24

Analogies and Roles in Human and Machine Thinking

September, 1981

IN our research in artificial intelligence, my graduate students Gray Clossman and Marsha Meredith and I have been looking at typical human thought processes in everyday life as well as in more limited domains, and everywhere we look, we seem to find that within the internal representations of concepts there are substructures that have a kind of independence of the structures of which they are part. Such a substructure is *modular*—exportable from its native context to alien contexts. It is an autonomous structure in its own right, and we call these modules *roles.* A role, then, is a natural "module of description" of something, a sort of bite-sized chunk that seems to be comfortable moving out of its first home and finding homes in other places, some of them unlikely at first glance.

One intriguing example is the "First Lady" role. Probably most Americans use this term more flexibly than they realize. They would most likely say, if asked, that the term means "the wife of the president", and not think any more about it. But if they were asked about the First Lady of Canada, what would almost surely pop into their mind is the name or image of Margaret Trudeau. They might reject the thought as soon as it occurred to them, but for us the important thing is that the thought of her would arise at all. First of all, people know her as the *former* wife of Pierre Elliott Trudeau. Second, Trudeau is not the president of Canada but its prime minister. How, then is "former wife of the prime minister" the same as "wife of the president"?

Before you answer, "Well, 'wife' and 'former wife' are related concepts, as are 'prime minister' and 'president'", consider who might be said to be the current First Lady of Britain. Whose name comes to your mind? Margaret Thatcher? Queen Elizabeth? They are women, but do they really play the role of First Lady? How about Denis Thatcher or Prince Philip? At first these suggestions seem silly, but in a strange way they start to seem compelling, particularly the thought of Denis Thatcher. In fact, I once clipped a newspaper article that portrayed Denis Thatcher as Britain's First Lady.

What kind of sense does this make? How can a male be a lady? Well, language is far slipperier than dictionary definitions would have you believe. Its slipperiness comes from the underlying slipperiness of concepts, in particular these elusive things we are calling roles.

Of course, you could argue that what "First Lady" *really* means is "spouse of the head of state", and so the First Lady role goes over without any trouble into "husband of the prime minister". But this won't do either. In Haiti, until recently the title of First Lady belonged to Simone Duvalier, the wife of the late former president, François ("Papa Doc") Duvalier. She is also the mother of the current president, Jean-Claude ("Baby Doc") Duvalier. Not long ago there was a bitter power struggle between Simone Duvalier and her daughter-in-law Michelle Bennett Duvalier, the wife of Baby Doc, for the title of First Lady. In the end, the younger woman apparently gained the upper hand, taking the title "First Lady of the Republic" away from her mother-in-law, who in compensation was given the lifetime title "First Lady of the Revolution".

Do you want to amend your suggestion so that it will say "spouse or parent, present or former, of a head of state, present or former"? You know perfectly well that we'll be able to come up with other exceptions. For example, imagine a meeting of the Pooh-Bah Club at which the Grand Pooh-Bah's favorite aunt was introduced as the First Lady of the club. Of course, the Grand Pooh-Bah is hardly a head of state, and so you could amend your definition to say "spouse or favorite relative, present or former, of the head, present or former, of any old organization". But suppose . . . Actually, I think I'll let *you* go on inventing exceptional cases. For any rule you propose, there is bound to be some conceivable way to get around it.

Worse yet, something terrible is happening to the concept as it gets more flexible. Something crucial is gradually getting buried, namely the notion that "wife of the president" is the most *natural* meaning, at least for Americans in this day and age. If you were told only the generalized definition, a gigantic paragraph in legalese, full of subordinate clauses, parenthetical remarks, and strings of *or*'s—the end product of all these bizarre cases—you would be perfectly justified in concluding that Sam Pfeffenhauser, the former father-in-law of the corner drugstore's temporary manager, is just as good an example of the First Lady concept as Nancy Reagan is. When this happens, something is wrong. The definition not only should be general, but also should incorporate some indication of what the *spirit* of the idea is.

* * *

Computers have a hard time getting the spirit of things; they prefer to know things to the letter. And so people spend an enormous amount of time talking to computers, writing long and detailed descriptions of ideas they could get across in *one good example* to anyone with half a brain. So a challenging question is how to get a computer to understand what is meant by "First Lady". For this we need to examine the idea of "roles" in detail.

In order to illustrate how the notion of "role" can be modeled in domains more formal than that of political protocol, I now will switch to one of my favorite domains: the natural numbers. I will present some puzzles that Gray, Marsha, and I have been thinking about. Each of them has a set of possible answers with varying degrees of plausibility or defensibility. We are working on a computer program that is able to see the rationale behind each possible answer, and thus is able to come up with the same set of "feelings" as a typical person would have, about what is a good answer and what is a bad one.

The domain of natural numbers might sound at first like a hard-edged, objective mathematical world, but actually it is a domain in which problems requiring extremely subtle *subjective* judgments can be formulated. We have given our program very little detailed arithmetical knowledge about the integers. The program does not, for example, recognize 9 as a square; in fact, it doesn't even know about multiplication! It does not know that 6 is even and 7 is odd. So what *does* it know? It knows how to count up or down —that is, it has a knowledge of successorship and predecessorship. Thus it recognizes that the sequence of numerals "12345" represents an upward counting process. It is also able to apply the notion of counting to structures it is looking at, as in "44444", which it could recognize as a group of five copies of the numeral '4'. It knows that 9 is bigger than 4, although it has no idea *how much* bigger. (Subtraction and other arithmetical operations are unknown to it.) You can think of our computer program as having the arithmetical sophistication of a five-year-old and an avid curiosity about number patterns. (By the way, it is not tied to or affected by decimal notation. The number 10 is not considered any more special than the number 9.)

Here is the first problem (invented, as were many of the following ones, by Gray). Consider the following structure, which we'll call A:

A: 1 2 3 4 5 5 4 3 2 1

Now consider the structure called B:

B: 1 2 3 4 4 3 2 1

The question is: *What is to B as 4 is to A?* Or, to use the language of roles: *What plays the role in B that 4 plays in A?*

Note that by asking it this way, we leave it to the puzzle solver to decide what role 4 actually does play in A. It would be analogous to asking "Who is the Nancy Reagan of Britain?", leaving it to the listener to figure out what conceptual role Nancy Reagan fills, and then to try to export that role to Britain. I have found that many people who balk at calling Denis Thatcher the "First Lady of Britain" are quite content with calling him the "Nancy Reagan of Britain". A curious point that this illustrates, and to which we will return, is this: If the role is left implicit, nonverbalized, it has more fluidity in the way it transfers than if it is "frozen" in an English phrase.

As a matter of fact, most analogies crop up in this type of nonverbal way. Seldom does someone say to you explicitly: "What is the counterpark of Central Park in San Francisco?" Usually it happens through a more implicit channel. When you are visiting San Francisco for the first time, you are driven through Golden Gate Park, and somehow it reminds you of Central Park. After the fact, you can point out some shared features: both are long thin rectangles; both contain lakes, curving roads, and excellent museums; and so on. Most analogies arise similarly—as a result of unconscious filterings and arrangings of perceptions, rather than as consciously sought solutions to cooked-up puzzles. To put it another way, to be *reminded* of something is to have unconsciously formulated an analogy.

Incidentally, when I first thought of writing about roles and analogies, I had in mind both the First Lady example and the numerical examples. As my thoughts evolved, I realized I was unconsciously developing a parallel in my mind between the First Lady example and the numerical examples. I'll call it a "meta-analogy", since it is an analogy between analogies. In this meta-analogy, I see structure A as corresponding to the United States, structure B to Britain, 4 to Nancy Reagan and the unknown number to the unknown person. We'll come back to the meta-analogy later on.

* * *

Let us now look at some possible answers to the first number-analogy problem. The most sensible answer is 3—and fortunately, it is also the most frequently given one. The usual justification is that 4 precedes the central pair (55) in A, and the corresponding central pair in B is 44, which is preceded by 3. Well, then, what would you say for C? What is to C as 4 is to A?

C: 1 2 3 4 5 6 6 6 6 5 4 3 2 1

The central pair in C is 66, which is flanked by 6's. Is 6, therefore, to C what 4 is to A? Well, most people probably would prefer 5, although it is perfectly *logical* to insist on 6. The preference for 5 comes, nonetheless, from a very sensible (and also logical) instinct to generalize the notion of "central pair" (itself, to be sure, a role) to "central plateau" (or whatever you want to call it). There are competing urges: first, to stay with the exact original concept, and second, to flex and bend when it "feels right", when it would seem rigid and stodgy to insist on established conventions over simple and "natural" extensions. But it is just these sorts of terms—"flex", "bend", "feels right", "rigid", "natural", and so on—that are so extraordinarily hard to put into programs, logical though programming might be.

Now let us investigate some other ways to make the role of 4 slip. Consider this structure:

D: 1 1 2 2 3 3 4 4 5 4 4 3 3 2 2 1 1

Here is a curious kind of reversal; now there is no central pair—yet everything else is in pairs. Some people might still pick 4, since it is next to the center. But what about 44, a *pair* rather than a single number? After all, as long as "pair" and "singleton" have switched places, we might as well go all the way and give an answer that reflects this perceptual turnabout. In fact, it would seem rigid and unimaginative to insist on sticking with single numbers when it is so obvious that the easiest way to perceive D is in terms of pairs:

1-1 2-2 3-3 4-4 5 4-4 3-3 2-2 1-1

Not just 4 but every part of A has a role, and there are corresponding roles in D. As you can see, within each role the concepts of pair and singleton have been switched.

Now is as good a time as any to return to my meta-analogy and to point out some correspondences between these problems and the First Lady problem. If you think of the president as "the highest, most central figure in the land" and his wife as "the one standing next to him", you will see that this characterization carries over almost literally to the numerical problems. In structure A, the highest, most central figure—the "president"—is 5 (or possibly the pair of 5's) and his "wife", standing next to him, is 4. In B, the president is 4 (or the pair of 4's) and his wife is 3. In C, the president is 6 (or the group of 6's), and his wife is 5. In D, the president is (for once) unambiguous (5), but to compensate, there is a dilemma concerning the identity of his wife. If you think of pairs as males and singletons as females, then D presents us with a case where the sexes are reversed, exactly as in the First Lady of Britain problem. The most reasonable answer seems to be the "spouse" (in this case, the husband) of 5, namely the pair 44.

Consider now the following couple of curious cases:

E: 1 2 3 4 5 6 7 8

F: 8 7 6 5 4 3 2 1

What can we make of these? A very rigid person might cling to the idea captured in the phrase "number to the left of the central pair", despite the fact that nothing at all distinguishes the central pair in either of these examples. Such a person would give the inane answers of 3 for E, and 6 for F. Such a person would do better to take up football instead of analogies, as Lewis Carroll's Tortoise once remarked to Achilles.

But what would be a wiser view of, say, E? How *should* one map E onto A? Any mapping is doomed to be imperfect, so how can we do it best (that is, with the least pain or frustration)? We might think of E, since it rises uniformly, as mapping onto the left half of A. This would involve a tacit judgment that it is all right to abandon the attempt to map E onto *all* of A, in return for the ease of mapping E onto a "natural" portion of A. That is a pretty subtle step to take, I would say. It would suggest 7 as the answer.

Well, what about F, then? Do we prefer 2 or 7? It depends on whether we choose to map F onto the left half or the right half of A. Mapping F onto A's left half involves mapping a *descending* sequence onto an *ascending* one. But either choice requires a willingness to let go of qualities that had seemed important, a willingness to bend gracefully under pressure. Fluid analogies are not a game for rigid minds!

These kinds of situations are difficult because in essence they call for splitting the role of 4 in A into two rival facets. In the mapping of A onto F, one of the rival facets sees 4's role in A as "one less than the president", whereas the other facet sees 4's role as "the next-to-rightmost element of a staircase". Thus one facet is primarily concerned with *magnitude* and the other primarily with *position.* The facet you find more important will determine your answer to F.

Pretty much this kind of split happened when you tried to decide whether the First Lady of Britain was Queen Elizabeth or Margaret Thatcher—or one of their husbands. Is being a figurehead or being a head of state more likely to make someone's spouse a First Lady? In the United States, these features coincide in one person (the president), but in Britain they do not. Consider the following target structures:

G: 5 4 3 2 1 1 2 3 4 5

H: 1 2 3 4 6 5 5 6 4 3 2 1

In G, what is most central is simultaneously lowest, and what is highest is simultaneously most peripheral! (G can be pictured as a valley and A as a mountain peak.) We have a "ceremonial figure" (the 5's flanking the structure) and we have a "head of state" (the two central 1's). Which one's spouse would better fill the role of First Lady? Or, to put it most simply: "What in G plays the role of 4 in A?" I personally would opt for 2 because it stands next to the central group. To me, centrality seems more important here than magnitude, just as political power seems more substantive than ceremonial show. Correspondingly, I would opt for Denis Thatcher rather than Prince Philip as "Britain's Nancy Reagan".

Now what happens when we tackle H? There are three "reasonable" possibilities (in the sense of appealing to law's proverbial "reasonable human"): 6 (flanking the central pair of 5's), 5 (being the next-to-largest number) and 4 (flanking the central "crater" 6556). Once again there is no Gloriously Right Answer, but there are certainly ideas that seem good and ideas that seem shaky. For instance, if someone suggested, "The answer is 4, because 4 is the fourth term of H, just as it is the fourth term of A", I would be nonplussed. That would be a childish generalization based on the most superficial of scans of both structures. It would be as childish as inferring that, since *our* national holiday falls on the fourth of July, other countries' national holidays would also have to fall on the fourth of various months. To see 4 as no more than the fourth element of A is to ignore all of A's structure. It is to see A as nothing richer than this:

o o o 4 o o o o o o

A good answer must take A's structure into account in a full, rich, and yet simple way. This means that, to the extent it is possible, all of A must be perceived in terms of interacting, mutually intertwining conceptual structures—roles that are mutually dependent, in the way that "family", "husband", "wife", "mother", "father", "daughter", "son", "brother", "sister", "relative", "in-laws", and so on are all interdependent concepts.

The word "role" makes us think of the theater. In a play, the various roles all mingle together in scenes. A scene is a larger-scale structure than an individual role; it is a place where several roles coexist and interact. In our analogy problems, one might try to conceive of the two structures involved as if they were two enactments of a single scene, portrayed by different directors working with different actors. Thus the *core roles* would exist and would be filled in both presentations, but at the same time each presentation would have minor aspects, or roles, unique to it. For example, the adaptation of the Greek legend of Orpheus and Eurydice into a contemporary context of carnival time in Rio de Janeiro is the basis for the movie *Black Orpheus.* Many original features cannot be *directly* exported, but with poetic modification they can be, and the director, Marcel Camus, met the challenge with great flair. In the movie there are, of course, many minor parts—extras—that add Brazilian flavor, yet they do not impair the analogy at all; in fact, they enrich it. This is the kind of thing that appeals deeply to human sensibilities, both intellectually and emotionally.

* * *

Now that you have seen some variations, I would like to return to our first puzzle and point out some of its hidden subtlety. First of all, the "central pair" notion, which functions as the keystone of structure A, is actually just a kind of by-product, an accidental artifact of the structure of A. To see what I mean, consider this question. How would you efficiently describe the structure of A (without quoting it digit by digit)? You would probably say it rises from 1 to 5 and then falls from 5 to 1, making two halves that are mirror images. Nowhere in this description was there any mention of some kind of central pair or central plateau. It was not needed; one will just appear automatically there when anyone follows your description. In fact, anyone who constructs a copy of A is very likely to see a central plateau, even if the concept was never suggested to them. To the mind's eye it appears something like this:

1 2 3 4 **5-5** 4 3 2 1

Somehow a new percept has been born in the center. It is, as I remarked above, the *keystone* of A. (Note that the concept of "the keystone of A" depends on, or implies, a mapping of A onto an arch—yet another analogy.)

Why don't we perceive the pair of 3's, say, as a unit as well? Probably simply because they do not touch. And consider this structure:

1 2 3 4 5 1 2 3 4 5

The central pair—"5 1"—doesn't pop out as being salient or important, does it? In A, though, the combination of *adjacency* and *equality,* particularly when supplemented by *centrality,* somehow makes the two central 5's merge into a unit in the perceiver's mind, albeit usually not at a conscious level. Still, if this perceptual shift did not happen, then the answer of 5 for C, based largely on equating the plateaus in A and C, would be considerably less compelling.

In the first puzzle, both A and B had obvious central plateaus. This suggested a good starting point for an overall mapping of A onto B: central plateau onto central plateau, start onto start, finish onto finish, and so on. But if we tried to complete this mapping, we would obviously run into trouble:

```
1 2 3 4 5 5 4 3 2 1
 \    | |   |    /
  1 2 3 4 4 3 2 1
```

We *must* have 1 in A mapping onto 1 in B, no? And the centers have to match up too, don't they? But where between 1 and 5 does the analogy break down? It seems that some kind of mapping of 4 onto 3, as is shown above, is satisfying to many people. But press them one step more, and they will shrug, grin, and give up.

Similarly, although you can ask for "the Nancy Reagan of Britain", it makes less sense to ask, "Who is the Maureen Reagan of Britain?" (Remember that Maureen Reagan is Nancy Reagan's stepdaughter.) Suppose the Thatchers had a biological daughter. Would she be the counterpart of Maureen Reagan? Or suppose Margaret Thatcher had a stepdaughter. Would she be the counterpart? Then again, suppose that

Margaret Thatcher had no daughter but that Denis Thatcher had *twin* stepdaughters. Would these twins, taken together, constitute the counterpart of Maureen Reagan? (How can two people fill a role defined by one person? Well, think of example D, where the pair of 4's played the role of a single 4 in A. Or think of many European countries, which have both a president and a prime minister.)

