

METAMAGICAL THEMAS

Variations on a theme as the essence of imagination

by Douglas R. Hofstadter

George Bernard Shaw once wrote (in *Back to Methuselah*): "You see things; and you say 'Why?' But I dream things that never were; and I say 'Why not?'" When I first heard this aphorism, it made a lasting impression on me. To "dream things that never were"—this is not just a poetic phrase but a truth about human nature. Even the dullest of us is endowed with this strange ability to construct counterfactual worlds and to dream. Why do we have it? What sense does it make? How can one dream, or even "see," what is visibly not there?

On my table sits a Rubik's Cube. I look at it and see a $3 \times 3 \times 3$ cube whose faces turn. I see—so it seems to me—what is there. Some people, however, looked at the cube and saw things that *weren't* there. They saw cubes with shaved edges, spherical "cubes," differently colored cubes, Magic Dominoes, $2 \times 2 \times 2$ cubes, $4 \times 4 \times 4$ and higher-order cubes, skew-twisting cubes, pyramids, octahedrons, dodecahedrons, icosahedrons, four-dimensional polyhedrons. And the list is not complete yet! Indeed, it is impossible to imagine closing the book on this rich idea.

How did it come about? How is it that, in looking directly at something solid and real on a table, people can see far beyond that solidity and reality, can see an "essence," a "core," a "theme" on which to devise variations? I must stress that the solid cube itself is not the theme (although it is convenient and easy to speak as if it were). In the mind of each person who perceives a Rubik's Cube there arises a *concept* we could call Rubik's cubicity. It is not the same concept in each mind, just as not everyone has the same concept of asparagus or Beethoven. The variations that are spun off by a given cube inventor are variations on that concept. In a discussion of perception and invention the distinction between an object and the concept of the object in someone's mind is crucial.

Now, when Sally Cubelover comes up with a new variation, let us sav the

brain, trying as hard as she can to "go against the grain" in order to come up with something original? Does she think to herself, "Golly, Rubik must have really exerted himself to come up with this totally new idea; therefore I too must strain my mind to its limits in order to invent something original"? Surely not. Einstein didn't go around racking his brain, muttering to himself, "How can I come up with a great idea?" Like Einstein (although on a lesser scale), Sally never needs to ask herself, "H'm, let's see, shall I try to figure out some way to spin off a variation on this object sitting here in front of me?" No, she just does what comes naturally.

The bottom line is that invention is much more like falling off a log than like sawing one in two. In spite of Thomas Edison's memorable remark, "Genius is 1 percent inspiration and 99 percent perspiration," we are not all going to become geniuses simply by sweating more or resolving to try harder. A mind follows the path of least resistance, and it is when it feels easiest that it is probably being its most creative. Or, as Mozart used to say, things should "flow like oil"—and Mozart ought to know. Trying harder is not the name of the game: the trick is getting the right concept to begin with, so that making variations on it is like taking candy from a baby.

Uh-oh—I've let the cat out of the bag. Let me, then, straightforwardly state the thesis I shall now elaborate: Making variations on a theme is really the crux of creativity.

On the face of it the thesis is crazy. How can it possibly be true? Aren't variations simply derivative notions, never truly original creations? Isn't the notion of a $4 \times 4 \times 4$ cube simply a result of "twiddling a knob" on the concept of Rubik's cubicity? You merely twist the knob from its "factory setting" of 3 to the new setting of 4, and presto—you've got it! An inner voice protests. That's just too easy. That's certainly not where *relativity* or Rubik's Cube came from, is

across a gap when an Einstein or even a Rubik comes up with a great idea, something that is patently lacking when Sally Cubelover twiddles a knob on the already existing notion of Rubik's Cube?

