

# VARIATIONS ON A THEME AS THE CRUX OF CREATIVITY

From Douglas R. Hofstadter's *Metamagical Themas* (Abridged)

## IGNORE REFERENCES TO OTHER CHAPTERS, OR TO FIGURES.

You see things; and you say "Why?" *But I dream things that never were; and say "Why not?"* -George Bernard Shaw in *Back to Methuselah*

When I first heard this beautiful line, it made a deep impression on me. It was in the spring of 1968, during the presidential campaign, and Robert Kennedy had made this line his theme. I thought it was wonderfully poetic, and I assumed he himself had dreamt it up. Only many years later did I find out I was quite wrong: Not only had he not made it up, but the character who utters it in the Shaw play is the snake in the Garden of Eden! How disturbing! Why couldn't it have been the way I thought?

"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 come up with counterfactual worlds and to dream. But why do we have this ability--in fact, this proclivity? What sense does it make? And how can one "see" what is visibly *not there*?

On my table sits a Rubik's Cube. I look at it and see a 3 X 3 X 3 cube whose faces turn. I see--so it seems to me--what is there. But some people looked at that cube and saw things that *weren't* there. They saw cubes with shaved edges, spherical "cubes", differently colored cubes, Magic Dominos, 2 X 2 X 2 cubes, 4X4X4 and higher-order cubes, skew-twisting cubes, pyramids, octahedra, dodecahedra, icosahedra, four-dimensional magic polyhedra. And the list is not complete yet! you wait!

How did this 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" upon 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* that we could call "Rubik's-Cubicity". It's not the same concept in each mind, just as not everyone has the same concept of asparagus or of Beethoven. The variations that are spun off by a given cube-inventor are variations on that concept. In a discussion of perception and invention, this distinction between an object and some mind's concept of the object is simple but crucial.

Now when Eve Rybody comes up with a new variation--let's say the 4 X 4 X 4--is it as a result of wracking her brain, trying as hard as she can to "go against the grain", so as to come up with something original? Does she think to herself, "Golly, that 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"? No, no, no! A thousand times no. Einstein didn't go around racking his brain, muttering to himself, "How, oh how, can I come up with a Great Idea?" Like Einstein (although perhaps on a lesser scale), Eve never needs to ask herself, "Hmm, 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. Despite Thomas Alva Edison's memorable remark, "Genius is 2 percent inspiration and 98 percent perspiration", we're not all going to become geniuses simply by sweating more or resolving to *try harder*. A mind follows its path of least resistance, and it's when it feels easiest that it is most likely 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 making variations on it is like taking candy from a baby.

Uh-oh--now I've given the cat away! So let me boldly state the thesis that I shall now elaborate: *Making variations on a theme is really the crux of creativity.*

---

On the face of it, this 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 4X4X4 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 Rubik's Cube, the *Rite of Spring*, relativity, or *Rome and Juliet* came from, is it? Isn't there a 'magic spark' that leaps across a gap when a Rubik or a Stravinsky or an Einstein or a Shakespeare comes up with a great idea, something that is patently lacking when an Eve Rybody merely twiddles a knob on an already-existing notion?"

Well, of course, inventing the notion of a 4X4X4 cube is far less deep than coming up with special or general relativity. I'd be the last to deny that. But that doesn't mean that the underlying mental processes are necessarily based on totally different principles. Of course, there is a boring sense in which the underlying mental processes in your brain, my brain, Eve's brain, and Einstein's brain are all "the same"--namely, they all depend on neural hardware. But it is not at such a microscopic, such a biological level that I mean it 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 that enables them to take place in brains.

Thus, a notion like "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 at last come to understand the brain, will we then be confident that we're on solid ground when we speak of a brain literally *containing concepts*? Or will such statements forever remain shaky compared to such hard-science facts as "At the back of each human brain there is a cerebellum"? Well, until words like "concept" have become terms as scientifically legitimate as, say, "neuron" or "cerebellum", we will not have come anywhere close to understanding the brain--at least not in my book.

However, it must be admitted that at present, words like "concept" are only metaphorical. They are protoscientific terms awaiting explication. But this is a very good 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 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 a set of knobs on it, just 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, containing a row of plastic knobs with little pointers on them, telling you what each one's setting is.