Issues like this crop up all the time in the pursuit of good analogies, and facing up to mismatches leads occasionally to productive insights. One could go on and press for even more detailed correspondences between entities in Britain and in the United States. What is the British counterpart of Watergate? Who plays the part of Richard Nixon? Of John Mitchell? Of Senator Sam Ervin? Of Senator Daniel Inouye? Of G. Gordon Liddy? Of Judge John Sirica? Of John Dean? Of Officer Ulasewicz? Of Alexander Butterfield? The less salient an object is inside a larger structure, the harder it is to characterize in an exportable way.

But what makes something salient? As a rule, it is its proximity, in some sense, to a "distinguished" element of the larger structure. Consider the following long structure:

1 1 1 1 1 1 1 1 1 2 2 2 2 3 3 3 4 3 3 3 2 2 2 2 1 1 1 1 1 1 1 1 1

The central 4 is probably the most distinguished individual numeral. Then, depending on how you perceive the sequence, different features will leap out at you. For instance, do you see it as "letters" or as "words" (larger-scale chunks of the sequence)? When I see it at the "word" level, the central group "3334333" seems just a shade less salient than the 4, and after that, perhaps, the two flanking groups of 1's. The two groups of 3's by themselves come next. Only then do the groups of 2's get recognized. On the other hand, when I perceive the sequence at the "letter" level, what is salient is quite different. After the central 4, probably the next most salient numbers to me are the first and last 1's, since they are very easy to describe—then maybe the first and last 2's. After that, the two 3's flanking the central 4—but at this point it starts to get a little harder to specify various items without resorting to such uninspired descriptions as "the fourth term".

A *distinguished* item is something we can get at via an elegant, crisp, exportable-sounding description. A *nearly distinguished* item is something we can get at by first pointing to a distinguished item, and then, in an exportable way, describing a short "jog" that leads to it. Just as in giving someone directions, some places are more salient, others are less so. Some buildings in New York City are inherently difficult to direct someone to, others are inherently easy. In the same way, some roles in a complex conceptual structure are highly distinguished and easily exportable, others are very hard to describe. Although they may have certain idiosyncratic qualities in their local context, nothing makes them stand out globally.

As you move progressively farther away from its central roles, any analogy becomes increasingly strained. For example, "Who is the Jackie Washington of Britain?" Should we begin by getting out the London telephone book and looking under "Washington, J."? Or should we look under "London, J."? Or is it a meaningless question, meaningless even to Jackie's best friend? After all, Jackie's role may just be too small and idiosyncratic within the structure of the United States. It is not exportable. The fact that Jackie is the manager of Gearloose's Hot Dog Stand in Duckburg does not help much, because one still has to figure out the identities of the British Duckburg and the British Gearloose—not to mention the British equivalent of hot dog stands!

The moral is a simple one: Don't press an analogy too far, because it will always break down. In that case, what good are analogies? Why bother with them? What is the purpose of trying to establish a mapping between two things that *do not* map onto each other in reality? The answer is surely very complex, but the heart of it must be that it is good for our survival (or our genes' survival), because we do it all the time. Analogy and reminding, whether they are accurate or not, guide all our thought patterns. Being attuned to vague resemblances is the hallmark of intelligence, for better or for worse.

* * *

The fact that we use words and ready-made phrases shows that we funnel the world down into a fairly constant set of categories. Often we end up with one word, such as "kitchen". In general, two kitchens will not map onto each other exactly, but we still are satisfied with the abstraction "kitchen". Generally speaking, a kitchen will have a sink, a stove, a refrigerator, cupboards, counters, drawers, and so on. In the United States, it is very common for people to assume that the garbage will be in a cupboard below the sink. The idea of "the cupboard below the sink" is a perfect example of an exportable role. In fact, isn't your sink the "president" of your kitchen? And . . .

Our language provides for mappings of many degrees of accuracy. Some people, when they see Bossie, see no further than "cow" and accordingly use that word; others notice that Bossie is female, and will say "heifer". Still others perceive the breed as easily as they perceive Bossie's "cowness" and talk about "that Angus heifer". A famous Dublin zookeeper, Mr. Flood, was once asked the secret of his great success in breeding lion cubs. "Understanding lions." said he. "And in what does understanding lions consist?" he was asked. His reply: "Every lion is different." This curious answer denies the category while taking advantage of it. But that is the nature of categories. Their validity can at best be partial. No matter at what level of detail you cut off your scrutiny, your perception amounts to filtering out some aspects and funneling the remainder into a single conceptual target, a mental symbol often labeled with just one word (such as "word") or stock phrase (such as "stock phrase"). Each such mental symbol implicitly stands for the elusive sameness shared by all the things it denotes.

Beyond the implicit analogies hidden in individual words or stock phrases, explicit analogies occur all the time on a larger scale in our sentences. We are quite uninhibited in comparing unfamiliar things with things we assume are more familiar. We see grids of all kinds as being similar to checkerboards. We see carefully charted actions in life as being similar to chess moves. We see the eye as a camera, the atom as a tiny solar system. Science is constantly being likened to a vast jigsaw puzzle (an analogy I have never cared for). In their eagerness to stretch and bend concepts, people turn proper nouns into common nouns, as in the statement "Brigitte Bardot is the French Marilyn Monroe." In such linguistic flexing, both *la Bardot* and the Monroe suffer somewhat in the interests of vivid imagery.

Then, going one step beyond the explicit linguistic level, there are the analogies and mappings that we use constantly to guide our thoughts on a larger scale. The perception of romantic dilemmas is one of the most striking places where mapping or analogical thinking dominates in an obvious way. When someone tells us of some romantic woe, we can usually map it immediately onto some experience of our own. In fact, we can probably draw some parallel between *any* romantic situation and any other one, and such a mapping will perhaps yield some insight if it is carried out well. Yet romances are incredibly detailed and idiosyncratic. The point is that we throw many details away; we skim off some abstractions and are careful not to try to carry the resemblance too far. And certainly we ignore the trivial aspects. A romance between Chris and Sandy can certainly map onto one between Pat and Chris or one between Sandy and Pat, despite the fact that names, hair colors, and other superficial features do not match!

The reason, then, for worrying about human analogical thought is that *it is there.* To ignore it would be like ignoring Everest in trying to understand mountain climbing.

* * *

Let us get back to some concrete problems in our more formal, numerical domain. Notice that there is an inherent kind of contradiction in setting up analogy problems—which, after all, are *informal* by definition—in a rather formal domain. But the nice thing is that it shows that the domain is actually just as slippery as any "informal" domain. Here are four further examples:

I: 1 2 3 3 4 5 6 7 6 5 4 3 3 2 1

J: 1 7 7 6 5 4 3 2 1

K: 6 9 7 3 9 4 1 6 6

L: 1 2 3 4 5 6 7 8 9 7 8 9 6 5 4 3 2 1

Example I involves what I enjoy referring to as a "governor", namely the pair 33. Here again, one role in A has been split into parts: 55 in A was not only the sole pair but also the *peak*, whereas in I, 33 is the sole pair and 7 plays the role of the peak. We are forced to choose between 2 (the wife of the governor) and 6 (the wife of the president). Actually, the governor has two "wives"—2 and 4, a "left wife" and a "right wife"—and so we have to choose between them, unless we go with 6 as being the wife of president 7.

Example J, beginning as it does with "1776", is a patriotic puzzle. (What is its British counterpart?) Its interest is primarily in that it draws attention to A's symmetry, which we had taken for granted. When we chose 4 as the president's wife, were we taking his *left* wife or his *right* wife? In A, of course, they coincided, so it didn't matter. But in J they differ: 1 would be the left wife, 6 would be the right wife. Because of a tendency to be influenced by our left-to-right scanning, we probably would choose the left wife under normal circumstances, but here, there is such great asymmetry that we pause. The regular descent from 7 to 1 corresponds far better to A's "staircase" structures than does the abrupt leap upward (from 1 to 7 in one step!). For that reason, 6 probably wins over 1, in this case.

Example K is a bit obscure, but it has been led up to by example J. In particular, example J drew attention to the fact that in A there are two 4's, not just one. Example K plays on the relation of those two 4's to each other. In A, there were two elements between the two 4's. We can take that property as defining the role of 4 in A. To be sure, that is not the *only* relation between the two 4's, but it is the most obvious. If you "turn off" everything in A but the 4's, you will get an image something like this: ooo4oo4ooo. That image makes the size of the interval between the 4's stand out. Given this way of looking at the role of A's two 4's, what in K corresponds? There is only one number that occurs exactly twice, and its two appearances are separated by two numbers. That number is 9. If you turn off everything but the 9's in K, you get this picture: o9oo9oooo. That may or may not be sufficient reason for you to choose 9 as the K-counterpart of A's 4.

Finally, consider example L. Here, the central-pair notion gets extended one further degree of abstraction. We go up, step by step, until we hit the second 7. Jolt! It takes us a moment to get our bearings, and when we recover, we realize that the central pair consists not of single integers but of "clumps" or "chunks": namely, two copies of the unit 789. We can aid the eye this way:

1 2 3 4 5 6 7-8-9 7-8-9 6 5 4 3 2 1

Now the answer seems glaringly obvious: It is 6! On the other hand, maybe we were supposed to get the hint offered us generously by the central pair. And what was that hint? It is that we could perceive the *whole structure* in triples, not just its center. In this case, L reparses into

123 456 789 789 654 321

Now the answer should be obvious—except that we are still left with a minor dilemma. Do we take the president's right-hand wife (654) or his left-hand wife (456)? I am biased by left-to-right chauvinism and would choose 456. Many people, however, refuse to see the sequence in triples and stick with 6.

Here is an innocent-seeming puzzle that points to still more complex issues:

M: 1 2 3 4 5 7 7 5 4 3 2 1

The way I see it, the best answer is 6. You might object, "Why not 5? 6 isn't even there!" True, but 6 is conspicuous by its absence. The 4 in A precedes the 5 not only typographically but also arithmetically: 4 is the numerical predecessor of 5. And what is 5 in A? It could be seen either as the *maximum* in A or as the number forming the *central pair* of A. Both carry over to M, yielding 7 as M's 5. Now, if you choose to see 4's rỏle in A abstractly and arithmetically rather than concretely and typographically, you can carry your vision directly over to M. Then candidate 6 must be considered a strong competitor to 5. In my mind, it wins.

This example opens up an whole new world of *levels of abstraction* in the perception of structures. To illustrate briefly, let me propose the following structures:

A′: 1 2 3 4 4 4 5 6 7 8 9 8 7 6 5 4 4 4 4 3 2 1

B′: 1 1 1 2 3 4 3 2 1 1

And here is the puzzle: What in B′ plays the role that 7 plays in A′? Well, 7 occurs twice in A′, but certainly it seems to play no *salient* role. As a numeral in A′, 7 has no outstanding characteristic, and so at first its role seems hard to export. However, 7 enters into the structure of A′ in another way, and a salient one at that. One of the most salient features of A′ is its large number of 4's. Count them. How many? Seven. Aha! Thus 7, in its capacity as an invisible *counting number* rather than as a visible *numeral,* plays a very distinguished role in structure A′. Still, is it possible to *export* this role to B′?

We have to decide how to characterize (in an exportable way) just what it is that 7 is counting. To insist that it must be the number of 4's seems a little parochial, to say the least. Who says that 4 is the 4 of B′? Perhaps a deeper and more fruitful way to look at matters is to see 4 as *A′'s most frequent term.* This leads us to look for the most frequent term in B′; this is 1. So 1 is the 4 of B′. Therefore the counterpart of 7 would be the number of 1's in B′—namely, 5—another "invisible" answer, in that 5 does not appear as a visible numeral in B′.

It is narrow-minded to insist that there is a big distinction between being present as a visible numeral and being present in a more abstract sense, for example, as a counting number. To put it another way, 5 is invisible in B′

only if you think of vision as having no cognitive component, as if we could perceive only *numerals.* In fact, with our eyes we are constantly "seeing" abstract qualities. When we look at a television program, we see more than flickering dots: we see people. Of course, somewhere deep down in the processing there are components of our visual system where the dots themselves are "seen" as dots, but ironically, we would hesitate to describe as "vision" what retinal and other cells do. Vision implies *going beyond the dots*; in other words, beyond the primitive visual level. We can "see" that a certain chess position is ominous, that a certain painting is by Picasso, that someone is in a bad mood, that this car won't fit in that garage, and so on. If we accept this notion that vision is imbued with a cognitive component, then we can agree to "look beyond the numerals". In that case, 5 *is* directly visible in B'!

By the way, I carefully drew up A' so that 7 would appear as a numeral in it (as well as counting the number of 4's). This threw in a complicating factor, something one had to ignore. I could have had A' have, say, 12 4's, in which case 12 would have "appeared" in A' only at the abstract level of a counting number and not as a numeral. But real life is seldom so considerate of would-be analogy-perceivers. For example, in thinking about the question "Who is the Nancy Reagan of Britain?", you might have felt that this was much harder than "Who is the First Lady of Britain?" because you may attach certain uniquely personal attributes to Nancy Reagan, over and above seeing her as the First Lady of the United States. What if I had asked for "the Eleanor Roosevelt of Britain"? Or, turning the tables, "Who is the Moshe Dayan of the United States?" I am almost tempted to answer "Douglas MacArthur", like Dayan a famous and successful general and a controversial political figure, but then I remember—MacArthur had two eyes! Dayan's eye patch is perhaps his most memorable feature.

It is interesting to go back to earlier examples, mapping them onto A' and asking, "What here plays the role of 7?" You will perceive those old structures through new eyes (or glasses). I leave a few challenging examples for you to map onto A and A':

P: 5 4 3 2 1 5 4 3 2 1

Q: 5 4 3 2 1 1 2 3 4 5 5 4 3 2 1

R: 1 2 3 4 9 8 7 6 5 4 3

S: 1 1 2 2 3 3 4 4 5 5 6 6 7 7 1 2 1 7 6 5 4 3 2 1

T: 1 2 3 4 1 2 3 1 2 1 2 1 3 2 1 4 3 2 1

U: 2 1 1 2 2 1 2 2 2 2 9 1 2 3 2

You might enjoy making up some examples of your own, which potentially might lead a solver to further unexpected modifications of the perceived role of 4 in A. For instance, can you devise an example in which it becomes sophisticated, rather than childish, to perceive 4 as the fourth element of A?

* * *

One of the purposes of these puzzles is to dispel the notion that the full, rich, intuitive sense of a role, such as that of 4 in A or that of First Lady, can be easily captured in words. In fact, it might be more accurate to assert the contrary: that precisely in its nonverbalizability lies its fluidity, its flexibility. This is a crucial idea. Consider how you would try to capture in some phrase the precise way you see "what 4 is doing" within A. No matter what phrase you give, someone will be able to concoct another example in which your phrase does not enable anyone to predict what you will perceive as being analogous to 4. A frozen verbal phrase is like a snapshot that gives a perfect likeness at one moment but fails to show how things can slip and move. There is something much more fluid in the way a mind represents the role internally. Various features are potentially important in defining the role, but not until an example comes up and makes one feature explicit does that feature's relevance emerge.

We make comparisons all the time. It does not seem particularly noteworthy when someone walks into your kitchen for the first time and says, "I like the way your kitchen is laid out better than the way mine is. My kitchen has windows over *there* and the stove is right *here,* so it is less convenient and the light isn't so good in the morning." Clearly the words "here" and "there" conceal implicit mappings of the two kitchens, otherwise the statement would be utter nonsense. Words like "this" and "that" and phrases like "that sort of thing" are even better at picking up intangible, flexible, implicit meanings that can be transported across the borders of situations differing widely from each other. And that's the name of the game, in thought.

Right now it seems that what artificial intelligence needs is a way to go beyond "delta function" programs: programs that are virtuosos in a very narrow domain but that have no flexibility or adaptability or tolerance for errors. I call these programs "AE programs": programs that have Artificial Expertise. The trouble with them is that they are always brittle and narrow. It seems that a careful study of judgmental processes in even so simple a domain as these curious number analogies would afford fascinating insights into how computer programs might be made to approach the flexibility and generality of our own minds.

To show what I mean, I would like to conclude with a verbatim transcript of a conversation I had with a friend a while back. It ran this way:

FRIEND: Last Friday afternoon I was over at the Pooh-Bah Club listening to a piece on the radio that I was *sure* was Shostakovich. When it ended and they announced it, sure enough, it was! I was thrilled, because that kind of thing has happened to me only a couple of times in my life!

ME: *That* kind of thing? You mean, being at the Pooh-Bah Club and hearing a piece on the radio that you thought was Shostakovich on a Friday afternoon?

FRIEND: You're so *dense*! When those *Scientific American* people hear about that, they probably won't want you to write any more articles for them.

ME: Yeah, I should have known that it didn't have to be on a Friday afternoon.

FRIEND: You should have known that it didn't have to be Shostakovich!

Quite coincidentally, a recently perfected natural-language computer program called CORTEX happened to be eavesdropping on us, and it just could not resist chiming in at this point, saying, "Oh say, that reminds me —something *really* similar happened to me the other day. I was at a club whose name is hyphenated, and the water cooler broke. Ain't that something!" Well, *that kind of thing* is what I would like to see artificial intelligence programs doing more of.

Post Scriptum.

Verdi is the Puccini of music.
—Igor Stravinsky

The knee is the Achilles' heel of the leg.
—Pasadena (Calif.) *Valley Values*

The AI work out of which this column grew was my "Seek-Whence" project. My original goal was to develop a program that would take as input a sequence of integers such as 1, 4, 9, 16, and that would detect the underlying law—thus, it would "seek whence" the sequence came, and would extend it. Over a period of time, it became obvious that certain aspects of the goal were more central to mentality than others. In particular, it became clear that the ability to quickly discover the law behind highly mathematical sequences (even lowly mathematical ones, like the squares) is a specialized skill that bears little relation to mind in general, but that the ability to quickly spot *patterns* (as in "1 2 2 3 4 4 5 6 6") is absolutely indispensable.