To be sure, inventing the notion of a $4 \times 4 \times 4$ cube is far less deep than coming up with special or general relativity. This does not mean, however, that the underlying mental processes are necessarily based on totally different principles. Of course, there is an obvious sense in which the underlying mental processes in your brain, my brain, Sally's brain and Einstein's brain are all "the same": they all depend on the neural hardware. But it is not this microscopic, biological level I mean when I suggest that the underlying mental processes in different brains are somehow the same. What I mean is that there are mechanisms, processes, call them what you will, that can be described functionally, without reference to the neural substrate enabling them to take place in brains.

Hence a notion such as twiddling a knob on a concept bears no relation to the activities of neurons in the brain, or at least no obvious relation. Well then, is there any reality to it, or is it just a metaphor? If someday we come to understand the brain, will we then be confident that we are on solid ground when we speak of a brain literally containing concepts? Or will such statements forever remain shaky and metaphorical *façons de parler* compared with such hard science facts as "At the back of each human brain there is a cerebellum"? Well, until words such as "concept" have become as scientifically legitimate as, say, "temperature," we will not have come anywhere close to understanding the brain—at least not in my book.

It must be admitted that at present words such as "concept" are only metaphorical. They are protoscientific terms awaiting explication. This, however, is an excellent reason to try to flesh them out as much as possible, to try to see what the metaphor of twiddling knobs on a concept involves. Pinning down the meaning of such a metaphor will help us to know much more clearly what we would ideally want from a "hard science" explanation of the brain.

This metaphor makes your imagination conjure up a vision of a tangible thing called a concept that literally has some kind of knobs on it, waiting to be twiddled. What I picture in my mind's eye is something that, instead of being built out of millions of neurons, is more like a metallic "black box" with a panel on it bearing a row of plastic knobs whose little pointers tell you what each one's setting is.

To make this image more concrete, let me describe a genuine example of such a black box with knobs. Back in



made piano rolls of all kinds of wonderful music. Nowadays you can buy phonograph records of those rolls being played back on player pianos, but you can do better than that. Many of the best rolls (made on a special kind of piano called a *Vorsetzer*) have been converted into digital cassette tapes, not tapes to be put into a tape recorder but tapes to be played on a piano equipped with a device called a *Pianocorder*. The *Pianocorder* "reads" the magnetic tape and converts it into instructions to the keyboard and pedals, so that your own piano then plays the piece. Each *Pianocorder* has a black box on the front of which is a control panel with a row of three knobs ("*Tempo*," "*Pianissimo*" and "*Fortissimo*") and one switch ("*Soft pedal*"). By twisting the "*Tempo*" knob you can make *Rachmaninoff* speed up; by twiddling the "*Pianissimo*" and "*Fortissimo*" knobs you can make *Horowitz* play more softly or *Rubinstein* more loudly. It is too bad there is not a knob labeled "*Pianist*" so that you could select who plays. It would be interesting to change in midstream from one pianist to another.

This device takes us one step toward realizing a dream of the Canadian pianist *Glenn Gould*. *Gould* is very much at one with the electronic age, and for

years he has been advocating the use of computers to enable people to control the music they hear. You begin with an ordinary recording of, say, *Gould* himself playing a concerto by *Mozart*. This is merely raw data for you to manipulate. On your space-age record player you have a bunch of knobs that enable you to slow the music down or to speed it up *ad libitum*, to control the volume of each separate section of the orchestra, even to correct for high notes played flat by the violinists. In effect you become the conductor, with knobs to control every aspect of the performance dynamically. The fact that it was originally *Gould* at the piano is, by the time you are done with it, irrelevant. By now you have taken over and made the performance your own. Presumably such systems would eventually evolve to the point where you could start with the written score, dispensing entirely with the acoustic recording stage.