Just to make this image more concrete, let me describe a genuine example of such a black box with knobs. Back in the old days of player pianos, good pianists made piano rolls of all sorts 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 to be played on tape recorders, but on pianos specially equipped with a device called a "Pianocorder". This "reads" the magnetic tape and converts it into instructions to the keyboard and pedals, so that your 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's too bad there's not a knob labeled "pianist" so that you can select who plays. After all, it would be interesting to change Horowitzes in midstream.

---

This device takes us one step toward realizing a dream of the unique Canadian pianist Glenn Gould. Gould is very tuned in to the electronic age, and for years has been advocating using computers to allow people to control the music they hear. You begin with an ordinary recording of, say, Glenn Gould himself playing a concerto by Mozart. But this is merely raw data for you to tamper with. On your space-age record player, you have a bunch of knobs that allow you to slow the music down or to speed it up *ad libitum*, to control the volume of all the separate sections of the orchestra, even to correct for high notes played too 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 Glenn Gould at the piano is, by the time you're done with it, irrelevant. By now you've totally taken over and made it your very own performance! Presumably, such systems would

eventually evolve to the point where you could start with the mere written score, dispensing entirely with the acoustic recording stage.

But why not carry this further, then? If we are allowing ourselves to fantasize, why not go as far as we can imagine? Why could our "raw data" be limited to the finite universe of already-composed pieces? Why could there not be a knob to control the mood of the composition, 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 really, this is too conservative. Why should we be limited to the finite universe of already-born composers? Why could there not be a knob to allow us to interpolate between composers, thus 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 a position halfway between 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's no trickier than solving any old analogy problem--you know, simple problems like this:

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 visual art as fugues are to music?

What is to German as Shakespeare's plays are to English?

What is to English as simplified characters are to Chinese?

What is to 1-2-3-4-4-3-2-1 as 4 is to 1-2-3-4-5-4-3-2-1

What is to *pqc* as *abc* is to *aqc*?

The truth is, of course, that analogy problems are staunchly resistant to mechanization. The knobs on most concepts are not so apparent as to allow us to just read their settings right off. The examples above simply carried a sensible thought to a ludicrous extreme. However, it is still worthwhile to look seriously at the idea that a concept can be considered as a "knobbed machine" whose knobs can be twiddled to produce a bewildering array of variations.

=====  
The Rubik's-Cube concept, with its "order" knob set at 3, produces an ordinary 3X3X3 cube--and with that knob set at 4, a 4X4X4. 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 X 3 X 2. So 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 X 7 X 7 Rubik's Cube, a 2 X 2 X 8 Magic Domino, even a 3 X 5 X 9 Rubik's Magic Brick (or, if you'll pardon me, a "Rubrick").

But wait a minute--if there really are just three knobs, then we're locked into three dimensions! Obviously we don't want that. So let's add a fourth knob to control tile length in the fourth dimensions. With this knob, we can now make a four-dimensional 2 X 3 X 5 X 7 Rubrick, as well as any Rubik's Tesseract that we might want. But needless to say, once we've gone through the gate from three dimensions to four, certainly we should expect to be able to go further. For any *n*, we could imagine *n*-dimensional Rubik's objects--for example, a 2X3X4X5X6X7X8 Hyper-Rubrick. But now 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 to 3, there need only be 3 more knobs. But if *n* is 100, we need 100 extra knobs!

No real machine has a variable number of knobs. Now this may sound like a somewhat trivial observation. However, it leads into some tricky waters. The point is that, if we wish 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 sprout, depending on the settings of other knobs. Or you can think of it this way, if you wish: on each, concept, there are potentially an infinite number of knobs, and at any moment, some new knobs may get revealed as a consequence of the settings of other knobs.

I'm not sure I like that view, however. It's too cut and dried, too closed and predetermined for my tastes. I am more in favor of a view that says that the knobs on any one concept depend on the set of concepts that happen to be awake simultaneously in the mind of the person. This way, new knobs can spring into existence seemingly out of nowhere; they don't all have to be present from the outset in the isolated concept. If we go back to Rubik, this would mean that *his* concept of Rubik's Cube didn't (and still doesn't) explicitly--or even implicitly--contain all the possible variations that people may come up with. Rubik anticipated, and even designed, many of the objects that have subsequently appeared at that we perceive as "variations on a theme"--but certainly, his mind did not exhaust that fertile theme. Once the concept entered the public domain, it started migrating and developing in ways that Rubik could never have anticipated.