Thus the Seek-Whence project retargeted itself on *structures composed of smallish integers;* and the major effort became one of figuring out how to perceive and "parse" such structures as these:

1 2 3 4 5 5 4 3 2 1
1 1 2 3 1 2 2 3 1 2 3 3
2 1 2 2 2 2 2 3 2 2 4 2

The latter two examples nicely illustrate one of the major problems to confront: that of boundary location. Where do you draw the lines separating one substructure from another one? And what are good ways of restructuring or regrouping if a first try fails? That whole area of concern is reflected in the project's very title, "Seek-Whence", which violates the conventional syllabic structure, namely, "see-kwents", regrouping it as "seek-wents". This kind of difficulty permeates the efforts to mechanize continuous speech recognition, and is very familiar to anyone who has been in the position of listening to a foreign language they have studied but don't know well. Often sounds will flow by so fast that you have absolutely no idea what is being said, simply because you cannot tell where the word breaks are; what is most frustrating is that this can happen even if everything being said would be perfectly clear if you saw it in writing (where word breaks are handed to you on a sil verplatter). The "parsing" of visual input is likewise permeated by boundary-location problems. Music is another such domain, and in fact the discrete multi-level patterns of melodies were among the biggest inspirations for the Seek-Whence domain.

At one point, in trying to get across the idea of Seek-Whence to someone who had a distaste for integers, I simply substituted letters for integers (*a* for 1, *b* for 2, etc.), and made up some parsing and analogy problems. For instance: "What is to *abcddcba* as *d* is to *abcdeedcba*?" Some people might say that this problem is *similar* or *analogous* to the first numerical analogy problem given in the column; I would say it is the *same* problem. Yes, in different clothing, if you like, but the same all the same. Numerals, capital letters, smalls—what's the difference? At least that was my feeling. Yet I found that I could usually awaken more interest in people if these analogy problems were presented in terms of letters instead of numbers. Groan!

From potentially infinite sequences and the *rules* behind them, my focus of attention gradually shifted to rather short sequences and the *roles* inside them (as I emphasized in the column). This concentration on roles and analogies then became so dominant that the Seek-Whence work revealed itself to be primarily a project on perception of analogies. Once this was out in the open, I decided to reify that concern by creating a new project, which I dubbed "Copycat", the idea being that being a copycat, when you're a child, is a universal and primordial experience in doing simple analogies. If I touch my nose and say to you, "Do this!", what will you do? Most people will touch their own nose. But why not touch mine? If I touch *your* nose, what will you do? Touch your own, or mine? And so on. A set of variations on this theme is shown in Figures 24-1 and 24-2. You can think of them as symbolizing the entire Copycat project.

One can be more flexible or less in how one interprets "Do this!" What does one take literally, what does one see as playing a role in a foreordained and familiar structure? What kinds of familiar structures is one willing to see as identical to each other? When is it necessary to start inventing new ways of perceiving a given situation in order to fit it into pre-existent frameworks, which then allow already-familiar roles to emerge? What remains firm, and what slips? What sticks, and what gives? These kinds of questions sound rather abstract, but when real analogies are manufactured, they are the chief

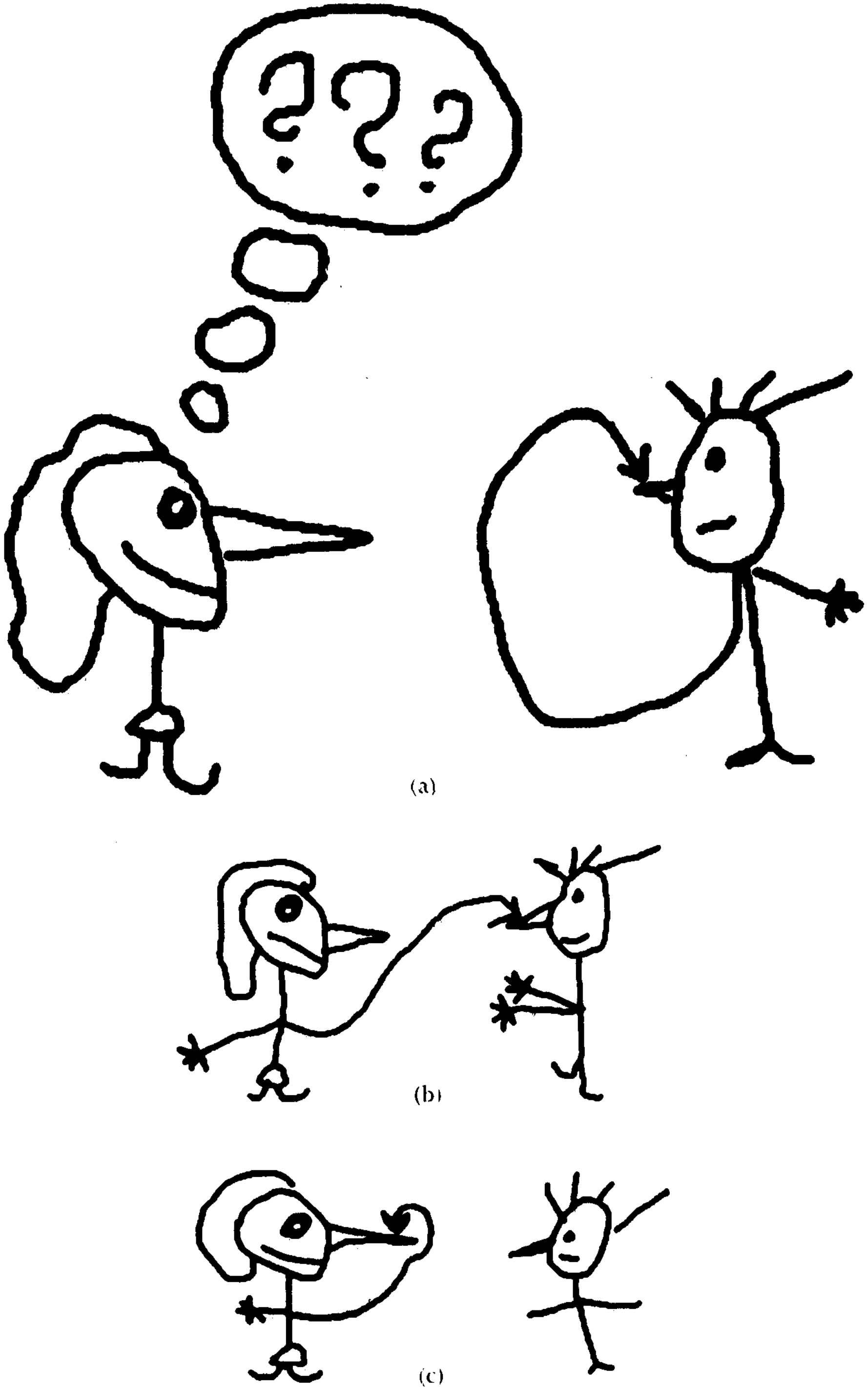
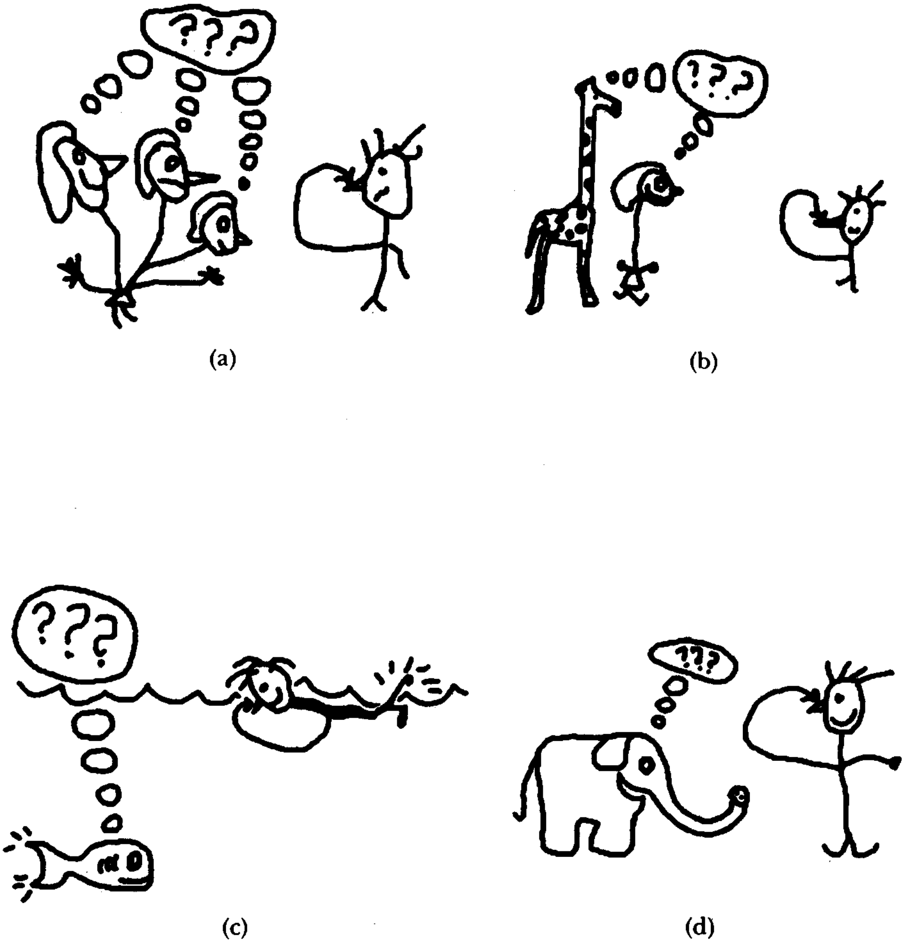


FIGURE 24–1. *The "Do this!" problem. In (a), Tom touches his nose and says to Annie: "Do this!" She wonders what she should do. In (b) and (c) you see how two Annie-clones respond. What would* you *do?*



(a)

(b)

(c)

(d)

FIGURE 24–2. *More "do-this" questions. In (a), a three-headed Annie is at a loss for what the best way to "do this" is. In (b), a long-necked Annie and her giraffe friend wonder what to do. In (c), Fanny the Fish—handless and noseless, but having gills—muses how to copy Tom. Finally, in (d), how in the world can poor little Elephannie do what Tom is doing?*

concerns of the analogy-maker, whether at a conscious level or not. Therefore these issues are the heart and soul of the Copycat project, and the purpose of this *P.S.* is mainly to show exactly how that is so.

* * *

One serious problem with studying real-world analogies like the "First Lady" examples is that they bring along too much baggage—too many complex tie-ins to all sorts of concepts. Another serious problem is that when we study real-world analogies that real people actually have made, we are nearly blinded to their essence because we have nothing to contrast them with. In order to see what makes human-made analogies good, we have to see the alternatives: analogies that humans would never make, even in jest. Can you imagine, for instance, an organism trying to understand the experience of being pregnant by likening it to being an elevator (or a football stadium, for that matter) containing one person? Something is very wrong with such an analogy—but what? Try to formulate principles of analogy-building that would suppress that type of analogy (whatever that means!) but that would recognize the insight in this one: "Try speaking English for a while without using the letters 'e' or 't' at all. Then you'll have an inkling of how Japan felt, having two of its major cities wiped out." Both these analogies involve mapping a thinking organism onto something very alien to it, and on a totally different scale. Yet one succeeds well and the other flops well. How come?

In such cases, filled to the brim with myriads of overlapping and softly blurring concepts, there is virtually no way to unravel what is really going on in a human understander's mind. To try to model all that at this early stage of research on natural and artificial minds would be as sillily ambitious as trying to master the most rich and idiomatic poetry of a foreign language without ever having bothered to tackle any prose in it at all. That would be arrogant and intellectually upside-down.

I believe that it is not merely preferable, but indispensable, to look at the analogy-making process in a pared-down domain, yet a domain where all the essential qualities of analogy-making remain. Newton couldn't have discovered his laws of motion if he had concentrated on trying to understand the laws governing waterfalls or hurricanes. Instead, he boiled the problem of motion down to the most pristine case he could imagine—planets coasting through a vacuum. This is the typical method of science: Isolate the crucial phenomena and study them in a pure context; then work your way upward towards phenomena in which two or more fundamental themes coexist.

This is what I sought to do in Copycat: To lay bare what I saw as the central problems of analogy without any extra clutter of real-world knowledge. Those central problems, as I see them, are:

- deciding how literally to take references (*i.e.*, deciding which parts of each situation already have literal counterparts in the other situation, and which parts need "literary" counterparts to be discovered or invented);
- deciding what structures are worth perceiving (*i.e.*, deciding which types of abstraction are worth bringing to bear as overarching frameworks to guide perception, so as to facilitate mapping pieces of one situation onto pieces in the other);
- perceiving roles inside structures (*i.e.*, selecting which aspects of the currently preferred organizing frameworks are most relevant and which are less relevant);
- deciding how literally to take roles (*i.e.*, recognizing which roles in each situation already have literal counterroles in the other situation, and which roles need "literary" counterroles to be discovered or invented);
- weighing rival ways of viewing a situation against each other and choosing the most elegant one (or, if you prefer, the simplest one).

The parallel passages about literary and literal *parts* and *roles* may seem a bit obscure, so let me motivate them briefly.

An entity in one situation can belong simultaneously to other situations. The sun is an example. From your point of view and mine, it is just the sun, a unique object. The sun is what I call a "part" of my world. Its counterpart in your world is the very same thing, not just something *like* it. Thus the sun is a part of *the* world. For another example, take Groucho Marx. When he died, I didn't have to put myself in my friends' shoes in order to understand what his death was like *for them.* His role in their world and his role in my world were so indistinguishable that to attempt such an empathetic mapping would have been foolishly extravagant.

But for a contrast, consider your beloved identical twin sister Glunka. Your connection to her is certainly far different from mine to her. In fact, I've never even met her, whereas *you've* known her all your life! Glunka is a very big part of your life, but from my rather distant and external point of view, her main identity is *as your twin*—after all, I know nothing else about her. Thus to me, she is the *filler of a role* in your life. If I were to learn that Glunka had died, I could hardly be expected to weep rivers, because she is not a part of my life. But that does not mean that I could not empathize, because I could project. The obvious mapping would see her counterpart as *my* identical twin sister. However, given that I am a male, no such person exists. Does that mean I am incapable of empathizing? I would be pretty inhuman if my ability to empathize were that weak. It is easy to slip from "identical twin sister" to "twin sister". Trouble is, I don't have one of those either. What can I do, then, to empathize? How about slipping along a different dimension—to my identical twin *brother*? That would be fine if he existed, but he doesn't, poor fellow. (And not because he died!) So I must loosen up still further, and try mapping your identical twin sister onto my twin brother, or just my brother. They don't exist either. Damn! In desperation, I try slipping to just plain old *sister*. Aha! This time it works. A sister I have (two, in fact), and in some loose sense, each of them plays a role in my life analogous to that of your identical twin sister in your life.

Of course, the analogy is weaker than it might be if I were twins (like you), but what can a body do? We make do with the best mapping we can find. The notion "my sister" is what I call the *counterrole* in my life to the notion "my identical twin sister" in your life. Of course, to someone else, the counterrole might be "my Siamese twin brother", or "my best friend, who I have known all my life", or even, for some people, "my car". The *filler* of a counter*role* is, of course, the counter*part.* In your life there are many parts and role fillers, as there are in mine. By discovering your life's counterparts to parts of mine, and your life's counterroles to roles in mine, you can project and identify.

How distinct are these concepts of role filler and part? Well, as concepts, they are quite distinct, but life constantly confronts us with blurry situations where people (or things) are simultaneously parts and role fillers. Think of your old and dear friend Millapollie, who is also familiar to me, but only mildly so. If you told me that Millapollie had died, how would I react? I would have dual approaches to the situation, one seeing Millapollie as a (very small) *part* of my life, and the other trying to find a *counterpart* in my life to Millapollie's part in *your* life. Thus I would try to find the filler of the counterrole in my life to the role that Millapollie plays in your life. Very probably, I would find myself flitting back and forth between these two visions of one and the same person. To effect such a part-role compromise is sometimes easy, but more often very tricky. Usually we are not even conscious of the conflicting pressures, but we muddle through anyway.

* * *

Cross-language comparisons may also help to make this idea more vivid. How eager I was, when learning French, to learn how to talk about baseball. I very much wanted to know how you say "pitcher", "catcher", "fly ball", "out", and so on. To be sure, such terms do exist in French, and it's fine to learn them, but it seems to me in retrospect to have been a misguided obsession for someone whose chief motivation was sheer fluency. In learning a foreign language, why place a high priority on learning how to talk about your *own* culture's idiosyncratic features? Instead, strive to learn the "corresponding" aspects of that culture—that is, the things that play *counterroles,* rather than the *translations* of many concepts unique to your culture. In my case, perhaps the appropriate move would have been to learn all about soccer and its terminology in French.

Of course, many terms transcend languages. It is important to know how to say "moon" in both languages, and it seems reasonable to assume that "the moon of France" and "the moon of the United States" are really the same, so that the moon is a shared part rather than a private role-filler. Now in a way this would seem true of any publicly visible entity, such as Algeria. And yet—are the Algeria of France and the Algeria of the United States really the same thing? Might Viet Nam not be "the Algeria of the United States", at least from a French perspective? Is Algeria an objective *part* of the world or something that plays a *role* in various national perspectives? What about Argentina? Australia? Antarctica? And such questions apply not only to proper nouns. What about wines, cheeses, languages?

