

(a) 10;
(b) 100;
(c) 1,000;
(d) 10,000;
(e) 100,000;
(f) 1,000,000;
(g) 10,000,000;
(h) 100,000,000;
(i) 1,000,000,000;
(j) 10,000,000,000;
(k) 100,000,000,000;
(l) 1,000,000,000,000?

In the United States, this last number, with its twelve zeros, is called a *trillion;* in most other countries it is called a *billion.* People in those countries reserve "trillion" for the truly enormous number 1,000,000,000,000,000,000—to us a "quintillion"—though hardly anyone knows that term.

What most people truly don't appreciate is that making such a guess is very much the same as looking at the chairs in a room and guessing quickly if there are two or seven or fifteen. It is just that here, what we are guessing at is the number of zeros in a numeral, that is, the logarithm (to the base 10) of the number. *If we can develop a sense for the number of chairs in a room, why not as good a sense for the number of zeros in a numeral?* That is the basic premise of this article.

Of course there is a difference between these two types of numerical horse sense. It is one thing to look at a numeral such as "10000000000000" and to have an intuitive feeling, without counting, that it has somewhere around twelve zeros—certainly more than ten and fewer than fifteen. It is quite another thing to look at an aerial photograph of a logjam (see Figure 6-1) and to be able to sense, visually or intuitively or somewhere in between, that there must be between three and five zeros in the decimal representation of the number of logs in the jam—in other words, that 10,000 is the closest power of 10, that 1,000 would definitely be too low, and that 100,000 would be too high. Such an ability is simply a form of number perception one level of abstraction higher than the usual kind of number perception. But one level of abstraction should not be too hard to handle.

The trick, of course, is practice. You have to get used to the idea that ten is a very big number of zeros for a numeral to have, that five is pretty big, and that three is almost graspable. Probably what is most important is that you should have a prototype example for each number of zeros. For instance: *Three* zeros would take care of the number of students in your high school: 1,000, give or take a factor of three. (In numbers having just a few zeros we are always willing to forgive a factor of three or so in either direction, as long as we are merely estimating and not going for exactness.) *Four* zeros is the number of books in a non-huge bookstore. *Five* zeros is the size of a typical county seat: 100,000 souls or so. *Six* zeros—that is, a million—is getting to be a large city: Minneapolis, San

Diego, Brasília, Marseilles, Dar és Salaam. *Seven* zeros is getting huge: Shanghai, Mexico City, Seoul, Paris, New York. Just how many cities do you think there are in the world with a population of a million or more? Of them, how many do you think you have never heard of? What if you lowered the threshold to 100,000? How many towns are there in the United States with a population of 1,000 or less? Here is where practice helps.