=====

There is a way that concepts have of "slipping" from one into another, following a quite unpredictable path. Careful observation and theorizing about such slippages affords us perhaps our best chance to probe deeply into the hidden murk of our conceptual networks. An example of such a slip is furnished to us whenever we make a typo or a grammatical mistake, utter a malapropism ("She's just back from a one-year stench at Berkeley") or a malapophor (a novel phrase concocted unconsciously from bits and pieces of other phrases, such as "He's such an easy-go-lucky fellow" or "Uh-oh, now I've given the cat away"), or confuse two concepts at a deeply semantic level (e.g., saying "Tuesday" but meaning "February", or saying "midnight" in lieu of "zero degrees"). These types of slip are totally accidental and come straight out of our unconscious mind.

However, sometimes a slippage can be nonaccidental yet still come from the unconscious mind. By "nonaccidental" here, I do not mean to imply that the slip is *deliberate*. It's not that we say to ourselves, "I think I shall now slip from one concept into a variation of it"; indeed, that kind of deliberate, conscious slippage is most often quite uninspired and infertile. "How to Think" and "How to Be Creative" books--even very thoughtful ones such as George Polya's *How to Solve It*--are, for that reason, of little use to the would-be genius.

Strange though it may sound, *nondeliberate yet nonaccidental slippage permeates our mental processes, and is the very crux of fluid thought*. That is my firmly held conviction. 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 very fabric of our lives.

To make this concrete, let me contrast an example of "deliberate" slippage with an example of "nondeliberate but nonaccidental" slippage. Imagine that one summer evening you and Eve Rybody 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 hypothetical variants of itself?

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

- I could have come with Ann Yone instead of Eve Rybody.
- We could have gone to a pancake house instead of a coffeehouse.
- The coffeehouse could have been nearly empty instead of full.
- It could have been a winter's evening instead of a summer's evening.

Now contrast your variations with one that I overheard one evening this past summer in a very crowded coffeehouse, when a man walked in with a woman. He said to her, "I'm sure glad I'm not a waitress here tonight!" This is a perfect example of a subjunctive variation on the given theme--but unlike yours, this one was made without external prompting, and it was made for the purposes of communication to someone. The list above looks positively mundane next to this casually tossed-off 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 not so much because I thought it was clever but mostly 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 glance to force you as listener to imagine a sex-change operation performed in world record time. But when you simply

*understand* the remark, you see that in actuality, 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 sort of "There-but-for-the-grace-of-God-go-I" feeling, which induces a quick flash to the effect of "Simply because I am human, I can place myself in the shoes of that harried waitress--therefore *I could have been* that waitress." Logical or not, this is the way our thoughts go.

So 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's just his flip way of saying, "Hmm, it sure is busy here tonight." And that's of course why nobody really is thrown for a loop by such a remark. Yet it was stated in such a way that it invites you to perform a "light" mapping of him onto a waitress, just barely noticing (if 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 very hard indeed to see what's amazing about the ordinary behavior of people. They cannot quite imagine how it might have been otherwise. It is very hard to slip mentally into a world in which people would not think by slipping mentally into other worlds--very hard to make a counterfactual world in which counterfactuals were not a key ingredient of thought.

Another quick example: I was having a conversation with someone who told me he came from Whiting, Indiana. Since I didn't know where that was, he explained, "Whiting is very near Chicago--in fact, it would be in Illinois if it weren't for the state line." Like the earlier one, this remark was dropped casually; it was certainly not an effort to be witty. He didn't chuckle, nor did I. I simply flashed a quick smile, 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. *Of course* Whiting would be in Illinois if the Illinois state line made it be so--but if *that's* all he meant, it is an empty remark, because it holds just as well for cities thousands of miles from Chicago. But clearly, the notion he had in mind was that there is an accidental quality to where boundary lines fall, a notion that there are counterfactual worlds "close" to ours, worlds in which the Illinois-Indiana line had gotten placed a couple of miles further east, and so on. And his remark tacitly assumed that he and I shared such intuitions about the impermanence and arbitrariness of geographical boundary lines, intuitions about how these lines could "slip".

Remarks like this betray the hidden "fault lines of the mind"; they show which things are solid and which things can slip. 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 in a nice, neat little control panel, forevermore 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 new knobs on concepts where others thought there were none, just as some people get to be good at perceiving mushrooms in a forest where others see none, even when they stare mightily.