Native English speakers are quite easily amused by very crude parodies of German, such as this sign posted near a computer terminal:

> **Alles Lookenspeepers!**
>
> Das Komputermaschine ist nicht für gefingerpoken und mittengrabben. Ist easy schnappen der Springewerk, blowenfusen, und poppencorken mit Spittzensparken. Ist nicht für gewerken by das Dummkopfen. Das rubbernecken Sightseeren keepen Hands in das Pockets—relaxen und watchen das Blinkenlights.

Our amusement is based on the peculiar way in which our language is rooted in the Germanic family. We tend to find many aspects of German gawky, comical, and old-fashioned. Obviously, German speakers will not easily see how their language has this quality to us. They will certainly not get a sense for our feelings of their language's gawkiness if they tack Germanic endings onto German words, use lots of "sch" sounds, and make long compound words—but neither will they do so by tacking on English endings, suppressing "sch" sounds, and breaking up compounds, because the historical and social connection between the two languages is not symmetric, and the effect, even if humorous, would not be analogous. But what, then, *would* be the analogue for native German speakers? What is "the German of German"?

Note that I seem to be implying that there is one best answer. Actually, I doubt there is. The connections between English and German are many and variegated. In some ways, English certainly *is* to German what German is to English. (This harks back to some problems of translating self-referential sentences, dealt with in Chapter 1 and its *Post Scriptum.*) In other ways, the assumption of symmetry is completely wrong. What would a German (or French, or whatever) parody of English (or Dutch, or whatever) look like?

How does English sound to a native Mandarin speaker studying it? I doubt I could ever know. Yet I am sure there is some fairly uniform reaction to English across the millions of Chinese people who have heard it. What would it be like to hear my own native language through ears that could not fathom it, or could penetrate it only superficially? Such an experience is denied to me—and yet, is it not exactly the same as my experience when I listen to Mandarin? Yes and no.

What entities in a given situation play the role of fixed stars, of mutually shared global points of reference? These are, in my terminology, parts. What entities are seen entirely in terms of their role relative to the perceiver, entirely as local occupiers of standard "slots"? These are role-fillers. What entities float midway between total globality and total locality, somewhere between being pure parts and pure role fillers? The answer is, of course, that *nearly all* entities are free-floating in this way, which is why the problems of analogy and translation are so deep and so deeply implicated in the mystery of mind and consciousness. The linguist George Steiner has provocatively explored these issues in his book *After Babel.*

Since a satisfactory discussion of the nature of analogy would take an entire book, I will not attempt to lay out the philosophy that the Copycat project is based on. The column gives you some good ideas (provided you can make that giant conceptual leap from numbers to letters). But it seems worthwhile presenting at least a few canonical examples from the Copycat project, since I feel they capture in a nutshell all that we are trying to do.

* * *

It should go without saying to readers of the column that Copycat deals with an alphabet of stripped-down letters. In particular, all a letter "knows" about itself is its predecessor and its successor (if it has such; *a* and *z* of course are special in that regard, each one lacking one). Letters do not know what they look like or sound like, or whether they are vowels or consonants. Since the Platonic alphabet has a starting point and a finishing point, unlike the integers, there is a kind of symmetry to it. There are two distinguished elements, namely, the endpoints *a* and *z*. These elements have identities on their own; they are somewhat like royalty. All other letters derive their identities, directly or indirectly, from these distinguished letters. Obviously *b* and *y* are like royal viziers, and *c* and *x* like vice-viziers.

I visualize the graph representing letters' "importances" as the arc of a suspension bridge, suspended at both ends from *a* and *z* and descending very steeply to a minimum at the center of the alphabet (see Figure 24-3). Thus in theory, the very least distinguished letters are *m* and *n*. However, practically speaking, all the letters in that general vicinity are pretty much equally nondescript. After all, if being nondescript were a salient property, then we would be caught in a paradox: *m* and *n* would be highly salient by virtue of being maximally undistinguished! But *m* and *n* do not know they are of minimal salience, and hence the paradox is obviated. In fact, any letters further in from the tips than *c* and *x* are pretty bland, and even those two aren't very exciting.

In the vast midwestern prairies of the Platonic alphabet, one step this way or that makes little difference. Poor *q* hardly knows what role it plays in society, since its only connections are to *p* and *r*, letters of equally little distinction. It's just as in human communities: most people are recognized in their own neighborhood, but as soon as they leave, they become

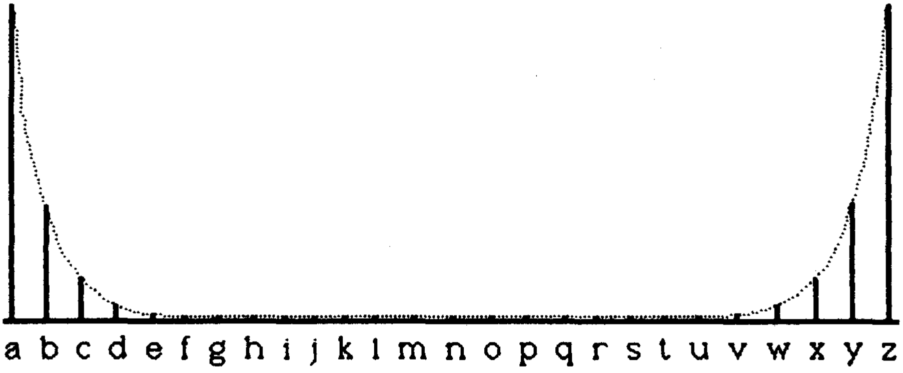


FIGURE 24–3. *A graphical representation of the effective saliencies of the letters in the Platonic alphabet of the Copycat world. The "San Francisco" and "New York" of the alphabet—*a *and* z*—are of course glamorous and salient. Nearby letters get some reflected glory, but as you leave the two "coasts", you fade into the drab middle regions, where no letter has much to draw attention to it.*

anonymous faces. The only thing you know about random strangers is precisely that: that they are strangers. Copycat letters are like people in that way.

For example, *q* recognizes that it doesn't know *k* or *x,* but they are so unfamiliar that it makes no distinction between their degrees of unfamiliarity. Only when you come quite close to *q* does it begin to act as if it recognizes you. But even then, whatever notice *q* might take of, say, *s* would not be *direct;* it would have to be mediated by *r,* which has a direct acquaintance with both letters. Generally speaking, connections decrease quickly as the number of intermediaries goes up, so that Platonic *q* loses virtually all "acquaintance" with letters much further from it than *s* or *o.* Still, there is a sort of exponentially decaying "halo" surrounding Platonic *q,* a residue of its interactions with its immediate neighbors (and *their* interactions with *their* neighbors, etc.), giving it a tapering-off set of "fringe acquaintances" (see Figure 24-4). The same phenomenon applies to all the Platonic letters, of course.

This "renormalization effect" (so called after the analogous effect in particle physics) is quite well captured by the following candid remarks made to the author by Platonic *q,* when queried about various letters:

"Mercy! I certainly don't recall hearing the name *m* before."
"Now then . . . I believe I've seen *n* somewhere around."
"Oh yes. I know *o,* though not terribly well."
"Positively. I'm old friends with *p*."
"Quit kidding! That's *me*!"
"Really a fine and true friend, is *r*."
"Sure, I know *s,* though not frightfully well."
"That's possible . . . Probably I've seen *t* somewhere around."
"Uhh . . . No, I definitely don't recall hearing the name *u* before."

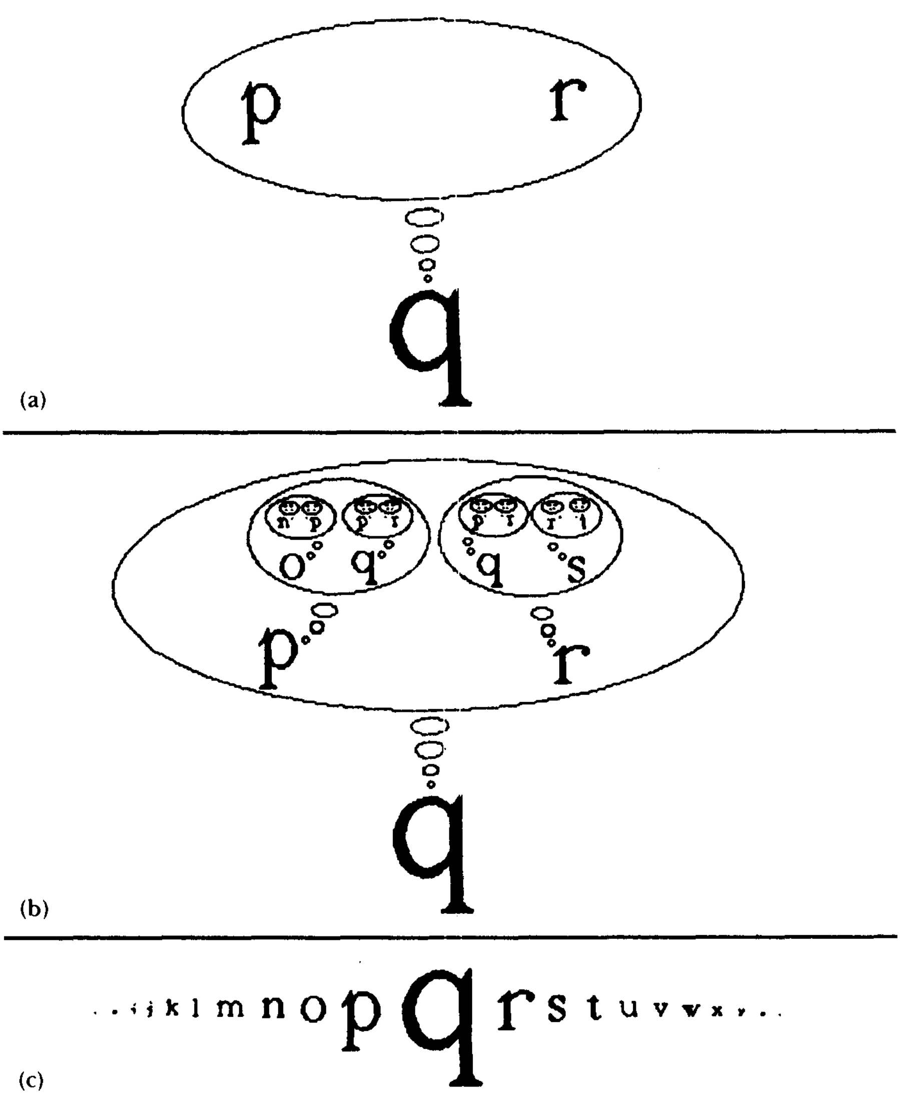


FIGURE 24-4. *The "renormalization effect" by which any letter (* q, *here) acquires a large set of* virtual *acquaintances despite having only two* true *acquaintances—its predecessor and successor. In (a), "bare"* q *dreams of its two neighbors, bare* p *and bare* r. *But those letters can in turn dream of* their *neighbors, and so on. This recursive dreaming-pattern is shown in (b). The upshot of it all is shown in (c): a* q*'s-eye view of the alphabet. This shows how* q *has an* effective *connection to every other letter of the alphabet, although the strength decays rapidly with distance, since it has to be mediated by all the intervening letters, and each extra link weakens the chain by some constant factor. This subjective vista, reminiscent of a sensory homunculus in an animal's brain, translates into a probabilistic statement for the running Copycat program: the further two letters are from each other, the less likely is any relationship they bear to each other to be noticed. Thus close letters (within two or so, as a rule of thumb) are pretty likely to be thought of in terms of their* roles *relative to each other, but further-apart letters are likely to be taken simply at face value, without any attempt to draw connections between them. This has vast repercussions on how analogies are made.*

Copycat's alphabetic world includes abstractions that lurk behind the scenes, and it is they that allow viewers to see coherence in *groups* of letters. The most basic of those concepts are two group-types based on the two simple relations that exist: (1) sameness and (2) successorship (or predecessorship, its mirror image). A group of neighboring elements linked by sameness is called a *copy-group,* or *C-group.* Here are a few C-groups:

aaa uuuuu cc wowowowowo

Notice that the concept includes the case where a *structure* (in this case, *wo*) is repeated. Degenerate cases include C-groups of length 1, such as *c* (or even *wo*), and worse yet, C-groups of length 0! Although it sounds perverse, sometimes a group consisting of zero copies of *e* is quite different from one consisting of zero copies of *f.* But this is a fine point, and only for advanced copycats.

The other group-type is that of *S-group* (and its mirror twin, *P-group*). An S-group is simply a group of neighboring successors, as shown below:

abc uvwxy cd pqrs

And if you flip these over so that they run backwards, then they are examples of P-groups:

cba yxwvu dc srqp

Needless to say, degenerate S-groups and P-groups of length 1 and 0 exist as well, but we need not worry over them. Our old friend "1 2 3 4 5 5 4 3 2 1" consists of course of a *numerical* S-group and P-group, back to back.

Beyond these most basic abstract constructs, there are more shadowy entities that move in the wings and exert intangible yet profound forces on perception of structures. These go by such names as *symmetry, uniformity, good substructures, boundary strength,* and so on. They are the kinds of forces that push you toward perceiving *abcpqr* as two groups of three, and *aabbcc* as three groups of two; they push you toward breaking *aakkkkggee* into five equal-length C-groups, and toward breaking *abcdefpqr* into three equal-length S-groups; they lend support to seeing all three of the structures *abcdcba, abcdabc,* and *abcdwxyz* as symmetric, although in three very different senses; and they make you feel quite uncomfortable with a structure such as *aaabbbqcc.* I will not try to spell out these elusive forces here, firstly because they are many, secondly because they are very abstract, and finally because people tend to grasp intuitively exactly what sorts of pressures are created by them anyway.

This concludes the presentation of the Copycat domain, nearly isomorphic to the Seek-Whence domain (except that in Seek-Whence, there is no analogue to *z*), and so without further ado, we may proceed to the analogies themselves. The basic set of analogies is actually a bunch of variations on a theme (just as in the column). That theme is the following "event":

abc changes into *abd.*

Note that it is up to you to decide what happened to *abc*; if you think that the *c* in it became a *d,* that is just fine, but it is your perception and not inherent in the change. Someone else might interpret it as a replacement of the entire "object" *abc,* lock, stock, and barrel, by *abd.*

Now, there are a number of possible counterpart situations that I could present and ask you to do "the same thing" in. It's hard to decide which one to try first, but I'll just plunge in:

What does *pqrs* change to?

Pretty easy, eh? I suppose you said *pqrt.* So does nearly everybody. That is because it is—in some rather elusive and wonderful sense—*the right answer.* Yet there are numerous other candidates that you might have considered. In fact, because it is almost impossible to appreciate the intricacy and fascination of the Copycat world unless you "live" there, I strongly urge you to pause here and think: What else could *pqrs* go to?

* * *

One possibility is *pqrd*; need I explain why? Then there is *pqrs,* produced from the input *pqrs* just as *abd* was produced from *abc*: by substituting a *d* for every *c.* Did we, or did we not, do "the same thing" here? Should you touch *your* nose, or *mine*? For that matter, why isn't *abd* or *pqds* the answer to this problem? Such questions of rigidity versus fluidity recur throughout analogy, and seem to resist formalization.

Speaking of rigidity versus fluidity, when I gave a lecture on analogies in the Physics Department at the California Institute of Technology several years ago, one Richard Feynman sat in the front row and bantered with me all the way through the lecture. I considered him a "benevolent heckler", in the sense that he would reliably answer each question "What is to X as 4 is to A?" with the same answer, "4!", and insist that it was a good answer, probably the best. It seemed to me that Feynman not only was acting the part of the "village idiot", but even was relishing it. It was hard to tell how much he was playing devil's advocate and how much he was sincere. In any case, I will never forget the occasion, since his arguing with me stimulated me no end, and at least from *my* point of view, it wound up being one of the best lectures I have ever given.

On a subsequent occasion, when giving another lecture on analogy, I remarked, quite innocently, "Last time I gave this lecture, Richard Feynman sat *right there*", and I pointed at a seat just to the left of center in the front row. No sooner had I said this than I realized the marvelous analogical transfer I had done so totally subconsciously. After all, at Caltech it had been a gigantic auditorium (this was a small classroom); the seats were in tiers (here they were just in ordinary rows); each row was very wide (here they were quite narrow); and I had been in California (now I was in Ohio). Yet pointing at one seat and saying "Feynman was sitting *there*" seemed to make eminent sense, in that context. (Isn't it equally sensible as claiming that the light bulb was invented in Dearborn, Michigan, merely because the New Jersey house that Edison did his work in has been transported to a historical park there?) Furthermore, it now occurs to me that "just to the left of center" is itself the key concept of many of the analogies in the column.

The more you look at the question of how to do "the same thing" to *pqrs*, the more possibilities you see. For instance, many people seem to like *pqst*, in which the first two letters are left alone and subsequent letters are replaced by their successors. Occasionally, people have suggesed *pqtu*, a rather ingenious notion based on seeing *rs* as a single unit whose successor is the unit *tu*. Somebody pointed out that *qrst* is a possibility, based on the idea of changing all letters but *a* and *b* to their successors. And one time someone sug-jested *dddd*, whose justification resides in the even more village-idiotic notion of changing all letters but *a* and *b* to *d*!

Some answers appear almost sick. Consider *abce*. The defense of this answer is that you take as many letters at the beginning of the alphabet as are in the target, then convert the final one to its successor. You can even come up with justifications for such queer answers as *abt* and, believe it or not, *pqre*.

* * *

When I call some answers sick, and others healthy by implication, there is something behind the metaphor. After all, there is a very serious question that always arises about analogy, but particularly strongly in such an abstract domain as this, and that is how one can ever speak with confidence about the rightness or wrongness of something that is so clearly subjective. The way I have come to view this is in terms of the *survival value* that an analogy-making capacity confers on its possessors. After all, our brains got to be the way they are only by helping our forebears to survive better than their rivals in this unforgiving world. And analogy-making is at, or close to, the pinnacle of our mental abilities.