Why not carry this farther, then? If we are allowing ourselves to fantasize, why not go as far as we can imagine? Why should our "raw data" be limited to the finite universe of existing pieces? Why should there not be a knob to control the mood of the composition and another to control the composer whose style it is to be written in? This way we could get a

new piece by our favorite composer in any desired mood. But that is too conservative. Why should we be limited to the finite universe of composers already born? Why could there not be a knob to enable us to interpolate between composers, making it possible for us to tune our music-making machine to an even mixture of *Johann Sebastian Bach*, *Giuseppe Verdi* and *John Philip Sousa* (ugh!), or to a position halfway between *Franz Schubert* and the *Sex Pistols* (super-ugh!)? And why stop at interpolation? Why not extrapolate beyond a given composer? For instance, I might want to hear a piece by "the composer who is to *Ravel* as *Ravel* is to *Chopin*." The machine would merely need to calculate the ratios of its knob settings for *Ravel* and *Chopin* and then multiply the *Ravel* settings by those same ratios to come up with a super-*Ravel*.

It is really no trickier than solving any trivial analogy problem: "What is to a triangle as a triangle is to a square?" "What is to Greece as the Falkland Islands are to Britain?" "What is to a water bed as ice is to water?" and other "easy" problems like that. The truth is, of course, quite the contrary: analogy problems are extremely tricky to mechanize. The knobs on most concepts are not so apparent that we can just read their settings right off. The examples above simply carried a thought to a ludicrous extreme. It is nonetheless worth while to look seriously at the idea that a concept can be considered as a machine whose knobs can be twiddled to yield a fabulous array of variations.

The *Rubik's Cube* concept with its "order" knob set at 3 gives rise to an ordinary $3 \times 3 \times 3$ cube, and with that knob set at 4 a $4 \times 4 \times 4$ cube. Come to think of it, doesn't there have to be a separate knob for each dimension so that you can twiddle each one independently of the others? After all, not all variations have to be cubical. The *Magic Domino* is $3 \times 3 \times 2$. Hence if we agree that there are three knobs defining the shape, then in the original cube they all just accidentally happened to have the same setting. Now, given these three knobs we can use our concept—our knobbed machine—to generate such mental objects as a $7 \times 7 \times 7$ *Rubik's Cube*, a $2 \times 2 \times 8$ *Magic Domino*, even a $3 \times 5 \times 9$ *Rubik's Magic Brick* (or, if you will excuse me, a *Rubrick*).

But wait a minute. If there really are only three knobs, we are locked into three dimensions. Obviously we do not want *that*. Then let us add a fourth knob to control the length in the fourth dimension. With this knob we can now make a four-dimensional $2 \times 3 \times 5 \times 7$ *Rubrick* and in addition any *Rubik's Tesseract* we might want. But needless to say, once we have gone through the

The LORD is my shepherd:
I shall not want.
He maketh me to lie down
in green pastures:
he leadeth me
beside the still waters.
He restoreth my soul:
he leadeth me
in the paths of righteousness
for his name's sake.
Yea, though I walk through the valley
of the shadow of death,
I will fear no evil:
for thou art with me;
thy rod and thy staff
they comfort me.
Thou preparest a table before me
in the presence of mine enemies:
thou anointest my head with oil,
my cup runneth over.
Surely goodness and mercy
shall follow me
all the days of my life:
and I will dwell
in the house of the LORD
for ever.

gate from three dimensions to four we should certainly expect to be able to go farther. For any n we could imagine n -dimensional Rubik's objects, for example a $2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8$ Hyper-Rubrick. But something peculiar has happened. We must now conceive of our machine—our concept—as having a potentially *unlimited* number of knobs on it (one for each dimension in n -dimensional space). If n is set at 3, there need be only three more knobs. But if n is 100, we need 100 extra knobs!

No real machine has a variable number of knobs. This may sound like a somewhat trivial observation, but it leads into some tricky waters. The point is that if we want to keep on using the metaphor of a concept as a machine with knobs on it, we have to stretch the very concept of "knob." New knobs must be able to materialize, depending on the settings of other knobs. Or you can think of it this way, if you like: On each concept there are potentially an infinite number of knobs, and at any moment some new knobs may be revealed as a result of the settings of other knobs.