=====

There is a curious book called *One Book Five Ways*, published in 1978 by William Kaufmann, Inc. It came about this way. As an educational experiment in comparative publishing procedures, a manuscript on indoor gardening was sent around to five different university presses, and they 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's more than one way to skin a cat.

Making this book was an extravagant foray into "possible worlds", the kind of thing that seems very hard to do. However as computers become more sophisticated and common, the notion of skinning a cat in nine 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 has become, instead, 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). We should use the power of computers to aid us in seeing the full concept--implicit "sphere of hypothetical variations"--surrounding any static, frozen perception.

I have concocted a playful 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 wish to enlist computers as our partners in this venture of inventing variations on a theme, which is to say, turning implicospheres into "explicospheres", we have to give them the ability to spot knobs themselves, not just to accept knobs that we humans 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, this induces modes of slippage in the whole. Generally, letting a constituent concept slip in its simplest ways is enough, since when more than one of these is done at a time, that can already create many unexpected effects. Gradually, as the space of possibilities of the new concept--the implicosphere--is traced out, the most common and useful of those slippages become more closely and directly associated with the new concept itself, rather than having to be derived over and over from its constituents. This way, the new concept's implicosphere becomes more and more explicitly explored, and eventually the new concept becomes old and reaches the point where it too can be used as a constituent of fresh new young concepts.

Some examples of this sort of thing were presented in my column for September, 1981 (Chapter 24). Now although September is almost October and 1981 is almost 1982, that doesn't quite mean that you have those examples at your mind's fingertips, or on the tip of your mind's tongue. So let me present 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 that of 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 September 1981 issue of *Scientific American*.

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

You are almost reading *Godel, Escher, Bach*.

You are almost reading a letter from me.

You are almost writing this column.

You are almost hearing my voice.

I am almost talking to you.

By now, the original concept is almost lost in a silly sea of "almost" variations--but it has been enriched by this 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 entity. 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's what lets you know that it has been firmly planted in the soil of your mind. It is not whether that 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 *Encyclopaedia Britannica's* first edition (1768-71) consisted of three volumes: Volume I: "A-B"; Volume II: "C-L," and then Volume III: the rest of the alphabet. In that edition, 511 pages were devoted to topics beginning with 'A', while the last volume had 753 pages altogether! (I guess that in those days there weren't yet many interesting things around that began with letters between 'M' and 'Z'.) Hearing this amusing fact instantaneously triggered the retrieval of the memory, implanted in me years and 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 playing along and all of a sudden the recording engineer would notice that there wasn't much room left on the plate, so the performers would be given a signal to hurry up, and as a result, the tempo would be faster and faster the further toward the center the needle came. I think it is obvious why the one triggered retrieval of the other. And yet--is it obvious?

On the surface, these two concepts are completely unrelated. One concerns printed matter, books, the alphabet, and so on while the other concerns plastic disks, sounds, performers, recording techniques, and so on. However, at some deeper conceptual level, 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's certainly not just one word. It will take you a while. And when you do come up with a phrase, 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 that I am unaware of, not yet having twiddled enough knobs on that concept. I don't yet 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 proven by the fact that it acts as a point of immediate reference; that my memory mechanisms are capable of using it as an "address" (a key for retrieval) under the proper circumstances. *The vast majority of our concepts are wordless* in this way, although we can certainly make stabs at verbalizing them when we need to.

=====  
Early in this column, I stated a thesis: that the crux of creativity resides in the ability to manufacture variations on a theme. I hope now to have sufficiently fleshed out this thesis that you 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 to the invention of the theme itself? This leads one back to that seductive notion that Einstein and other geniuses are "cut from a different cloth" from ordinary mortals, or at least that certain cognitive acts done by them involve principles that transcend the everyday ones. This is something I do not believe at all. If you look at the history of science, for instance, you will see that every idea is built upon a thousand related ideas. Careful analysis leads one to see that what we choose to call a new theme is itself always some sort of variation, on a deep level, of previous themes. The trick is to be able to see the deeply hidden knobs!

Newton said that if he had seen further than others, it was only by standing on the shoulders of giants. Too often, however, we simply indulge in wishful thinking when we imagine that the genesis of 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 were necessarily shallow and mundane.