It seems to me that people do not generally recognize how deeply implicated the analogical capacity is in decisions that affect the course of their lives. On a global level, it is evident, once pointed out. Is the embroilment of the United States in Lebanon "another Viet Nam"? How about in El Salvador? How does the American invasion of Grenada map onto the Soviet invasion of Afghanistan, or the British invasion of the Falklands? Is the Soviet Union more like an irrational paranoid person, or someone rational who has been badly bullied recently? Does the current arms race have valid precedents in history to which it can be compared?

On a more local scale, our system of law very obviously sanctifies analogy as the ultimate justification for making a reasonable and even a wise decision. The term "precedent" is just a legalistic way of saying "well-founded analogue". Two cases that at a surface level have nothing to do with each other (a bank robbery, say, and a kidnapping) may be mapped onto each other in exquisite detail at a more abstract level, with the napped kid being the loot, for instance. Lawyers attempt to sway the jury by bringing in new ways of looking at the situation that discredit their opponents' analogies, as well as by making and buttressing their own rival analogies. (Peter Suber has written a nice article connecting Copycat and Seek-Whence analogies with legal reasoning. It is called "Analogy Exercises for Teaching Legal Reasoning" and can be gotten from him at the Philosophy Department of Earlham College in Richmond, Indiana 47374.)

In our private lives, most of our important judgments are made by conscious or unconscious analogy. Should I fight this bureaucracy or accept some annoying inconvenience? Should I buy this computer or wait for a better one to come along at the same price? Should we have children now or wait a few years? Should I retire or continue working beyond retirement age? Questions concerning what to buy, what to think of someone, whom to marry, whether to move to a new city, how to talk to someone who has suffered a calamity, and on and on—all of them are influenced in a myriad ways by prior experiences of the same general sort. And remember that even in cases where there is not any obvious analogy guiding the judgment, all the categorization of the situation is being made by a mind exposed to many thousands of words, and the purpose of words is to label situation types and thus implicitly to make use of stored analogical mappings.

As was discussed in the column, the boundary line between making creative analogies and recognizing pre-existent categories is very blurry. It is signaled when we feel a desire to pluralize a proper noun ("your Einsteins and your Mozarts") or to prefix a proper noun by a definite article ("the Podunk of Albania"). Most common words hide an enormous degree of analogical abstraction. For instance, the abstractions "female" and "male" are not nearly as simple as most people think, especially when you consider how they extend to plants. What makes Middle Eastern pita, Indian puri, French baguettes, and American Wonder all be examples of the concept "bread"? When you migrate from nouns toward verbs and prepositions, the difficulties escalate. What do all "*x*-is-on-*y*" situations have in common?

All of this points out how analogies determine the course of our lives in the *present.* But I would go much further than that. In pre-civilized days, when people (or proto-people) lived in caves and hunted bison, analogy played no less important a role. Samenesses that we have absorbed into our perception as being obvious were great insights back then. For instance, the idea that one could chart out a plan for trapping a wild beast by drawing a map on the ground must have been a fabulous advance. All that is involved, in some sense, is a change in scale—one of the most obvious of analogical transforms, yet when it was first invented, it must have been revolutionary. On the other hand, proto-humans who tried burying meat underground in an attempt to imitate squirrels' underground hoarding of acorns might thereby seriously damage their chances for survival. Some analogies help, others hinder.

Our current mechanisms for analogy-making must certainly have emerged as a consequence of natural selection. Good mechanisms were selected for, bad ones were selected against, way back when, in the old times when you and I were but monkeys and rodents scampering about in tree branches ('member?). The point, then, is that far more than being just a matter of taste, *variations in analogy-making skill can spell the difference between life and death.* That's why "right answer" means something even for analogies; it's why analogies are only to *some* degree a matter of taste.

* * *

This finally gets us back to the rivalry among answers in the *pqrs* case. The domain does admittedly appear abstract and of course it is totally decoupled from the cruel world. People who prefer *dddd* are not suddenly going to get swallowed by a tiger or topple off a cliff. But people who genuinely believe in *dddd*'s superiority over *pqrt* will still have a rough time in life, because they lack the means to size up a situation and catch its essence in their mind's mesh, letting the trivial pass through. Something about their analogy-making mechanisms is defective.

To be sure, there is room for argument about answers in this mini-world, just as there is in a courtroom. But just as a lawyer who suggested that killing a human being is analogous to breaking a window because both are nasty or because both can be done with a brick would lose the case in a snap, so anyone who prefers *dddd* or *abt* to *pqrt* can be safely assumed to be totally off base. There are absurd answers, there are good answers, and there are in-between ones, just as there are degrees of edibility of food. Some foods lead to no survival, some to bare survival, and others to comfortable survival; the same is true of analogies.

One can liken the various levels of quality to the concentric circles surrounding a bull's-eye on a target (see Figure 24-5). In the middle are the totally edible foods (or insightful analogies); further out are semi-edible substances, such as grass, hay, or ants (weak analogies) and worse, leather or wood (dubious analogies); and then way out are the completely inedible things, such as nails, shards of glass, or Anglican cathedrals (these correspond to analogies that lead to disaster, such as forming a higher-level category that lumps tigers together with zebras simply because both have stripes). In the very center, to be sure, one can argue about taste and it is indubitably good for the human race that there are people who see analogies differently in that region, but you cannot get too far-fetched.

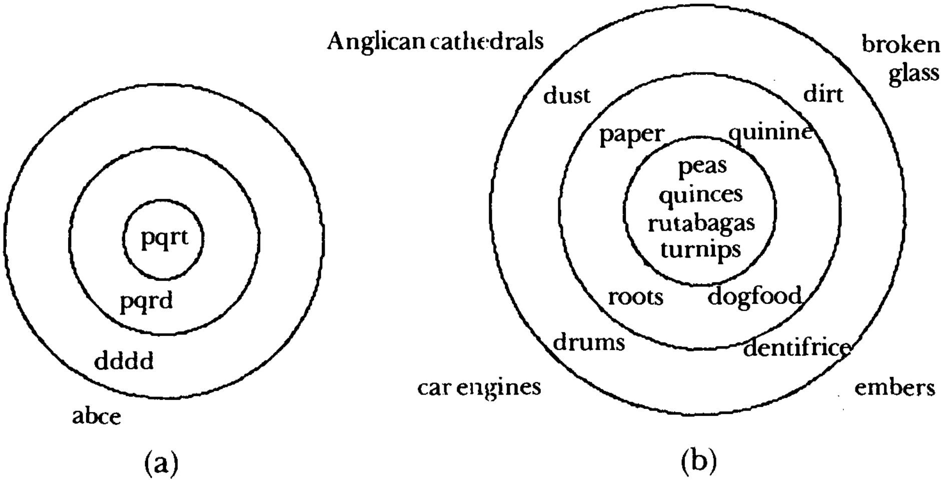


FIGURE 24–5. *Targets representing the survival values of various actions taken by (a) a mind seeking answers to the analogy "If* abc *changes to* abd, *what does* pqrs *change to?", and (b) a mind choosing things for its body to ingest. To be sure, there is room for debate in the bull's-eye region (hard to say whether having a fried egg sandwich or a plate of spaghetti is better for you), but as you edge further out, it gets less and less debatable. Fried dust just doesn't match up to pea soup! Similarly for analogies. Among the various answers to any analogy problem, some will be decidedly weaker than others, even if you find that no one answer emerges as the clear victor.*

There is a radius beyond which analogies will be very likely to bring bad consequences to their proposers, at least if they are acted upon.

It is for this kind of reason that I unbudgeably believe that there are *better* and *worse* answers to analogies, whether in life or in the Copycat domain. Elegance is more than just a frill in life; it is one of the driving criteria behind survival. Elegance is just another way of talking about getting at the essence of situations. If you don't trust the word "elegance" in this context, then you may substitute "compactness", "efficiency", or "generality"—in short, *survivability.* Insight into the mechanisms behind this sense of elegance is the goal of the Copycat project. And personally, I would not shy away from equating elegance with wit. I would venture that one cannot be a successful Copycat without a sense of humor.

* * *

Having seen numerous wild and woolly (and sometimes witty) answers to the question "What happens to *pqrs*?", you might now wish to see some alternate targets. They are extremely important, because they bring out a variety of new ways of perceiving the original "*abc* goes to *abd*" change. Here are a few fascinating challenges that I urge you, once again, to actually devote some time to considering. All of them are based upon our old stock event, *abc* goes to *abd.*

1. What does *cab* go to?
2. What does *cba* go to?
3. What does *pct* go to?
4. What does *pxqxrx* go to?
5. What does *aabbcc* go to?
6. What does *aaabbbcck* go to?
7. What does *srqp* go to?
8. What does *spsqsrss* go to?
9. What does *abcdeabcdabc* go to?
10. What does *bcdacdabd* go to?
11. What does *ace* go to?
12. What does *xyz* go to?

Each one of these questions shakes some fundamental assumptions about how you should perceive the original change. Does it necessarily affect just one letter? Need the object affected be at the righthand extremity? Does the changed piece always get replaced by its successor? In short, what should be taken *literally,* and what *slipperily* ? Although I would love to do so, I am not going to discuss all twelve examples here, for that would take a good long time. Each one merits at least a page on its own. (A set of "answers" is given at the end of this *P.S.*) I will discuss just one of them, number 12, in some depth. It has real beauty and raises all the central issues, so I hope you will give it some thought before reading on.

May I go on now? All right. Many people are inclined to say that *xyz* should go to *xya.* But who said the alphabet was circular? To make that leap, you almost need to have had prior experience with circularity in some form, which we all have. For instance: The hours of a clock form a closed cycle, as do the days of the week, the months of the year, the cards in a suit, the digits 0–9, and so on. But not all linear orders are cyclic. The bottom rung on a ladder is not above the top rung! The Empire State Building's top floor is not the same as its basement! It is a premise of the Copycat world that *z* has no successor. Sure, a machine could *posit* that *a* is the successor to *z,* but to do so would be an act of far greater creativity than it would be for you, because you have all these prior experiences with wraparound structures. The Copycat program does not. Therefore, let us consider *xya* as admirable but simply too daring, and look for something more humble yet no less apt. What else remains? Again, I urge you to think about this before reading on. This is the crucial point where there simply is no substitute for your own experimentation.

* * *

Okay. You've thought it over. You've got an answer, perhaps even a few, ranked more or less according to the pleasure they give you. Great! Some people suggest *xy,* pure and simple. Since *z* has no successor, they just let the third term drop away, as if it had fallen off the edge of the world. Some think, "Why not just leave *z* alone, producing *xyz*?" Some say, "Since the rightmost letter has no successor, why not slip over to the next-to-rightmost one and take *its* successor, thus producing *xzz*?"

Those are all right, but far more insightful answers are possible. To find them, one can let the unexpected "snag" (namely, the problem of trying to take the successor of a successor-less object) trigger a search for something crucial possibly overlooked earlier. What people tend to see at this stage is the potential correspondence of *a* and *z*—the two extreme letters of the alphabet, our two twin "monarchs". If *z* is the *a* of *xyz*, then what is the *c*? Quite obviously, it is *x*. Now the question arises: "What to do to *x*?" Should we take its successor, thereby producing *yyz*? To my mind, there is something almost repulsively rigid about this suggestion. After all, the very fabric out of which *abc* was constructed has now been reversed in our new way of looking at *xyz*. Leftward motion has seized the role of rightward motion, and with it, predecessorship has taken on the role that successorship played in *abc*. Therefore elegance, in the form of a *drive toward abstract symmetry,* very strongly pushes for the answer *wyz*. Now that's a beautiful answer, in my estimation.

There is one other answer that I have encountered fairly often, and that I admire and decry at one and the same time. That is *wxz*. To be sure, it has the same inner structure as does *abd*: a jump of size one followed by a jump of size two. That much is good, but there is something peculiar about *wxz* nonetheless. To illustrate my ambivalence toward this answer, I will relate a micro-allegory.

Arphabelle Snerxis built a lovely house, ultra-modern in every respect. Then one day she capriciously removed her snazzy, sleek, new doorknob and replaced it by a most conspicuous creaky, rusty, old doorknob. Now Zulips Twankler, a great admirer of Arphabelle Snerxis, happened to have built a lovely house, old-fashioned in every respect; and when he saw Arphabelle's action, he determined to do "the same thing" to *his* house. And how did he do that? You might guess that Zulips Twankler removed his creaky, rusty, old doorknob and replaced it by a most conspicuous snazzy, sleek, new one! But no, Zulips left his creaky, rusty, old doorknob intact and instead he tore down the rest of his lovely house, old-fashioned in every respect. Then Zulips built another quite different but also lovely house, ultra-modern in every respect—except for the creaky, rusty, old doorknob. And *that's* how Zulips Twankler did "the same thing" to his house as Arphabelle Snerxis did to her house (except that when he'd finished, it wasn't *his house* any more—it was a different house altogether).

In this "analogory", Zulips let the identity of his house slip in a manner determined by its doorknob's properties, while for most people it would seem more natural to have the slippabilities reversed. There is a parallel in the fight between answers *wxz* and *wyz*. The former sees preservation of the literal intervals (1, then 2) as necessary at all costs, and it allows the bulk

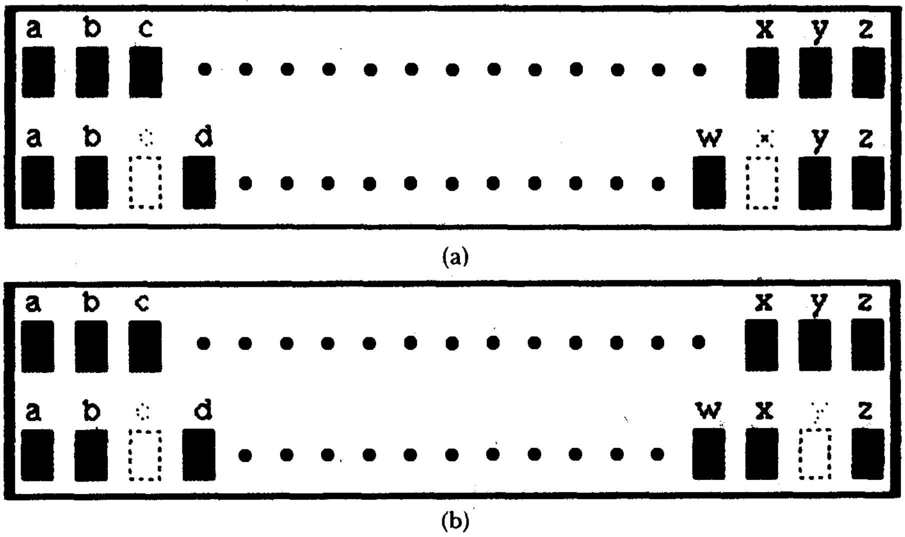


FIGURE 24-6. *A visual comparison of two good answers to the question "If* abc *changes to* abd, *what does* xyz *change to?" In (a),* wyz *is depicted. The total symmetry of this answer is its virtue. In (b),* wxz *is depicted. Its virtue is that its spacing imitates the spacing of* abd *literally, even slavishly. The question is, which of these answers creates a better overall analogy? Is it wise or is it a cop-out to intone* "De gustibus non est disputandum" *and leave it at that?*

of the original entity *xyz* to be shifted in order to achieve those aims. The latter sees only one small role (analogous to the doorknob) as singled out for modification, and keeps the bulk intact. The contrast is vividly portrayed in Figure 24-6. Another way to see the contrast is this: *wyz* is based on a better imitation of the *change* from *abc* to *abd,* while *wxz* is a better imitation of the *end product, abd.* Which do you find the more satisfying or elegant action? For me it is *wyz,* hands down. Still, I find a strange charm in the Zulipian answer. This is one of those cases where two different answers (three, if you count *xya,* that deeply creative answer that comes so easily to us old circularity hands) can lie inside the innermost "delicious and nutritious" circle, within which *de gustibus non disputandum est.*

There are a host of other "possible" answers to this problem, but many of them lie in the risky outer circles and are dangerous to their proposers. (Or, to speak more precisely, the *mental mechanisms* that would allow them to be considered seriously are dangerous to their owners, since those same mechanisms could produce very untrustworthy suggestions for courses of action in cases that are *not* decoupled from consequences.) Some of these answers are so far-fetched that they are actually quite humorous. In fact, some of them are so outlandish that I would claim no conceivable survivor of evolution's harsh pruning would ever come up with them, unless deliberately for humorous purposes. Let us look.

First of all, not all that funny but very much in the Feynman "village idiot" spirit, is *xyd*. Just plain old dullsville. What can you say to such an answer? Certainly, it has more merit than *pqrd* did earlier, for here, at least, there is an excuse for such rigidity: *z* is a trouble spot, whereas *s* was not. Another answer, definitely more pitiful, is *dyz*. This one just goes "Thud!" very loudly in my mind's ear. Anyone who would seriously suggest this answer has seen the light but then dropped the ball, mixed-metaphorically speaking. That is, they go quite a long ways in mapping *a* onto *z*, *c* onto *x*, but then they seemingly totally forget this level of sophistication and revert to an infantile literality: "Replace whatever plays the *c* role by *d*." It is so inept that it is funny. In fact, I would say that in the answer *dyz* to this analogy problem there is the germ of a rich theory of humor, based on the idea of level-mixing slips of this sort. Now if only I could say just exactly what I mean by "slips of this sort", I'd be in business . . .