I am not sure I like that view, however. It is too cut and dried, too closed and predetermined for my taste. I am more in favor of a view holding that the "knobs" on any one concept depend on the set of concepts that happen to be active simultaneously in the mind of the person. This way new knobs can spring into existence seemingly out of nowhere: they do not all have to be present from the beginning in the isolated concept. If we go back to Rubik, it would mean that *his* concept of Rubik's Cube did not (and still does not) explicitly—or even implicitly—incorporate all the possible variations people may come up with. Rubik anticipated, and even designed, many of the objects that have subsequently appeared and that we perceive as variations on a theme, but his mind did not exhaust that fertile theme. Once the concept entered the public domain it began to develop in ways Rubik never could have anticipated.

There is a way concepts have of slipping from one into another, following a quite unpredictable path. This slippage affords us perhaps our deepest visions into the hidden nature of our conceptual networks. Sometimes the slippage is totally accidental, as it is when we make a typographical error or a grammatical mistake, choose the wrong word, create a malapropism or concoct a phrase out of other phrases. Sometimes it is nonaccidental but comes straight out of our unconscious mind. By "nonaccidental" I do not mean to imply it is deliberate. It is not that we say to ourselves, "I think I shall now slip from one concept into a variation of it," since this kind of deliberate conscious

slippage is most often quite uninspired and infertile. "How to think" and "how to be creative" books (even thoughtful ones such as George Pólya's *How to Solve It*) are, for that reason, of little use to the would-be genius.

Strange though it may sound, non-deliberate yet nonaccidental slippage permeates our thought processes and is, I believe, the very core of thinking. This subconscious manufacture of subjunctive variations on a theme is something that goes on day and night in each of us, usually without our slightest awareness of it. It is one of those things that like air or gravity or three-dimensionality tend to elude our perception because they define the fabric of our lives.

To make this concrete let me contrast an example of "deliberate" slippage with an example of "non-deliberate but nonaccidental" slippage. Imagine that one summer evening you and Sally Cubelover have just walked into a surprisingly crowded coffeehouse. Now go ahead and manufacture a few variants on that scene, in whatever ways you want. What kinds of things do you come up with when you deliberately "slip" this scene into variants of itself?

If you are like most people, you will come up with some pretty obvious slippages, made by moving along what seem to be the most obvious "axes of slippability." Typical examples are:

It could have been a winter evening instead of a summer one.

You could have come with Adam Spherehater instead of Sally Cubelover.

You could have gone to a Chinese restaurant instead of a coffeehouse.

The coffeehouse could have been almost empty.

Now contrast your variations with one I overheard one summer evening in a crowded coffeehouse when a man walked in with a woman. He said to her: "I'm glad I'm not a waitress here tonight." This is a perfect example of a subjunctive variation on the given theme, but unlike yours it emerged spontaneously and for the purpose of communication. The list above looks positively mundane next to this casual remark. And the remark was not considered to be particularly clever or ingenious by his companion. She merely agreed with the thought by saying "Yeah." It caught my attention less because I thought it was clever than because I am always on the lookout for interesting examples of slippability.

I found this example not just mildly interesting but highly provocative. If you try to analyze it, it would appear at first to force you as a listener to imagine a sex-change operation done in record time. But when you simply *understand* the remark, you see that in reality there was no intention in the speaker's mind of bringing up such a bizarre image.

His remark was much more figurative, much more abstract. It was based on an instantaneous perception of the situation, a kind of there-but-for-the-grace-of-God-go-I reaction, that induces a quick flash to the effect of "Simply because I am human I can place myself in the shoes of that waitress; therefore I *could have been* that waitress." Logical or not, this is the way our thoughts go.

Thus when you look carefully, you see that this particular thought has practically nothing to do with the speaker, or even with the waitresses he sees. It is just his slip way of saying, "Boy, it sure is crowded here tonight." And that, of course, is why nobody really is thrown for a loop by such a remark. Yet the remark was made in such a way that it invites you to lightly "map" him onto a waitress, just barely noticing (if you notice it at all) that there is a sex difference. What an amazingly subtle thought process is involved here! And what is even more amazing (and frustrating) to me is how hard it is to point out to people how amazing it is. People find it hard indeed to see what is amazing about the ordinary behavior of people. They cannot quite imagine how it might have been otherwise. It is hard to slip mentally into a world in which people would *not* think by slipping mentally into other worlds, very hard to create a counterfactual world in which counterfactuals were not a key ingredient of thought.