My own mental image of the creative process involves viewing the organization of a mind as consisting of thousands, perhaps millions, of overlapping and intermingling implicospheres, at the center of each of which is a conceptual skeleton. The implicosphere is a flickering, ephemeral thing, a bit like a swarm of gnats around a gas-station light on a hot summer's night, perhaps more like an electron cloud, with its quantum-mechanical elusiveness, about a nucleus, blurring out and dying off the further removed from the core it is. If you have studied quantum chemistry, you know that the fluid nature of chemical bonds can best be understood as a direct 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 that allow 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".

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. This view emphasizes the coming together of two concepts, while bypassing discussion of the internal structure of a *single* concept. In Koestler's view, something new can happen when two concepts "collide" and fuse---something not present in the concepts themselves. This is in keeping with Koestler's philosophy that wholes are somehow greater than the sum of their parts.

↳ SYSTEMS = Koestler

By contrast, I have been emphasizing the idea of the internal structure of *one* concept. In my view, the way that concepts can bond together and form conceptual molecules on all levels of complexity is a consequence of their internal structure. What results from a bond may surprise us, but it will nonetheless always have been completely determined by the concepts involved in the fusion, if only we could understand how they are structured. Thus the crux of the matter is the internal structure of a single concept and how it "reaches out" toward things it is not. The crux is not some magical, mysterious process that occurs when two indivisible concepts collide; it is a consequence of the divisibility of concepts into subconceptual elements. As must be clear from this, I am not one to believe that wholes elude description in terms of their parts. I believe that if we come to understand the "physics of concepts", then perhaps we can derive from it a

Reductionist = Hofstadter

"chemistry of creativity", just as we can derive the principles of the chemistry of atoms and molecules from those of the physics of quanta and particles. But as I said earlier, it is not just around the corner.

Recently I happened to read a headline on the cover of a popular electronics magazine that blared something about "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?"

---

### *Post Scriptum*

*knobs, knobs, everywhere--Just vary a knob to think.*

Some readers objected to the slogan of this column--that making variations on a theme is the crux of creativity. They felt--and quite rightly--that making variations (i.e., twisting knobs) is as easy as falling off a log. So how can genius be that easy? Part of the answer is: For a genius, it is easy to be a genius. Not being a genius would be excruciatingly hard for a genius. However, this isn't a completely satisfactory answer for people who pose this objection. They feel that I am unwittingly implying that it is easy for *anybody* to be a genius: after all, a crank can crank a knob as deftly as a genius can. The crux of their objection, then, is that the crux of creativity is not in *twiddling* knobs, but in *spotting* them!

Well, that is exactly what I meant by my slogan. Making variations is not just twiddling a knob before you; part of the act is to manufacture the knob yourself. Where does a knob come from? The question amounts to asking: How do you see a *variable* where there is actually a *constant*? More specifically: what might vary, and how, might it vary? It's not enough to just have the desire to see something different from what is there before you. Often the dullest knobs are a result of someone's straining to be original, and coming up with something weak and ineffective. So where do good knobs come from? I would say they come from *seeing one thing as something else*. Once an abstract connection is set up via some sort of *analogy* or *reminding-incident*, then the gate opens wide for ideas to slosh back and forth between the two concepts.

*atom - solar system...*

A simple example: A friend and I noticed a fuel-delivery truck pulling into a driveway, and on it was very conspicuously printed "NSF", standing for "North Shore Fuel". However, to us those letters meant "National Science Foundation" as surely as "TNT" means "trinitrotoluene" to Eve Rybody. Now, we could have just let the coincidence go, but instead we played with it. We envisioned a National Science Foundation truck pulling up to a research institute. The driver gets out of the cab, drags a thick flexible hose over to a hole in the wall of a building and inserts it, then starts up a loud motor, and pumps a truckload of money--presumably in large bills--into the cellar of the building. (Wouldn't it be nice if grants were delivered that way?) This vision then led us to pondering the way that money actually *does* flow between large institutions: usually as abstract, intangible numbers shot down wires as binary digits, rather than as greenbacks hauled about in large trucks.

This very small incident serves well to illustrate how a simple reminding-incident triggered a series of thoughts that wound up in a region of idea-space that would have been totally unanticipable moments before. All that was needed was for an inappropriate meaning of "NSF" to come to mind, and then to be explored a bit. Such opportunities for being reminded of something remote--such *double-entendre* situation--occur all the time, but often they go unobserved. Sometimes the ambiguity is observed but shrugged off with disinterest. Sometimes it is exploited to the hilt. In this example, the result was not earthshaking, but it did cast things in a new light for both of us, and the image amused us quite a bit. And this way of exploiting serendipity--that is, exploiting coincidences and unexpected perceived similarities--is typical of what I consider the crux of the creative process.