Equally scramble-brained is *dba*. Its hypothetical serious proposer has clearly seen that, in some abstract sense, *xyz* is a "mirror image" of *abc*. An attempt is therefore made to take a mirror image of *abd*, but somewhere in the shuffle, the *type* of mirror image involved got forgotten, and for it was substituted the crudest possible notion of "reversal". The result is another heavy-handed thud. By the way, nobody has ever seriously suggested to me either the clunky *dyz* or *dba*, or even *yyz*, interestingly enough. Oh well, I guess such people's ancestors must have all gotten gobbled up by dreaded human-eating zebras.

* * *

These analogies, as must be abundantly clear by now, are themselves analogous to real-world analogies. Or better yet, they are *allegories* or *fables* for analogy-land children. They capture the essence of the dilemmas that analogy-makers face over and over again. Pressures for literality vie mightily with pressures for "literarity"—that is, for high-flown abstractions that disdain rigidity of reference. In an analogy where real insight is needed, often some initial forays are made that reveal some literal-minded ways of seeing the situation, but they ring false or overly crude, and so one does not stop there. Instead, one continues searching, guided by a number of small cues, and at some unpredictable moment, something simply snaps, and a host of things fall into place in a new conceptual schematization of what is going on. What was once important becomes suddenly trivial, and a new essence emerges, an organizing concept or set of concepts that seem far superior.

But beware! You should not take any of this to imply that being literal-minded is to be avoided at all costs. If jumping to rarefied levels of abstraction were always preferable, then we would wind up never making any distinctions between situations. Every situation's optimal description would be: "Something happens." It would be better to say "woman" than "Mary", "person" than "woman", "animate object" than "person", "physical entity" than "animate object", and ultimately, just "thing" would emerge triumphant. (Actually, even that choice would be nonexistent, for those different words would not exist, as they would merely make petty distinctions that vanish at enlightened higher levels.) Rigidity comes in many forms, and a rigid drive toward abstraction is no less stupid than a rigid refusal to abstract. The clearest and cleanest statement of the problem that analogy poses is that there are always fights between forces pushing for literality (in its many forms) and forces pushing for abstraction (in its many forms). How, in specific circumstances, those forces compete and interact and in the end come up with some sort of optimal compromise is the problem. And if you go away from this chapter with one thought, please let it be this: There is no fixed mathematical recipe for reconciling all the different forces pushing and pulling you in analogies.

* * *

In the Copycat world, we have remarkably fine control over how pressures interact, and over the strengths of various rival answers. For instance, we can "tune" a given analogy by varying knobs on it, until gradually we home in on the most refined possible version of it. We can also take two possible answers and make each one seem preferable by very subtly "tweaking" the analogy itself. It is a a highly pleasing esthetic exercise to seek the perfect balance point where an analogy is teetering on the brink and can tip either way, so that about half the people to whom we give it see it one way and half see it the other way.

A lovely example of this idea is provided by a very simple analogy using a knob that twiddles the part-*vs.*-role proportion perceived by a typical person. This example may help clarify that distinction a little, as well. Consider the following three variations on a familiar theme:

1. If *abc* goes to *abd,* what does *pqrs* go to?
2. If *abc* goes to *abe,* what does *pqrs* go to?
3. If *abc* goes to *abf,* what does *pqrs* go to?

Line 1 is familiar by now. There, nearly everyone instantly proposes *pqrt*; very few people think of, let alone prefer, *pqrd.* The reason is that the *c-d* conversion seems so obviously to be a "leap" from one letter to its successor that we ourselves leap to that conclusion. But consider line 3. Here we are confronted by the much larger *c-f* gap, too large for us (or the Copycat program) to have any intuition for. It seems so arbitrary that instead of seeking any "whence" behind it, we just accept it at face value, saying to ourselves, "Okay, replace the rightmost letter by *f,* eh?" And indeed, most people are happy with the answer *pqrf.*

To some, this answer may seem overly Feynman-like, but I must reiterate that in the Copycat world, there is no simple connection between distant letters. You can't just "subtract" *c* from *f* the way you can subtract 3 from 6. Subtraction is an unknown concept, just as in Seek-Whence. The connection between *c* and *f*, to the extent there is one, is a *conceptual* one rather than a mathematical one: *f* is the successor of the successor of the successor of *c*, and that is just too topheavy a notion to have much charisma.

All right, if line 1's *c-d* leap has charisma and line 3's *c-f* leap lacks it, what about the intermediate case of line 2? Here we are poised between two analogies that push us in opposite directions. If we were to follow line 1's example, we would see the *c-e* leap *intensionally*, a fancy way of saying that we would see *e* as a *role-filler* (as the successor of the successor of *c*—a bit gawky but still plausible). But if we were to follow line 3's example, we would see the same leap *extensionally*, meaning that we would see *e* as a mere *part* of the event, filling no role other than being itself (which it could hardly help doing). This is not gawky, but it is so literal that it provides no insight into why the given change occurred. So which do you prefer—the intensional view of *e* as gawky role-filler, or the extensional view of *e* as arbitrary part? The former leads you to answer *pqru*, the latter to answer *pqre*.

Although line 2 may not be your personal balance point, there is surely a point at which you will switch over from one view to the other. It would seem highly compulsive if, given the question

abc goes to *abv*; what does *pqrs* go to?

somebody insisted that the *v* must somehow "come from" the *c*, and tried to force a vision of some connection when there really is none. Furthermore, even violating the spirit of Copycat and seeing *v* as the order-19 successor of *c* is not of much help, for what is the order-19 successor of *s*? It gets us right back to successor-of-*z* problems, very messy territory.

Seeking the balance point of analogies is an esthetic exercise closely related to the esthetically pleasing activity of doing ambigrams, where shapes must be concocted that are poised exactly at the midpoint between two interpretations (see Figures 13-6 and 13-7). But seeking the balance point is far more than just esthetic play; it probes the very core of how people perceive abstractions, and it does so without their even knowing it. It is a crucial aspect of Copycat research.

* * *

A few more choice problems are given below for would-be copycats. I do not have the space-time to discuss them all here; I propose them simply because each one has a chance of affording you a small but thrilling moment of blinding insight, if you look at it in just the right way. Our view of the best answers is given in the *Post Post Scriptum*.

1. If *aqc* goes to *abc,* what does *pqc* go to?

2*a*. If *efgh* goes to *fghi,* what does *mvr* go to?

2*b*. If *efgh* goes to *fghi,* what does *uuuuu* go to?

3. If *beq* goes to *bqe,* what does *abcdefpqr* go to?

4. If *xyzabc* goes to *xyzqabc,* what does *abcxyz* go to?

5. If *aaqqkkkk* goes to *zaazqqzkkzkkz,* what does *abcdefstu* go to?

6. If *eeeffghhiii* goes to *eeeefffgghhhiiii,* what does *eefhii* go to?

7*a*. If *eqe* goes to *qeq,* what does *abcdcba* go to?

7*b*. If *eqe* goes to *qeq,* what does *aaabccc* go to?

7*c*. If *eqe* goes to *qeq,* what does *eqg* go to?

It must be emphasized that the selection of Copycat problems presented here is but a tiny fraction of all those that I, together with David Rogers, a postdoctoral fellow working on the Copycat project, have come up with. This selection is biased toward analogies that have spice and tang, as opposed to bland ones such as "If *bbb* goes to *bbbb,* what does *eee* go to?" There are of course innumerable ones of this boring sort, ones that have obvious answers and that present no serious challenges to adult humans. Now, we do not in the least disdain such analogies in the project. Indeed, it is a tremendous challenge to make a program that could handle these seemingly easy cases reliably. They are amazingly deceptive in their subtlety. But it is not of much interest to people to go down a long list of (not actually but seemingly) trivial analogies, so that explains the censorship in my choices for you.

Still, it must be admitted, analogies that seem to require a deep perceptual shift after an initially unsatisfactory first stab are the ones that beguile us, for they seem to promise insight into that mystery of mysteries: insight. I must admit to the belief, or at least the strong intuition, that all the depth of scientific discovery, even the profoundest discovery, is wrapped up in the mechanisms for solving these simple problems in which conflicting pressures push around one's percepts and concepts, letting things bounce against each other until, all at once, something falls into place and then, presto! A sense of certainty crystallizes, so powerful that you *know* you have found the right way to look at things. I firmly believe, in short, that "mini-breakthroughs" and "maxi-breakthroughs" have precisely the same texture. That's the faith underlying Copycat.

It may seem arrogant or blasphemous to compare the trivial alphabetic insights of a copycat with the genius of an Einstein discovering special relativity, yet I do not think the comparison is all that silly. What characterized Einstein's unique view of space and time was that he had decided that certain things were more unslippable than others: in particular, that the speed of light was unslippable but the notion of absolute simultaneity of events separated in space was slippable. To be perhaps more accurate, Einstein didn't *decide* that simultaneity was slippable, but was *forced* into that conclusion, since his stronger intuitive belief in the invariance of the speed of light simply compelled him to accept its consequences, strange and counterintuitive though they might be. (Note that counterintuitive consequences can flow from intuitive grounding.) Einstein did not begin with the idea of simultaneity being nonabsolute, but when he had to confront that possibility, he let it slip. This fluidity of mind, guided by a certainty about the deepest, most unslippable concepts, gave rise to the creative insights of special relativity.

There is an old song whose lyrics go this way:

When an irresistible force such as you,
Meets an old immovable object like me,
You can bet, as sure as you live,
Something's gotta give, something's gotta give, something's gotta give!

Yes, something's gotta give, but what? A reliable nose for what might slip and what ought not marks the difference between a great mind and a small one. If the Copycat research can unearth the basis for judgments exhibiting creative, artistic slippability even in our tiny domain, we will be ecstatic, for in our opinion, that would put us well on the road to understanding where full-scale artistic creativity comes from. Now *that* may sound arrogant, but firstly, we are not expecting it to happen just around the corner, and secondly, it is just an expression of our faith that we have not lost the essence of the larger problem in boiling it down this far. If Newton saw whirling planets in falling apples, why can we not see great leaps in small slips?

* * *

One can look to literature as well as science to find cases where finding the right things to slip yields highly creative solutions to hard problems. One example I gave in the *Post Scriptum* to Chapter 1 was the problem of translating into French (or the foreign language of your choice) the title of the book *All the President's Men.* A word-for-word translation would be as dull as dishwater, as flat as old ginger ale whose carbonated kick has long since evaporated. In order to keep the title alive in French, you must seek out a line well known to readers of French that carries the same subliminal imagery. Need it be a line in the canonical translation of "Humpty Dumpty"? Need it even be a line from *Mother Goose*? Of course not. The essence of the situation does not reside in those particulars. So slip, baby, slip! But how?

The crux of the matter is to find a line alluding to a famous irreversible downfall. If it is a line from Pascal's *Pensées,* so be it. If it is from a popular song of recent years, so be it. You may have to go further afield to find an appropriate line. There may be no line of the sort in the popular French-speaking consciousness, in which case more radical solutions must be sought. There is no clean, clear recipe guaranteed to work. By the way, I do not have any idea if that book has ever been translated into other languages, and if so, what solutions its translators found. But this type of problem is absolutely standard, since these days particularly, book titles of that style, making an oblique allusion to some well-known phrase, are a dime a dozen.

I must admit to some twinges of shame for having leapt aboard the title-as-pun bandwagon, when Daniel Dennett and I chose the title *The Mind's I* for our anthology. Good but non-native speakers of English usually are confused by this title, and often read the final "I" as the roman numeral for "one", which makes absolutely no sense, yet it is the best they can do, being unfamiliar with the idiom "the mind's eye".

Just to give a hint of how creative solutions can be found for such titles, I'll give one possible French translation for our title (even though it is not yet certain whether the book will ever be translated into French). Jacqueline Henry, one of the two translators into French of *Gödel, Escher, Bach,* came up with *Vues de l'esprit*—literally, "Views of Spirit", which clearly gets across one main purpose of the book: to focus on the nature of mind from many angles. But at the same time, it has a more idiomatic meaning, namely: grandiose dreams such as are dreamt by visionaries (and lunatics)—in short, *visions* or possibly even *hallucinations.* This too has its own kind of appropriateness, since a basic theme of the book is that much of the mystery surrounding mind, spirit, and soul is caused by a kind of hallucination: the hallucination that there is some *thing* called "I". Therefore, the French double meaning is elegant and, though it does not replicate in French the exact effect of the English *double entendre,* it is effective and thought-provoking. What more could you ask?

Incidentally, if I were writing this *P.S.* in French, I would of course talk about books in French, not ones in English, whose titles are hard to translate. Thus a proper translation of this very passage would involve a good deal of literarity. In fact, I have one particular book title in French in mind: *Le corps a ses raisons*—a book about health and physical fitness, which actually came out in English under the feeble title *The Body Has Its Reasons.* Can you do better? Hint: You need to be familiar with a famous saying by Pascal, namely, *Le cœur a ses raisons que la raison ne connaît point.*

* * *

In a certain way, translation is the quintessential form of analogy. You are given a fixed overarching framework—the home language and culture—and within it, a novel structure has been erected—a book title or sentence, for instance. Your task as translator is to replicate, as best you can, the overall "feel" of that small structure, but in a different overarching fixed framework—the target language and culture. This description obviously recalls the allegory of Arphabelle and Zulips, where the frameworks are their respective houses, and it applies equally clearly to the "Do this!" examples.

My own mental image that best gets at the nature of translation involves

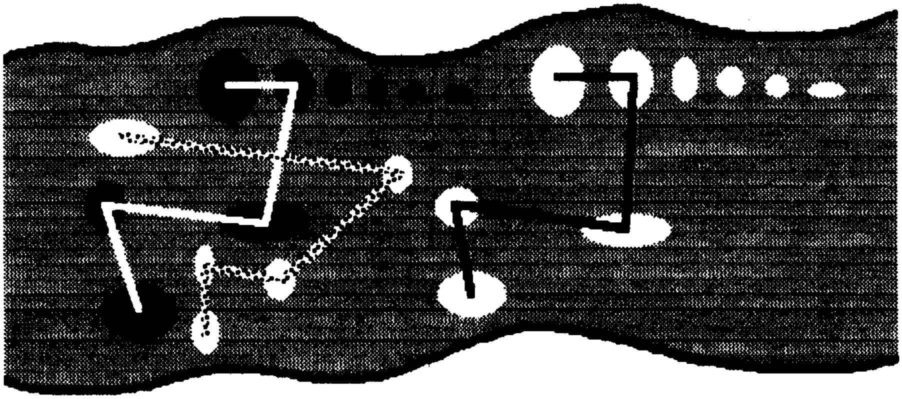


FIGURE 24-7. *A metaphor for translation. A stream (symbolizing reality) has two sets of stepping-stones (symbolizing the basic ingredients of a language, such as words and stock phrases) in it. The black stones (Burmese, say) are arranged in one way, and the white stones (say, Welsh) in some other way. A pathway linking up a few black stones (a thought expressed in Burmese) is to be imitated by a "similar" pathway joining up white stones (translated into Welsh). One possibility is the speckled pathway, located at nearly the same part of the stream as the original pathway but not terribly similar in shape to it (a fairly literal translation), while a rival candidate (a more literary translation, needless to say) is the pathway located a distance upstream and resembling the original in some more abstract way, including patterns in some of the "overstones" of the main stones (the similar archipelagos in Burmese and Welsh stones running roughly parallel to the far bank).*

picturing each language as a fixed set of stepping-stones in a stream (see Figure 24-7). Suppose you are translating from Burmese to Welsh. A Burmese utterance is a pathway from one place to another via the black stones. They seem to be located in convenient enough places, and you can get pretty much wherever you want to go. But when it comes to translating what you have said into Welsh, you find that the Welsh stepping-stones—the white ones—are often not quite in the same places as the Burmese ones, and even in the cases where they are just about in the same places, they are shaped differently, and so you can't treat them as identical to the Burmese stones you are familiar with. You must tread with great care, and sometimes you will find that there are gaps in one language's set of stones that don't exist in the other, so that some routes are easier to mimic than others. The most literal translations involve sticking as close as you can to the original route, at the stone-by-stone level. Of course, no mimicking route is exactly the same as the original.

It may be, however, that the essence of a particular route lies not in *where* it starts or ends in the stream, but in its *shape*. It may be that in the particular region of the stream where the original path was traced, the Burmese stones very easily allow many shapes to be traced out but the Welsh stones happen to be sparse. At various places upstream or downstream, however, the converse is true. If you believe that the essence of the idea resides more in its shape than in its absolute location relative to the stream bed, then you won't mind moving upstream or downstream a bit, in order to gain that flexibility. Less metaphorically speaking, this means that sometimes the overt topic of a passage can slip as long as something more central—style, perhaps, or metaphorical allusion—is preserved.

A critical idea is the following one: The longer a passage is, the less the graininess of the underlying medium is going to be noticed. In purely geometric terms, what I am saying is this. The larger the curves of the pathway are in comparison to typical inter-stone distances, the less it matters which of the two grids of stepping-stones you are using. This can be illustrated very elegantly by thinking of trying to approximate a circle by filling in various squares on a normal (8×8) chessboard. Clearly you would make the circle as big as you can within the confines of the board, so as to round off the effects of the squareness. And if you could make the board bigger, you would. On a 100×100 chessboard, you could draw a very fine approximation to a circle, and on a 1,000,000×1,000,000 board, no one would know the difference. Furthermore, nobody would even be able to tell, in such a case, whether the underlying board was a square lattice, a hexagonal lattice, or what. But if you go back down to circles whose size is on a par with that of the lattice grain, then of course the lattice becomes very visible.

For this reason, I feel safe in suggesting that translating a novel's *title* may sometimes be the most challenging aspect of translating the whole novel. After all, the overall message of most novels is on such a vastly larger scale than the grain size of either language involved that small jogs here and there (where the idiosyncratic placing of the stepping-stones forces you to take an awkward zigzag) can be compensated for in other places, and in the larger picture such jogs will balance or cancel each other out. Recall that I said something similar about computer languages and AI programs—the grain size of the ideas in a big program is far larger than that of any conceivable computer language.