Here, briefly, is another example: I was having a conversation with someone who told me he came from a town in Indiana named Whiting. Since I did not know where the town was, he said it was near Chicago and then added, "Whiting would be in Illinois if it weren't for the state line." Again the remark was made casually; it was certainly not an attempt to be witty. He didn't chuckle, nor did I. I simply smiled, signaling my understanding of his meaning, and then we went on. But try to analyze what this remark means! On a logical level it is somewhat like a tautology. Whiting would undoubtedly be in Illinois if the Illinois state line made it so, but then if we are letting the border of Illinois slip, what is to prevent it from enclosing Toronto or even Peking? But psychologically the remark is quite sensible, relying implicitly on some shared intuition about the impermanence and arbitrariness of geographic boundary lines, an intuition about how state lines could indeed slip. It wouldn't ever occur to us—except in discourse such as the present one—to slip the Illinois state line around Peking. Yet it *did* occur spontaneously in that man's mind. He was thus revealing some deep qualities of his mental representation of Whiting.

Remarks such as this one betray the hidden "fault lines" of the mind; they

show which things can slip and which cannot. And yet they also reveal that nothing is reliably unslippable. Context contributes an unexpected quality to the knobs that are perceived on a given concept. The knobs are not displayed on a neat little control panel, forever unchangeable. Instead changing the context is like taking a tour around the concept, and as you get to see it from various angles more and more of its knobs are revealed. Some people get to be good at perceiving fresh knobs on concepts where others thought there were none, just as some people get to be good at seeing mushrooms in a forest where others see none.

It may still be tempting to think that for each well-defined concept there must be an "ultimate" or "definitive" set of knobs such that the abstract space traced out by all possible combinations of the knobs yields all possible instantiations of the concept. A case in point is the concept of the letter *A*. The typographically naive might think that here there would be only four or five knobs to twiddle. The more you look into letterforms, however, the more elusive any attempt to define them mathematically becomes. One of the most valiant efforts at "knobbifying" the alphabet has been the letterform-defining system called Metafont, developed at Stanford by the computer scientist Donald E. Knuth.

Knuth's purpose is not to arrive at an ultimate mathematical definition of the letters of the alphabet (I suspect he would laugh at the very notion) but to allow a user to create "knobbed" letters; we could call them letter schemas. This means you can choose for yourself what the variable aspects of a letter are, and then, with the aid of Metafont, you can easily devise knobs allowing those aspects to vary. That would include just about anything you could think of: the length of a line in a letter, widenings or taperings of lines, the shape of a curve, the presence or absence of serifs and so on. The full power of the computer is then at your disposal; you can twiddle away to your heart's content, and the computer will generate all the products your knob settings define.

Going further than dealing with letters in isolation, Knuth allowed letters to *share* parameters. That is, a single "master knob" can control a feature common to a group of related letters. Then, although there may be hundreds of knobs when you count the knobs on all the control panels of all the letters of the alphabet, there will be far fewer master knobs that have a deep and pervasive influence on the entire alphabet. What happens, in effect, is that by twiddling the master knobs alone you have a way of drifting smoothly through a "space" of typefaces.

Perhaps Knuth's greatest virtuoso feat with Metafont is what he did with Psalm 23, which in English consists of 593 characters (including spaces). Knuth had defined a full set of characters that shared 28 master knobs. He began his printed version of the psalm with all 28 knobs at their leftmost settings. Then, character by character, he inched his way toward the rightmost settings, turning each knob $1/592$ of the way, so that by the time he reached the final character the extreme opposite end of the spectrum had been reached. In one sense every letter in the psalm is in a different typeface. And yet the transition is so smooth that it is locally undetectable. This example is drawn from Knuth's inspiring article "The Concept of a Metafont" (*Visible Language*, Vol. 16, No. 1, pages 3-27; Winter, 1982).