=====

Serendipitous observation and quick exploration of potential are vital elements in the making of a knob. What goes hand in hand with the willingness to playfully explore a serendipitous connection is the willingness to censor or curtail an exploration that seems to be leading nowhere. It is the flip side of the risk-taking aspect of serendipity. It's fine to be reminded of something, to see an analogy or a vague connection, and it's fine to try to map one situation or concept onto another in the hopes of making something novel emerge--but you've also got to be willing and able to sense when you've lost the gamble, and to cut your losses. One of the problems with the ever-popular self-help books on how to be creative is that they all encourage "off-the-wall" thinking (under such slogans as "lateral thinking", "conceptual blockbusting",

"getting Whacked on the head", etc.) while glossing over the fact that most off-the-wall connections are of very little worth and that one could waste lifetimes just toying with ideas in that way. One needs something much more reliable than a mere suggestion to "think zany, out-of-the-system thoughts".

romantic striving to be original will usually get you nowhere. Far better to relax and let your perceptual system and your category system work together unconsciously, occasionally coming up with unbidden connections. At that point, you--the lucky owner of the mind in question--can seize the opportunity and follow out the proffered hint. This view of creativity has the conscious mind being quite passive, content to sit back and wait for the unconscious to do its remarkable broodings and brewings.

The most reliable kinds of genuine insight come not from vague reminding experiences (as with the letters "NSF"), but from strong analogies in which one experience can be mapped onto another in a highly pleasing way. The tighter the fit, the deeper the insight, generally speaking. When two things can both be seen as instances of one abstract phenomenon, it is a very exciting discovery. Then ideas about either one can be borrowed in thinking about the other, and that sloshing-about of activity may greatly illumine both at once.

A mapping-recipe that often yields interesting results is *projection of oneself* into a situation: "How would it be *for me*?" This can mean a host of things, depending on how you choose to inject yourself into the scene, which is in turn determined by what grabs your attention. The man who focused in on the bustling activity in the coffeehouse and said, "I'm sure glad I'm not a waitress here tonight!" might instead have been offended by the sounds reaching his ears and said, "If I were the owner here, I'd play less musak"--or he might have zeroed in on someone purchasing a brownie and said, "I wish I were that thin." People are remarkably fluid at seeing themselves in roles that they self-evidently could never fill, and yet the richness of the insights thus elicited is beyond doubt.

=====

When I first heard the French saying *Plus ca change, plus cest la meme chose*, it struck me as annoyingly nonsensical: "The more it changes, the samer it gets" (in my own colloquial translation). I was not amused but nonetheless it stuck in mind for years, and finally it dawned on me that it was full of meanings. My favorite way of interpreting it is this. The more different manifestations you observe of one phenomenon, the more deeply you understand that phenomenon, and therefore the more clearly you can see the vein of sameness running through all those different things. Or put another way, experience with a wide variety of things refines your category system and allows you to make incisive, abstract connections based on deep shared qualities. A more cynical way of putting it, and probably more in line with the intended meaning, would be that superficially different things are often boringly the same. But the saying need not be taken cynically.

Seeing clear to the essence of something unfamiliar is often best achieved by finding one or more known things that you can see it *as*, then being able to balance these views. Physicists have long since learned to juggle two views of light: light as waves, light as particles. They know that each contains a grain of the essence of light, that neither contains it all, and they know when to think of light which way. Don't be fooled by people who knowingly assure you that physicists don't depend on crude images or analogies as crutches, that everything they need is contained in their formulas. The fallacy here is that *which* formula to apply, *how* to apply it, and what parts of it to *neglect* are all aspects not covered in any formula, which is why doing physics is a great art, despite the fact that there are formulas all over the place for Eve Rybody and her brother to use.

Seeing anything as waves suggests immediate knobs: wavelength, frequency, amplitude, speed, medium, and a host of other basic notions that define the essence of undularity. Seeing anything as particles suggests totally different knobs: mass, shape, radius, rotation, constituents, and a host of other basic notions that define the essence of corpuscularity. If you choose to see, say, people as waves or as particles, you may find some of these suggested knobs quite interesting. On the other hand, it may not be fruitful to do so. Good analogies usually are not the product of an off-the-wall suggestion like this, but spring to mind unbidden, from the deep similarity-searching wells of the unconscious.