But a title is another story. A title is tiny. Its grain size is barely above that of the stepping-stones themselves. It consists of a pathway just a handful of stones long, and the challenge is great when it contains subtlety of any sort —which is the case for most titles, as it is for proverbs, epigrams, and so on. As they say in Italian, *Traduttore, traditore*—which, literally as well as literarily translated, means "Translator, traitor." In this curious case, the English version is a perfect counterexample to its own claim, but generally speaking, the Italian epigram is right on target, and pithily expresses the idea that no translation—no analogy—is perfect. Perhaps a better English translation of it would therefore be: "Transductor, treasoner."

* * *

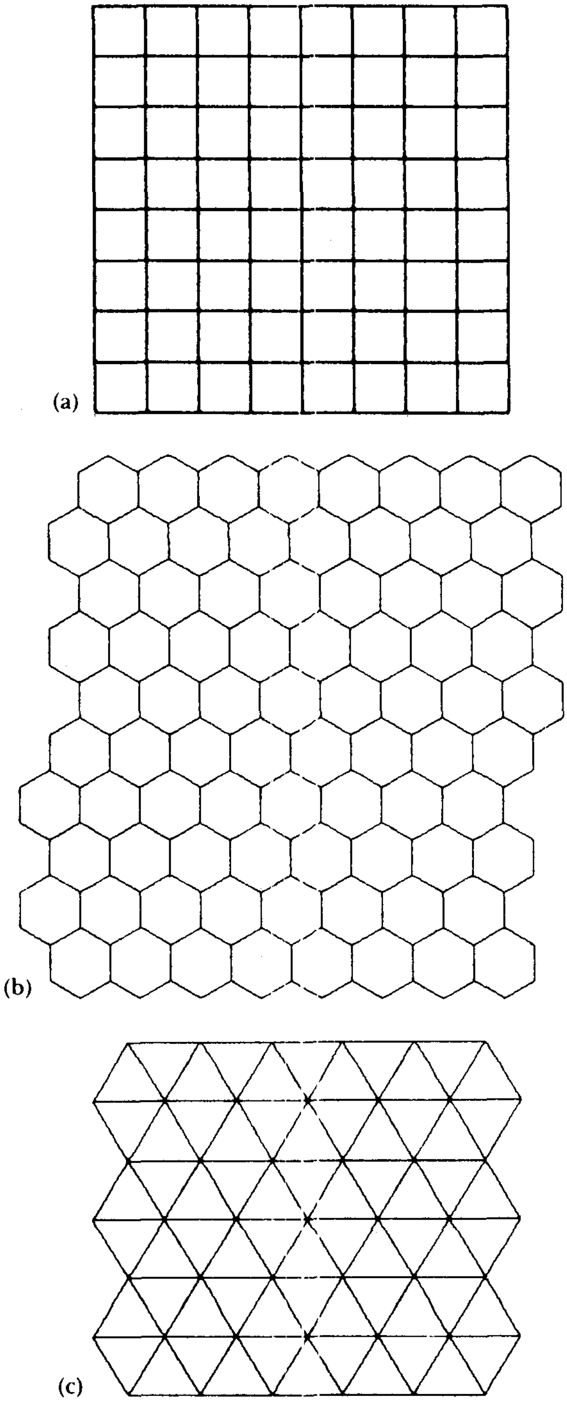


FIGURE 24–8. *Three lattices on which chess-like games could be played. In (a), a square lattice; in (b), a hexagonal lattice; in (c), a triangular lattice.*

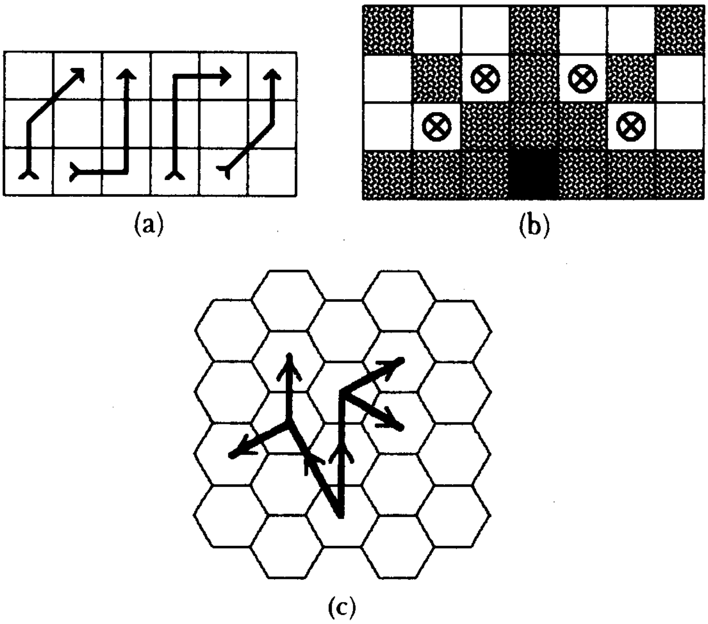


FIGURE 24–9. *Various ways to think about the knight's move. In (a), it is built out of rook-move and bishop-move primitives. In (b), it is portrayed as the closest spot not immediately accessible to rook or bishop. In (c), we make some first stabs at the knight's move on a hexagonal lattice. Are some of these possibilities more defensible than others?*

The idea of approximating a given shape (such as a circle) on a coarse-grained grid (such as a chessboard) provides a wonderful way of framing many analogy issues. But the target shape need not be subtly curvilinear for the analogy problem to be deeply challenging. If you are trying to export even a very simple shape from one grid—its "natural habitat"—into another grid, and if it does not export literally to the target grid, then something's gotta give, and that is the hallmark of a hard analogy problem.

Since we are talking about chessboards, let us use a chess example. The underlying grid of chess is a square lattice. Suppose we pick as our target grid the hexagonal or triangular lattice (see Figure 24-8), and ask, "What is the knight's move on this lattice?" We are immediately forced to confront the question, "What is the essence of the knight's move in the only case we really know?" There are a number of ways of thinking about it (see Figure 24-9). Which of the following, if any, is the most insightful characterization of the knight's move?

(1) a rook step of length two followed by a single perpendicular rook step;
(2) a single rook step followed by a perpendicular rook step of length two;

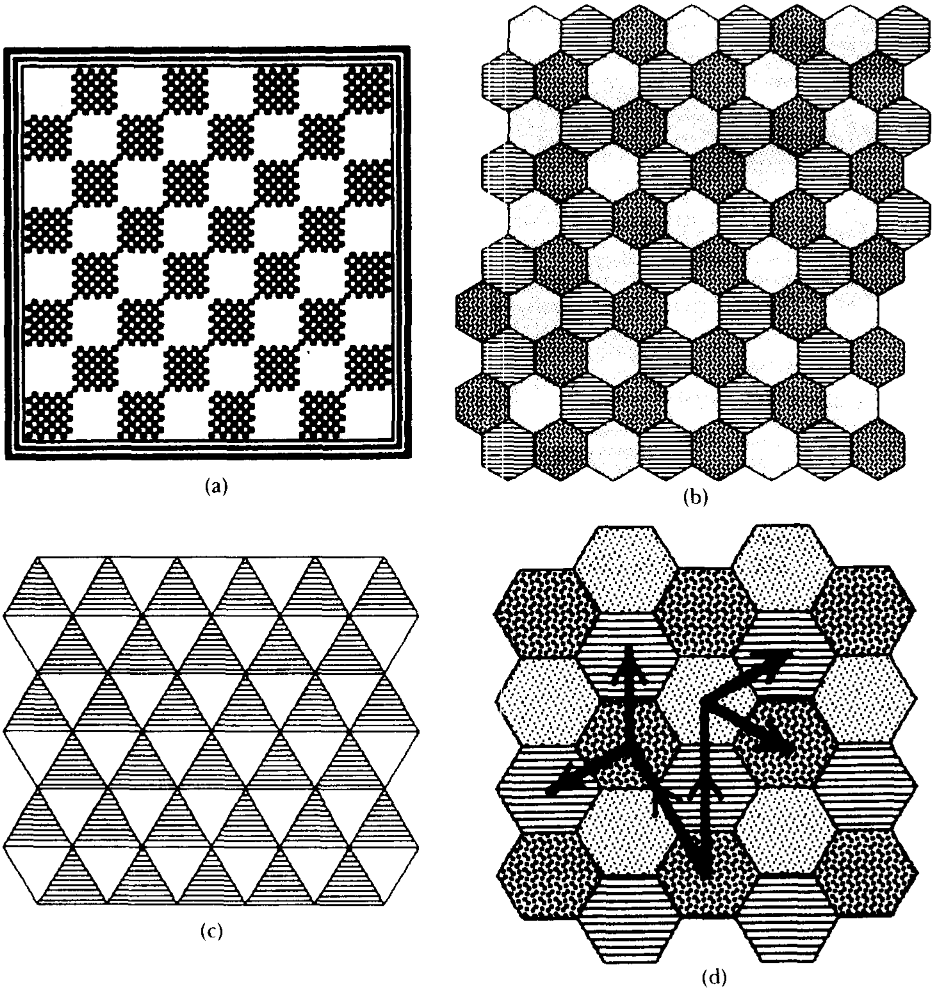


FIGURE 24–10. *Recognizing that the coloring of the board plays a significant role in defining the moves of chess pieces. In (a), the standard coloring pattern of a square lattice. In (b), the most natural coloring of a hexagonal lattice. Notice that it involves* three *colors. In (c), the most natural coloring of a triangular lattice. Finally, in (d), the guesses of Figure 24-9 are now shown on a colored-in hexagonal lattice. How does this affect their plausibilities?*

FIGURE 24–11. *In (a), a board for unidimensional chess, known as* chass. *Its optimal number of squares is to be determined. In (b), a wider board for* quasi-*unidimensional chess. This variant is known as* chäss *or* chæss.

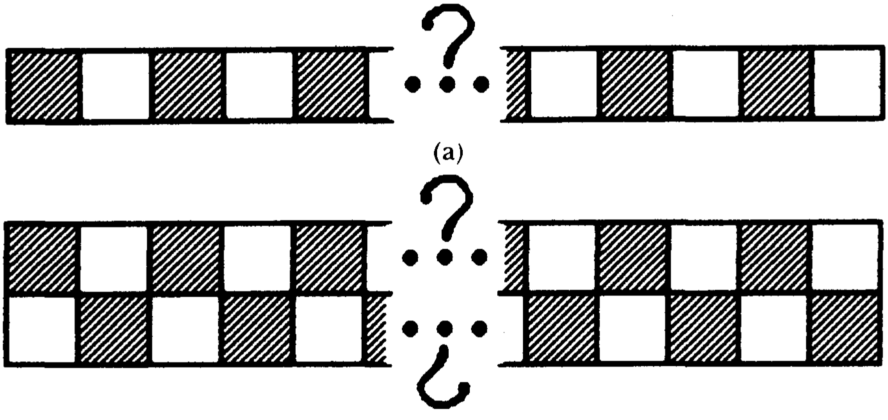
(3) a rook step of length one extended by a bishop step of length one;
(4) a bishop step of length one extended by a rook step of length one;
(5) a normal pawn move followed by a pawn's move in "capture mode";
(6) the shortest move that no other piece can make.

Or are all of these simply *aspects* of the essence of the knight's move? Which aspects are more central, then? Which would you be willing to relinquish first? Which never? Are you sure? How slippable are the following aspects, all of which do apply in the square grid?

* When a knight moves, it must land on a square of a different color.
* A knight must be able to jump over or around pieces.
* A knight can make a tour of the entire chessboard, landing on every square.
* The starting and stopping squares of a knight's move should lie on opposite sides of a straight line that contains one edge of each.
* The knight's move should not resemble any other piece's move.
* All knight moves should be congruent except for rotation and reflection.
* A knight must have about the same power as a bishop and less power than a rook.

Once you have tried extending the concept of the knight's move to another lattice, then you begin to sense all the subliminal features that add up to define its highly composite identity, features that you most likely never would have thought about without this pressure. For example, coloring the lattices was a revelation for me, and turned out to be the royal road to finding elegant knight's-move solutions (see Figure 24-10).

While working on these two puzzles, I dreamt up what sounded at first like an absurd challenge: to compress chess into one dimension. In other words, take a chain of squares of length N (to be determined) and find moves for rook, bishop, knight, queen, king, and pawn (see Figure 24-11).



(a)

(b)

Also consider how to place the pieces on the board at the game's start, and determine *N*. This droll exercise gave rise to a number of very stimulating discussions among AI colleagues and chess-playing friends of mine. The sense of sheer analogical elegance had to compete with realism about what choices might make the game more interesting. Along the way, one unexpected suggestion arose: What about a board *two* squares wide, rather than just one? On that type of grid, the knight's move seemed trivially obvious—but my "obvious" solution turned out to have a fatal flaw, which we patched in a most intriguing way.

One of the more amusing spinoffs of those discussions was a quest for the name of the game (as they say). One-dimensional chess was dubbed *chass,* since 'a' is the first vowel. Two-dimensional chess retained its name, 'e' being the second vowel. (But what would seven-dimensional chess be called?) The 2×*N* game, delicately poised between one- and two-dimensionality, was yclept *chäss.* And what about the names for chess on the nonstandard two-dimensional lattices? Here are some solutions that fell into a delightful pattern:

> *Chesh:* chess played on a *hexagonal* lattice;
> *Chest:* chess played on a *triangular* lattice; and of course,
> *Chess:* chess played on a *square* lattice.

And please—when you are about to deliver the death blow to your opponent's king in a game of chass, don't forget to triumphantly exclaim, "Chackmate!"

* * *

I found these chess-extension puzzles to be beautiful not only as puzzles, but much more rewardingly, as examples of issues in analogy. When, for example, I settled on my answer for chesh, it felt like not just *an* answer, but *the* answer: *the* knight's move on a hexagonal lattice. This reminded me strongly of the feeling of absolute certainty that one gets in mathematics when one sees a familiar phenomenon recurring in a new way in an unfamiliar domain. One says, "Aha! So *this* is how Wiggler's Lemma generalizes! The twistoploppic theomorphism is the same for *even* clackdoodles, but becomes a hypertwistoploppic pseudotheomorphism for *odd* clackdoodles. That's so *beautiful*!"

Examples galore of this feeling must have arisen in the minds of the people who extended the Magic Cube concept to other polyhedra, other dimensions, other ways of slicing. And once you have made or acquired a new "cube" (such as the Skewb or IncrediBall), you will want to know how to export a known *algorithm,* broken up into its fundamental *operators,* from a familiar cube. What is the *essence* of each operator? One senses a deep invariant lying somehow "down underneath" it all, something that one can't quite verbalize but that one recognizes so clearly and unmistakably in each new example, even though that example might violate some feature one had thought necessary up to that very moment. In fact, sometimes that violation is what makes you sure you're seeing *the same thing,* because it reveals slippabilities you hadn't sensed up till that time.

No better example exists than the way mathematicians extended the concept of *exponentiation*—putting x to the y power. At first, y had to be a positive integer. Then it was realized that $x^0=1$ fits exactly into the pattern, so zero was allowed as an exponent. Immediately, it was seen how the same pattern would suggest—nay, *require*—that x^{-1} be the reciprocal of x. By then the generalization ball was off and rolling. Fractional powers came along very quickly: 1/2 as an exponent meant you should take the square root, 1/3 meant the cube root, and so on. Then on to real numbers. But why stop there? Various abstract representations of what it meant to exponentiate had now been formulated, allowing one to transcend earlier, primitive notions of what it meant. Pretty soon, not only could complex numbers be exponents, but so could $n \times n$ matrices, functional operators, and God knows what else! This conceptual supernova was still very much centered on one core, and blurry though the implicosphere around it might be, the vastness of this implicosphere's size only made the conceptual core stronger, firmer, realer.

Another example: There is clearly only one sensible 4×4×4 Magic Cube. It is *the* answer; it simply has the *right spirit.* The same holds for the four-dimensional cube, discovered by Kamack and Keane. Similarly, Scott Kim once generalized the concept of the "impossible triangle" (a two-dimensional drawing read by three-dimensional viewers as a three-dimensional object that cannot exist) up one dimension, so that it became the "impossible skew quadrilateral" (a three-dimensional sculpture read by four-dimensional viewers as a four-dimensional object that cannot exist). Later, he found out that Roger Penrose, the inventor of the impossible triangle, had likewise "added 1" to his three-dimensional illusion and come up with exactly the same construction as Scott had—only Penrose did it fifteen years earlier. Clearly, then, this was *the* corresponding paradox, manifesting the same deep essence as the original and simpler one. Here again we see the aptness of that wonderful saying, *Plus ça change, plus c'est la même chose.*

That feeling of encountering an absolute and almost divine truth and reality behind highly abstract analogical connections is particularly prevalent in mathematics, but it can also arise in other areas of life. When a "reminding" feeling becomes so strong that you want to use the *same word,* that is when you start getting religious about your discovery. Golomb's "quarks" on the cube, for instance, seem to have some "essence of quarkness" about them. Is this *one phenomenon* manifesting itself in two different ways, or is it simply a pretty coincidence? Such questions can occasionally not be answered, but very often our minds come to conclusions on such matters without our ever noticing it. Reification of new categories in words is a telltale signal, and one of the most important of mental events.

* * *

Some people might look upon the exercise of translating the knight's move into an alien grid as an amusing but trifling game, and maintain that such things are far from real-world concerns. Actually, in recent years, problems not too different from this have become the stock-in-trade of people working on the computerization of typefaces, where the idea is to pack as much of the spirit of a typeface (such as Helvetica) into the smallest possible number of "pixels" (on-off dots, usually arranged in a square lattice, though that is not necessary). Can one make an 'a' that is recognizably a *Helvetica* 'a', using just 35 pixels arranged in a 5×7 array? This is certainly beyond feasibility. But how few can you get away with? When does at least a hint of "Helveticality" start to appear? (See Figure 24-12.) And just what is this "Helveticality" spirit that is so elusive? How much harder to capture than "essence of knight's move" is it?