One of Knuth's main theses is that with computers we are in the position of being able to describe not just a thing in itself but *how that thing would vary*. Metafont epitomizes this thesis. In a sense the computer, rather than simply blindly reproducing fixed letterforms, has a crude "understanding" of what it is drawing, created by the designer who "knobbified" the letters. And yet one should be careful not to fall under the illusion, easily created by Metafont's extraordinary power, that these 28 master knobs—or any finite set of knobs—actually span the entire space of all possible typefaces. This is about as far from the truth as would be the assertion that the set of all possible types of human faces could be captured in a computer program with 28 knobs.

Even the space of all versions of the letter *A* is only barely explored when you twiddle *all* the knobs in Knuth's representation of *A*. not just the 28 master knobs it shares with other letters but the many "private" knobs it has as well. Even 1,000 knobs would not suffice to cover the variety of *A*'s people can recognize easily. Some examples of the richness of the full space of *A*'s are given in the illustration on page 21.

There is a crucial distinction to be made here. A machine with one off-on switch (the most trivial kind of knob) for each square in a 200×200 grid will certainly define any of the *A*'s shown, but it will not exclude *B*'s, *Z*'s or pictures of your grandmother. It is another matter entirely to define a set of knobs whose twiddling covers all the *A*'s shown, all the interpolations between them (as well as extrapolations in all possible directions), yet never leads you out of the space of recognizable *A*'s. This is far trickier! Similarly, it is a nearly trivial project to write a computer program that in principle writes all possible sequences and combinations of tones in all possible rhythmic patterns, but it is a far cry from writing a program that produc-

es only pieces in the style of Bach. Putting on the constraints makes the program unutterably more complex.

What Metafont gives you, rather than the full space of the *A*'s in all type faces, is a *subspace*, and such a tightly related one that it is perhaps best to call it a *family*. Nobody would be able to predict the existence of butterflies from having studied only ants, wasps and beetles. Likewise nobody would be able to predict the full magnitude of the *concept* of *A* from seeing the family traced out by the finite number of knobs in any realistic Metafont program for *A*.

The next stage beyond Metafont will be a program that on its own can extract a set of knobs from a set of given input letters. This, however, is a program for the distant future. At present it takes a highly trained and perceptive type designer months to convert a set of letterforms into Metafont programs with knobs flexible enough to warrant the trouble taken. It would be relatively easy to do it in some crude mechanical way, but what one wants is for stylistic unity to be preserved even as the master knobs are twiddled. Therefore the task of mechanizing the production of Metafont programs amounts to the mechanization of artistic perception. It is hardly around the corner.

There is a curious work called *One Book Five Ways*, published in 1978 by William Kaufmann, Inc. It came about as follows. As an educational experiment in comparative publishing procedures, a manuscript on indoor gardening was sent to five different university presses. The presses all cooperated in coming up with full publication versions of the book, which turned out to be stunningly different at all conceivable levels. William Kaufmann had the bright idea of publishing pieces of the various versions side by side; what resulted was this elegant "metabook." It brings home the meaning of the old saying that there is more than one way to skin a cat.

Making the book was an extravagant foray into "possible worlds," the kind of thing that seems very hard to do. One of Knuth's points, however, is that as computers become commoner and more sophisticated the notion of skinning a cat in 10 different ways will gradually become less extravagant. Once your "cat" has been represented inside a powerful computer program it is no longer just one cat; it is a "cat schema," a mold for many cats at once, and you can skin them all differently (or at least until the cat schema runs out of lives).

Text formatting and computer typesetting present us easily with many alternative versions of a piece of text. Metafont shows us how letterforms can glide into alternative versions of themselves. It is now up to us to continue this trend

of extending our abilities to see farther into the space of possibilities surrounding what *is*. We should use the power of computers to aid us in seeing the full concept—the implicit “sphere of hypothetical variations”—surrounding any static, frozen perception.