Once you have decided to try out a new way of viewing a phenomenon, you can let that view suggest a set of knobs to vary. The act of varying them will lead you down new pathways, generating new images ripe for perception in their own right. This sets up a closed loop:

- \* fresh situations get unconsciously framed in terms of familiar concepts;
- \* those familiar concepts come equipped with standard knobs to twiddle;
- \* twiddling those knobs carries you into fresh new conceptual territory.

A visual image that I always find coming back in this context is that of a planet orbiting a star, and whose orbit brings it so close to another star that it gets "captured" and begins orbiting the second star. As it swings around the new star, perhaps it finds itself coming very close to yet another star, and ficklely changes allegiance. And thus it do-si-do's its way around the universe.

The mental analogue of such stellar peregrinations is what the loop above attempts to convey. You can think of concepts as stars, and knob-twiddling as carrying you from one point on an orbit to another point. If you twiddle enough, you may well find yourself deep within the attractive zone of an unexpected but interesting concept and be captured by it. You may thus migrate from concept to concept. In short, knob-twiddling is a device that carries you from one concept to another, taking advantage of their overlapping orbits.

The cycle shown above spells out what I intend by the phrase "making variations on a theme", and it is this loop that I am suggesting is the crux of creativity. The beauty of it is that you let your memory and perceptual mechanisms do all the *hard* work for you (pulling concepts from dormancy); *all you* do is twiddle knobs. And I'll let you decide what this odd distinction is between something called "you" and the hard-working mechanisms of "your memory".

=====

The concept of the "implicosphere" of an idea--the sphere of variations on it resulting from the twiddling of many knobs a "reasonable" amount is a difficult one, but it is absolutely central to the meaning of this column. One way of thinking about it is this. Imagine a single gnat attracted by a bright light. It will buzz about, tracing out a three-dimensional random walk centered on that light. If you keep a photographic plate exposed so that you can record its path cumulatively, you will first see a chaotic broken line, but soon the image will get so dense with crisscrossing lines that it will gradually turn into a circular smear of slowly increasing radius. At the outer edges of the smear you might once in a while make out an occasional foray of the lone bug. For a while, the territory covered expands, but eventually this gnat-o-sphere will reach stable size. Its silhouette, instead of being a sharp-edged circle, will be a blurry circle whose approximate radius reveals something about how gnats are attracted by lights.

Now if you simply think of this translated into idea-space, You have roughly the right image. Of course, not all implicospheres have the same radius. Some people's implicospheres tend to have bigger radii than other people's do, and consequently their implicospheres overlap more. This can be good but it can be overdone. Too much overlap and all you have is a mush of vaguely associated ideas, an overdone and tasteless mental goulash. Too little overlap and you have a very thin, watery mind, one with few big surprises (except for the meta-level surprise of having so few surprises). There is, in other words, an optimum amount of overlap for useful creative insight. This is the kind of thing that cannot be taught, however. It would be like trying to train a gnat to control the size of the spheres it traces out. Or if you prefer, it would be like trying to train an entire swarm of gnats to form spheres of a particular size whenever they cluster around lamps. The problem is, it is already preprogrammed in gnats how much they are attracted by lights, by each other, and so on.

When a new idea is implanted in a mind, an implicosphere grows around it. Since this means, in essence, the linking-up of this new idea with older ideas, I call it "diffusion in idea-space". My canonical example of this phenomenon, although it is a rather grim one, has to do with the recent spate of random murders inspired by the spiking of Tylenol capsules with strychnine. It was the Food and Drug Administration's response that so intrigued me, because it implicitly revealed a theory of how this idea would diffuse in the idea-space of a typical potential murderer. The FDA imposed a set of packaging regulations on manufacturers, with various types of products being given various deadlines for compliance. The idea was that your potential murderer could slip from the idea of Tylenol to that of aspirin in a week's time, but it would take the expanding sphere longer to hit the brilliant idea that it could be just any over-the-counter drug. Not just the FDA seemed to think this way; also radio talk-show hosts seemed to love speculating about what drug might be chosen next--but I never heard them worrying about ordinary food in grocery stores. Yet why should it give a stochastic killer any less joy to kill by spiking a jar of mustard than by spiking a drug? In fact, if your goal in life is to see masses of random people die, there are all sorts of routes you can take that don't involve ingestion at all. A friend of mine took a train from Washington to New York and enroute her train smashed into a washing machine full of rocks that had been placed on the

tracks by some do-badder. Was this part of the Tylenol-murders implicosphere in the mind of the person who did it? I doubt it, but it is possible.