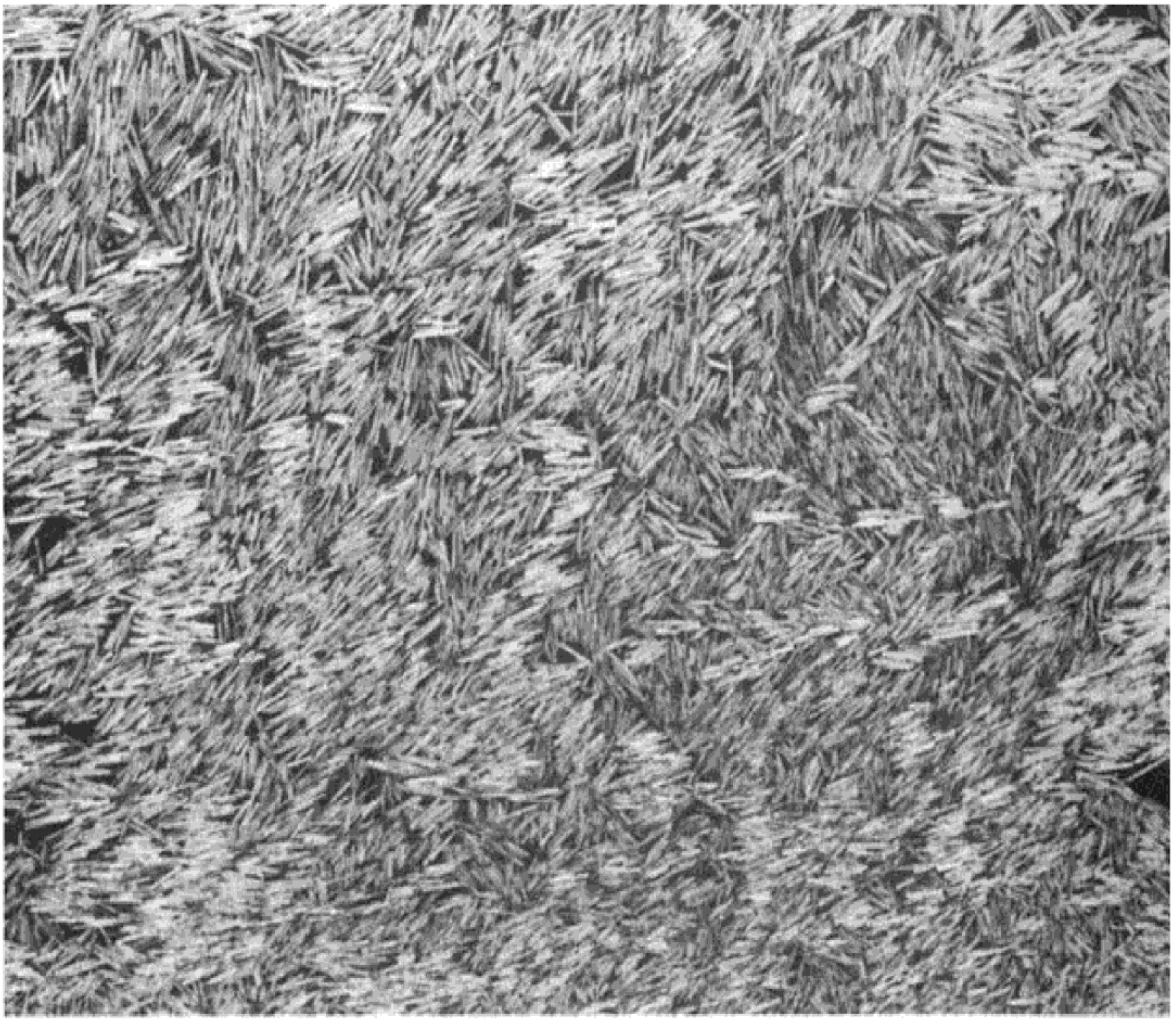


FIGURE 6–1. *Aerial view of a logjam in Oregon. How many logs?* [*Photo by Ray Atkeson.*]

I said that you should have one prototype example for each number of digits. Actually, that is silly. You should have a few. In order to have a concrete sense of "nine-zero-ness", you need to see it instantiated in several different media, preferably as diverse as populations, budgets, small objects (ants, coins, letters, etc.), and maybe a couple of miscellaneous places, such as astronomical distances or computer statistics.

Consider the famous claim made by the McDonald's hamburger chain: "Over 25 billion served" (or whatever they say these days). Is this figure credible? Well, if it were ten times bigger—that is, 250 billion—we could divide by the U.S. population more easily. (This is apparent if you happen to know that the U.S. population is about 230 million. For the purposes of this discussion, let us call the U.S. population 250 million, or 2.5×10^8—a common number that everyone should know.) Let us imagine, then, that the claim were "Over 250 billion served". Then we would compute that 1,000 burgers had been cooked for every person in the U.S. But since we deliberately inflated it by a factor of 10, let us now undo that—let us divide our answer by ten, to get 100. Is it plausible that McDonald's has prepared 100 burgers for every person in the U.S.? Sounds reasonable to me; after

Speaking of yearly deaths, here is one we are all used to sweeping under the rug, it seems: 50,000 dead per year (in this country alone) in car accidents. If you count the entire world, it's probably two or three times that many. Can you imagine how we would react if someone said to us today: "Hey, everybody! I've come up with a really nifty invention. Unfortunately, it has a minor defect—every twelve years or so it will wipe out about as many Americans as the population of San Francisco. But wait a minute! Don't go away! The rest of you will love it, I promise!" Now, these statistics are accurate for cars. And yet we seldom hear people chanting, "No cars is good cars!" How many bumper strips have you seen that say, "No more cars!"? Somehow, collectively, we are willing to absorb the loss of 50,000 lives per year without any serious worry. And imagine that half of this—25,000 needless deaths—is due to drunks behind the wheel. Why aren't you just fuming?

* * *

I said I would be a little lighter. All right. Light consists of photons. How many photons per second does a 100-watt bulb put out? About 10^{20}—another biggie. Is it bigger or smaller than the number of grains of sand on a beach? What beach? Say a stretch of beach a mile long, 100 feet wide and six feet deep. What would you estimate? Now calculate it. How about trying the number of drops in the Atlantic Ocean? Then try the number of fish in the ocean. Which are there more of: fish in the sea, or ants on the surface of the earth? Atoms in a blade of grass, or blades of grass on the earth? Blades of grass, or insects? Leaves on a typical oak tree, or hairs on a human head? How many raindrops fall on your town in one second during a terrific downpour?

How many copies of the Mona Lisa have ever been printed? Let's try this one together. Probably it is printed in magazines in the United States a few dozen times per year. Say each of the magazines prints 100,000 copies. That makes a few million copies per year in American magazines, but then there are books and other publications. Maybe we should double or triple our figure for the U.S. To take into account other countries, we can multiply it again by three or four. Now we have hit about 100 million copies per year. Let us assume this held true for each year of this century. That would make nearly ten billion copies of the Mona Lisa! Quite a meme, eh? Probably we have made some mistakes along the way, but give or take a factor of ten, that is very likely about what the number is.

"Give or take a factor of *ten*"!? A moment ago I was saying that a factor of *three* was forgivable, but now, here I am forgiving myself *two* factors of three—that is, an entire order of magnitude. Well, the reason is simple: We are now dealing with larger numbers (10^{10} instead of 10^5), and so it is permissible. This brings up a good rule of thumb. Say an error of a factor of three is permissible for each estimated factor of 100,000. That means we are allowed to be off by a factor of ten—*one* order of magnitude—when we get up to sizes around ten billion, or by a factor of 100 or so (*two* orders of magnitude) when we get up to the square of that, which is 10^{20}, about 2.5 times the size of Rubik's constant. This means it would have been forgivable if Ideal had said, "Over a *billion* billion combinations", since then they would have been off by a factor of only 40—about 1.5 orders of magnitude—which is within our limits when we're dealing with numbers that large.

Why should we be content with an estimate that is only one percent of the actual number, or with an estimate that is 100 times too big? Well, if you consider the base-10 logarithm of the number—the number of zeros—then if we say 18 when the real answer is 20, we are off by only ten percent! Now what entitles us to cavalierly dismiss the magnitude itself and to switch our focus to its logarithm (its order of magnitude)? Well, when numbers get this big, we have no choice. Our perceptual reality begins to shift. We simply *cannot* visualize the actual quantity. The numeral—the string of digits—takes over: our perceptual reality becomes one of numbers of zeros. When does this shift take place? It begins when we can no longer see, in our mind's eye, a collection of the right order of magnitude. For me, this "perceptual logjam" begins at about 10^4—the size of the actual logjam I remember in the photograph. It is important to understand this transition. It is one