I have concocted a name for this imaginary sphere: I call it the “implicosphere,” which stands for “implicit counterfactual sphere,” referring to things that never were but that we cannot help seeing anyway. (The word can also be taken as referring to the sphere of implications surrounding any given idea.) If we want to enlist computers as our partners in this venture of inventing variations on a theme, which is to say turning implicospheres into explicospheres, we must give them the ability to spot knobs themselves, not just to accept knobs we human beings have spotted. To do this we will have to look deeply into the nature of slippability, into the fine-grained structure of those networks of concepts in human minds.

One way to imagine how slippability might be realized in the mind is to suppose that each new concept begins life as a compound of previous concepts and that from the slippability of those concepts it inherits a certain amount of slippability. That is, since any of its constituents can slip in various ways, modes of slippage are induced in the whole. Generally letting a constituent concept slip in its simplest ways is enough, since when more than one slippage comes at a time, it can already create many unexpected effects. Gradually, as the space of possibilities of the new concept—the implicosphere—is traced out, the commonest and most useful of the slippages become more closely and directly associated with the new concept itself rather than having to be derived repeatedly from its constituents. In this way the new concept’s implicosphere becomes more and more explicitly explored. Eventually the new concept becomes old, and it reaches the point where it too can be used as a constituent of a fresh, new, young concept.

Some examples of this kind of thing were presented in “Metamagical Themas” for September, 1981. Although September is almost October and 1981 is almost 1982, you may not have those examples at your mind’s fingertips, or on the tip of your mind’s tongue. Let me give a few more examples of slippage of a new notion based on slipping some of its parts in their simplest ways. The notion I have chosen is the one of you yourself sitting there reading this very column at this very moment. Here are some elements of the implicosphere of that concept:

You are almost reading the Septem-

You are almost reading a piece by Richard Hofstadter, the historian.

You are almost reading a column by Martin Gardner.

You are almost reading this column in French.

You are almost reading my book *Gödel, Escher, Bach: an Eternal Golden Braid*.

You are almost writing this column.

I am almost talking to you.

By now the original concept is almost lost in a sea of “almost” variations, but it has been enriched by the exploration, and when you come back to it, it will have been that much more reified as a stand-alone concept, a single entity rather than a compound one. After a while, under the proper triggering circumstances, this very example may be retrieved from memory as naturally and effortlessly as the concept of “fish” is.

This is an important idea: the test of whether a concept has really come into its own, the test of its genuine mental existence, is its retrievability by that process of unconscious recall. That is what lets you know the concept has been firmly planted in the soil of your mind. It is not whether the concept appears to be “atomic,” in the sense that you have a single word to express it by. That is far too superficial.

Here is an example to illustrate why. A friend told me recently that the first edition of *Encyclopaedia Britannica* consisted of three volumes: Volume 1 was A through B, Volume 2 was C through L, and Volume 3 covered the rest of the alphabet. A was given 511 pages and M through Z were given 753 pages altogether! This amusing fact instantaneously triggered the retrieval of another memory, implanted in my mind years ago under totally unremembered circumstances, of how records used to be made back in the days when there was no magnetic tape and the master disk was actually cut during the live performance. The performers would be singing or playing along and all of a sudden the recording engineer would notice that there was not much room left on the disk, and so the performers would be given a signal to hurry up. As a result the tempo would be faster the closer to the center of the disk the needle got. I think it is obvious why the one concept triggered the retrieval of the other. But then again, is it really obvious?

On the surface the two concepts are completely unrelated. One concerns printed matter, books, the alphabet and so on; the other concerns wax disks, sounds, performers, recording techniques and so on. At some deeper conceptual level, however, these really *are* the same idea. There is just one idea here, and this idea I call a conceptual skeleton. Try to verbalize it. It is cer-

tainly not just one word. It will take you a while. And when you do come up with a phrase, the chances are it will be awkward and stilted—and still not quite right!