In its own gruesome way, the generalization of the Tylenol murders resembles that of the expanding implicosphere of the be--and that of any idea that arises. Ideas, whether evil or beneficial, have their own dynamics of spreading in and long minds.

=====

Slippage of thought is a remarkably invisible phenomenon, given its ubiquity. People simply don't recognize how curiously selective they are in their "choice" of what is and what is not a hinge point in how they think of an event. It all seems so natural as to require no explanation.

I dropped a slice of pizza on the floor of a pizza place the other evening. My friend Don, who was less hungry than I was, immediately sympathized saying, "Too bad I didn't drop one of my pieces--or that you didn't drop one of mine instead of one of yours." Sounds sensible. But why didn't he say, "Too bad the pizza isn't larger"? His choice revealed that to his unconscious mind, it seemed sensible to switch the role-filler in a given event, as if to imply that a pizza-slice-droppage had been in the cards for that evening, that God had flipped a coin and, unluckily for me, it had come out with me as the dropper instead of Don--but that it might have come out the other way around.

Some hypothetical replacement scenarios--I like to call them "subjunctive instant replays"--are compelling, and come to mind by reflex. They are not idle musings but very natural human emotional responses to a common type of occurrence. Other subjunctive instant replays have little intuitive appeal and seem far-fetched, although it hard to say just why. Consider the following list:

Too bad they didn't give us a replacement piece.

Lucky we weren't in a really fancy restaurant.

Too bad gravity isn't weaker, so that you could have caught it before it hit the ground.

Lucky it wasn't a beaker filled with poison.

Too bad it wasn't a fork.

Lucky it wasn't a piece of good china.

Too bad eating off floors isn't hygienic.

Lucky you didn't drop the whole pizza.

Too bad it wasn't the people at the next table who dropped *their* pizza.

Lucky there was no carpet in here.

Too bad you were the hungry one, rather than me.

I'll leave it to you to generate other subjunctive instant replays that he might have come up with. There is a rough rank ordering to them, in terms of plausibility of springing to mind. It's the rhyme and reason behind that ordering that fascinates me.

Why do people find it not only plausible but even compelling to make remarks like the following?

If Jesse Jackson were a white man, he'd be elected President.

If Jesse Jackson were a white man, he'd be running for dogcatcher.

These two sentences came from random voters, as quoted in *Newsweek*. I wonder what slips in people's minds when they imagine a white Jesse Jackson. Do they envision a preacher in a Baptist church? Is this person an ardent fighter for civil rights? Or, conversely, an ardent fighter against the quota system? Similarly, what does a high-school boy mean when he says, "If I were my father, I wouldn't lend me the car"? Does he ever notice that if he were his father, he would *ipso facto* be his own son? Or need that be so? Would the two have exchanged roles? The point is, there are a host of questions left completely open here, yet no one balks for a second at such counterfactuals. In fact, they are common currency, they are daily bread, they are the meat and potatoes of communication. But some types of counterfactuals never (or hardly ever) come up, while others, equally reality violating, are a dime a dozen.

Daniel Kahneman and Amos Tversky, cognitive psychologists, have made studies of how much emotion people generate upon reading stories of just-missed airplanes or just-caught airplanes--especially ones that crash. These kinds of near misses, whether fortunate or unfortunate, tug at our hearts and do so in nearly universal ways. Something about these slippability examples is truly at the core of what it is to be human and to experience the world through the filter of the human mind.

We are left with some interesting questions:

Which counterfactuals are likely to be triggered in a human mind by various types of events in the world?

Why are some events perceived to be "near misses", while others are not?

Why are some deaths of innocent people viewed as more tragic than other deaths of innocent people?

At such points where deep human emotion, identification with other beings, and perception of reality meet lies the crux of creativity--and also the crux of the most mundane thoughts. Spinning out variations is what comes naturally to the human mind, and is it ever fertile!