Both of the cited instances of this conceptual skeleton—in itself nameless, majestically nonverbalizable—are floating about in the implicosphere that surrounds it, along with numerous other examples I am unaware of, not yet having twiddled enough knobs on that concept. I do not, of course, even know which knobs it has, but I may eventually find out. The point is that the concept itself has been reified; this much is proved by the fact that it acts as a point of immediate reference, that under the proper circumstances my memory mechanisms are capable of accessing it directly. The vast majority of our concepts are wordless in this way, although we can certainly make a stab at verbalizing them when we need to.

Early in this column I stated a thesis: that the crux of creativity lies in the ability to manufacture variations on a theme. I hope now to have sufficiently fleshed out this thesis for you to understand the full richness of what I meant when I said “variations on a theme.” The notion encompasses knobs, parameters, slippability, counterfactual conditionals, subjunctives, “almost” situations, implicospheres, conceptual skeletons, mental reification, memory retrieval—and more.

The question may persist in your mind: Aren't variations on a theme somehow trivial compared with the invention of the theme itself? This leads one back to the seductive notion that Einstein and other mighty creators are cut from a different cloth from the one used to make ordinary mortals, or at least that certain cognitive acts of such creators involve principles transcending the everyday ones. This is something I do not believe at all. If you look into the history of science, for instance, you will see that every idea is built on a thousand related ideas. Careful analysis leads one to see that what we choose to call a new theme is itself always some kind of variation, on a deep level, of earlier themes.

Newton said that if he saw farther than others, it was only because he stood on the shoulders of giants. Too often, however, we simply indulge in wishful thinking when we imagine that a clever or beautiful idea was somehow due to unanalyzable, magical, transcendent insight rather than to any mechanisms, as if all mechanisms by their very nature are necessarily shallow and mundane.

My own mental image of the creative process involves viewing the organization of a mind as consisting of thousands, if not millions, of overlapping and intermingling implicospheres, at the

center of each of which is a conceptual skeleton. The implicosphere is a flickering, ephemeral thing, rather like the electron cloud, with its quantum-mechanical elusiveness, around an atomic nucleus. If you have studied quantum chemistry, you know that the fluid nature of chemical bonds can best be understood as a consequence of the curious quantum-mechanical overlap of electronic wave functions in space, wave functions belonging to electrons orbiting neighboring nuclei. In a metaphorically similar way, it seems to me, the crazy and unexpected associations allowing creative insights to pop seemingly out of nowhere may well be consequences of a similar chemistry of concepts with its own special types of “bonds” that emerge out of an underlying “neuron mechanics.”

The novelist Arthur Koestler has long been a champion of a mystical view of human creativity, advocating occult views of the mind while at the same time eloquently and objectively describing its workings. In his book *The Act of Creation* he presents a theory of creativity whose key concept he calls bisociation: the simultaneous activation and interaction of two previously unconnected concepts. That view emphasizes the coming together of two concepts, bypassing discussion of the internal structure of a single concept. This is in keeping with Koestler's philosophy of believing wholes are somehow greater than the sum of their parts.

In contrast, I have been emphasizing the idea of the internal structure of one concept. In my view the way concepts can bond together and form conceptual molecules at all levels of complexity is a consequence of their internal structure. The crux of the matter is the internal structure of a single concept and how the concept “reaches out” toward things it is not. I am not one to believe wholes elude description in terms of their parts. If we come to understand the “physics of concepts,” then perhaps we can derive from it a “chemistry of creativity,” just as we can derive the principles of chemistry from those of physics. But again it is not just around the corner. Alan Turing's words of cautious enthusiasm about artificial intelligence remain as apt as they were in 1950, when he wrote them in concluding his famous article “Computing Machinery and Intelligence”: “We can only see a short distance ahead, but we can see plenty there that needs to be done.”

Recently I happened to read a headline on the cover of a popular electronics magazine that blared something to the effect of CHIPS THAT SEE. Bosh! I'll start believing in chips that see as soon as they start seeing things that never were and asking “Why not?”