Attempting to compress a visual form into smaller and smaller arrays of pixels forces one to confront ever more deeply the question of its essence. What can one afford to release, and what must be held onto? An analogous *aural* analogy problem is very obvious to state, yet seldom explored: Can one translate a complex piece of music from major into minor, or vice versa? Musicians will immediately recognize that the major and minor scales here play the roles of underlying *lattices*, so that we are undeniably dealing with a lattice-conversion problem. Mechanical methods will carry you a certain distance, to be sure, but for any complex piece there will always remain a lot of sticky and idiosyncratic knots. For instance, what if a piece in a major key turns minor for a short stretch? Should its minor-key "translation" turn major at the corresponding point? This example is just the tip of the iceberg in the major-minor translation game. To get into the right spirit, you might try humming to yourself such old favorites as the popular song "Awful Days Are Here Again" (traditionally sung by mournful Democrats right after they lose an election) and Frédéric Pichon's celebrated Baptismal March (from his piano sonata in B-flat major) . . .

Another musical analogy problem arises when one tries to arrange a piece of music for a new instrument or group. Can George Gershwin's very pianistic preludes for piano be adapted for guitar, for example? Could one convert the wonderful Mendelssohn violin concerto into a piano concerto? Each instrument forms a kind of grid, and inter-grid transfer of essence is the problem.

From vision and hearing, we now move to a more conceptual domain: pieces of writing. The task of compressing a piece of text one has written into fewer and fewer words forces one to struggle to define the essence of what one is trying to get across. Up to a point, a piece of text may actually be improved by having some fat trimmed here and there, just as a university or government agency can undoubtedly benefit now and then from a severe budget crunch—but this can be carried too far, and meaning will certainly begin to suffer. A fascinating exercise is to try to pack a page of one's writing into half a page, then into a quarter page, and so on, until one has gone

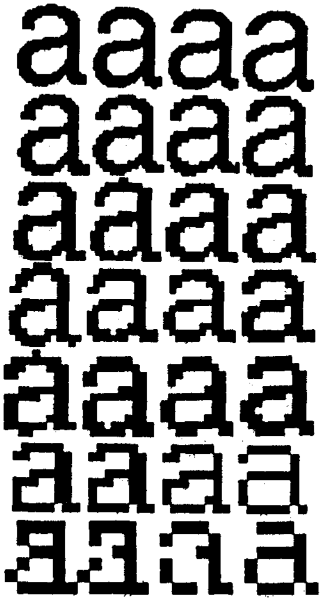


FIGURE 24–12. *Helveticality emerging from the gloom. Proceeding from bottom to top, we have a series of increasingly fine-grained dot matrices within which to maneuver. Clearly, both the 'a'-ness and the Helveticality get easier and easier to recognize as you ascend—especially if you look at the page from a few feet away. Proceeding from left to right, we have a series of increasingly letter-savvy programs doing the choosing of the pixels to light up. (As a matter of fact, the rightmost column is a very light touch-up job of the third column, done by a human.)*

The leftmost column is done by a totally letter-naïve program. It takes the curvilinear outline of the target shape and turns on all pixels whose centers fall within that outline.

The second and third columns are the work of an algorithm that has information about zones likely to be characteristic and critical for recognizability. It mathematically transforms the original outline so that the critical zones are disproportionately enlarged (the way your nose is enlarged when you look at yourself in a spoon). It then applies the naïve algorithm to this new outline (pixels light up if and only if they fall inside). This amounts to an interesting trade-off: sensitivity in the critical zones is enhanced at the sacrifice of sensitivity in less critical zones. Consequently, some pixels are turned on that do not fall inside the letter's true *outline, while some that* do *fall inside that outline remain off. It's a gamble that usually pays off, but not always, as you can see by comparing the first and second letters in, say, the third row.*

The difference between the second and third columns is that in the second column, the critical zones are crude averages fed to the program and don't even depend on the letter involved. In the third column, however, the program inspects the curvilinear shape and determines the zones itself according to its knowledge of standard letter features such as crossbars, bowls, posts, and so on. Then it uses these carefully worked-out zones just the way the second algorithm uses its cruder zones: by distorting the true outline to emphasize those zones, and then applying the naïve algorithm to the new outline.

But no matter how smart a program you are, the problem gets harder and harder as you descend towards typographical hell: matrices too coarse to capture essential distinctions. En route *to hell, more and more sacrifices are made. Helveticality goes overboard first, then 'a'-ness; and from then on, entropy reigns supreme. But just before that point is the ultimate challenge—and only people can handle it, so far.* [*Computer graphics by Phill Apley and Rick Bryan.*]

down to a phrase of only a few words. This can be seen as both a translation problem and an analogy problem. It is not usually considered either one, but just think: One is trying to "say this" in an ever sparser and tighter language, an ever more severely constricted grid.

In a similar vein, learning to write in the language called "Nonsexist" is a great exercise in translation and analogy, as is trying to become fluent in 'e'-less English, referred to earlier. Both provide you with a somewhat modified set of stepping-stones, and force you to invent and then get accustomed to many new types of constructs in order to say things that are easily said in the more prevalent mode of speaking. It is very hard to become totally fluent in either language.

* * *

A significant problem these days, related to that of capturing "Helveticality" in a low-resolution grid, is that of producing original and esthetically pleasing low-resolution typefaces—in other words, instead of imitating a known curvilinear typeface, inventing a new typeface whose natural habitat is, say, a 5×7 or 10×12 grid, all of whose letters are in "the same style" within that tiny world. Many human designers have discovered solutions of great ingenuity, but machine designers? There are none.

Letter Spirit, an AI project of mine currently on the back burner (it is impatiently waiting for Copycat to come to a boil), has as its aim to produce a program that can do just that: Given one or two low-resolution letters as inspiration, complete the alphabet in "the same style"—or rather, the same *spirit.* Instead of using pixels (points) as the primitive components of letters, however, I chose to use short straight-line segments on a fixed grid containing just vertical, horizontal, and 45-degree diagonal segments. I call those primitive segments "quanta". Figure 24-13 shows the tiny grid permitted, and the stunning variety of 'a's that one can realize within it. Actually, I estimate there are well over a thousand ways to realize grid-bound designs possessing some degree of 'a'-ness; some will definitely hit the bull's-eye while others will clearly be way out on the fringes of the 'a'-sphere. Then of course there will be many shapes that hover simultaneously near the fringes of two or more Platonic letters' spheres of influence. Such shapes are anathema to the human visual system, which greatly desires unambiguous category membership; they should be likewise antithetical to the Letter Spirit program.

The Letter Spirit grid, although seemingly a constraint, actually inspires flights of fancy that total freedom would not (a fundamental and general lesson about the deep connection between constraints and creativity). Figure 24-14 gives a sampling of a few "gridfonts" inspired by various stylistic quirks in one letter or another. Once again, this is only the tip of the iceberg. There are thousands of intriguing gridfonts to be designed and savored. As of this writing, I have designed about 150 of them. You could

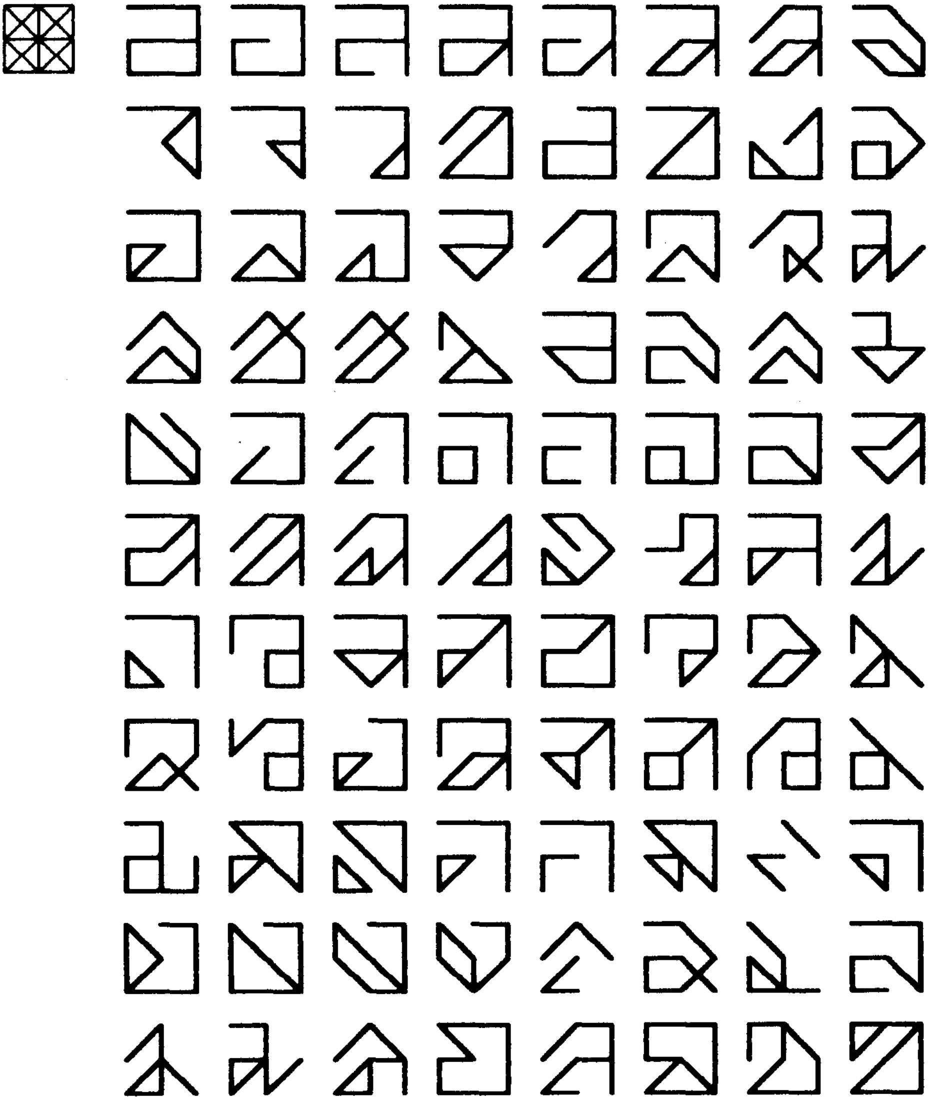


FIGURE 24–13. *87 'a's composed of horizontal, vertical, and 45-degree "quanta" in the Letter Spirit world. How many more shapes recognizable as 'a's do you suppose lurk in the given grid? (Compare this figure with Figures 12-2, 12-3, and 12-4.)*

say I'm addicted! Seven complete gridfonts by me are exhibited in this book, below the introductory paragraphs to the seven sections.

The Letter Spirit project was distilled from a far more ambitious dream: that of producing a program able to create genuinely artistic, curvilinear, full-fledged typefaces when inspired by one or more sample letterforms. I

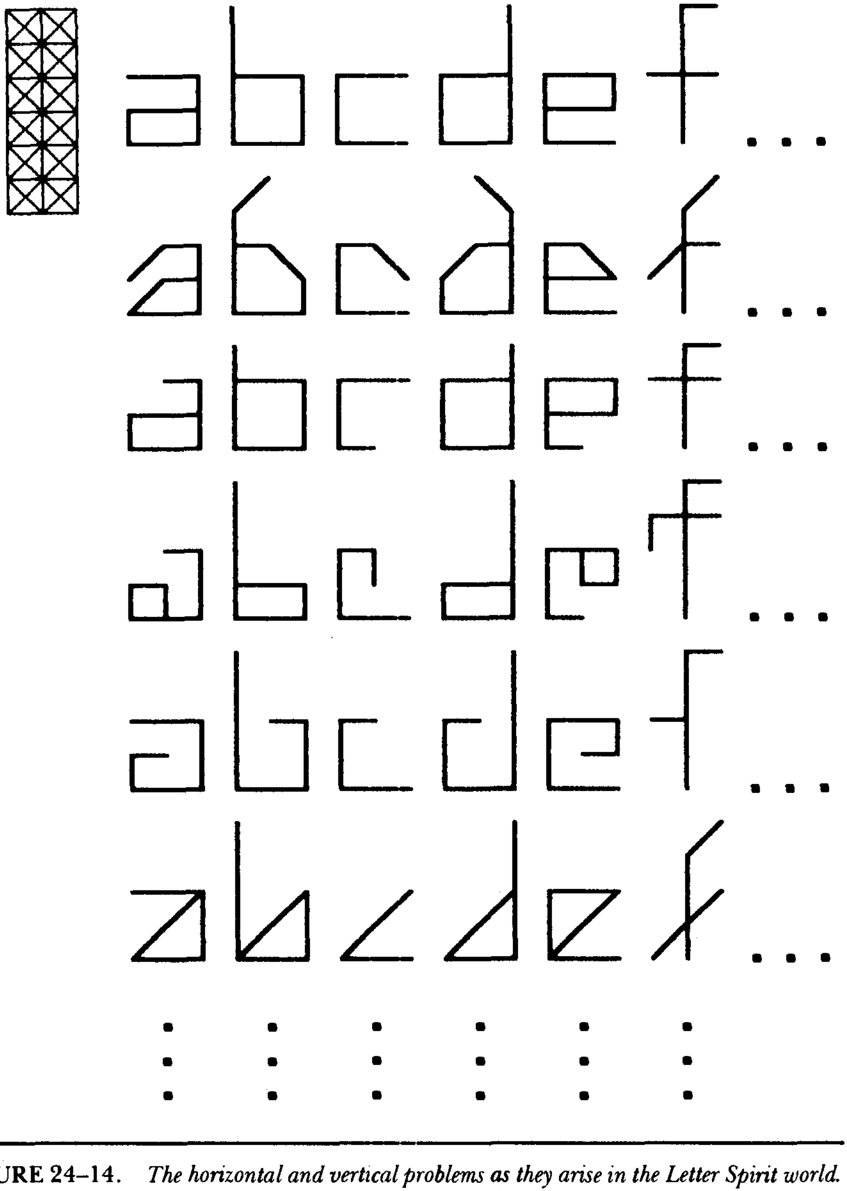


FIGURE 24–14. *The horizontal and vertical problems as they arise in the Letter Spirit world. (Compare this figure with Figure 13-8.) The central problem of the Letter Spirit project is to characterize what it is that items with "the same spirit"* (i.e., *in the same row) have in common with each other, so that in general, given a sample letter or two, the program can "get the hang of it" and then go ahead and design all the remaining letters in the same spirit, thus creating an esthetically pleasing "gridfont". Readers are encouraged to try their hand at completing the six gridfonts whose beginnings are shown here, and inventing their own.*

wrote "far more ambitious", yet in a way that's not right. After all, during each boiling-down step (and there were quite a few between the initial conception of the project and the final arrival at the grid), I assured myself that it truly preserved the *essence* of the full typeface problem and merely eliminated some superficial aspect of it. So in some sense I do believe that Letter Spirit actually encapsulates *the* central problem not only of typeface design, but indeed of art and creativity in general. And I am prepared to stand behind that claim, as long as you give me some grains of salt to defend myself with.

I recently had a fascinating visit at Bitstream, a Cambridge, Massachusetts firm specializing in the digitization of human-designed typefaces. A typical task they must do over and over again is to take a given font and adapt it from one high-resolution lattice to another (for example, from 200×200 pixels per letter to merely 100×100). For this, they have specialized graphics hardware that works just fine. This grid-to-grid conversion is an easily mechanizable analogy, or translation, task. However, when they want to take a given font and adapt it from a high-resolution lattice to a *medium*-resolution one (say, down to 15×15), their graphics machine produces an unacceptably crude solution, filled with ragged edges and spurious pixels of all sorts. To improve on this, Bitstream purchased an expensive Lisp machine and developed a complex program for this purpose. Some of the results of that work are shown in Figure 24-12. The point is, severe compression requires far more brute hardware and sophisticated software than does gentle compression. Finally, when they need to compress a font from a high-resolution lattice down into a truly coarse-grained one (say, 10×10), they turn the task over to human designers, because people alone can handle the many interacting perceptual forces that emerge at this level of resolution.

At first, this may sound counterintuitive, but really it makes perfect sense. With high-resolution grids, the graininess of the underlying medium all but disappears, and it is child's play to convert from one grid to another. It wouldn't even matter if the target grid were hexagonal, as long as it were sufficiently fine-grained. But compression down into very coarse grids forces one to deal with the conceptual and perceptual essence of visual forms—and essence, if anything, is the central problem of analogy. In fact, a sense of essence, in essence, is, in a sense, the essence of sense, in effect.

* * *

Any analogy can be viewed as an attempt to reproduce in one metaphorical grid a form that exists in another metaphorical grid. The more coarse-grained the two "grids" are, the more ingenious the analogy-maker has to be to perform the mapping. Roles and substructures must be extracted and weighed and mapped against each other. Shifts of all sorts, up and down in abstraction as well as sideways in conceptual similarity, must be able to take place. The analogy-maker attempts to judge proposed solutions for their elegance, but in the end, only their performance in the world determines their success.

The Copycat domain might appear less charming a domain than the nose-touching and chess domains, less grabbing than the Letter Spirit domain. But that is a superficial viewpoint. To make progress in science, one has to make sure that the phenomena under study are truly isolated. I am banking on having carried out the job of isolation very well, and now comes the stage of making the model. That project is ongoing, and its method of attack—its vision of how to build a system that would run on a real machine —is an esoteric and complex one. To relate that would be another very long story. It is the domain itself that has been the subject of discussion here.

I feel confident that this tiny alphabetic world allows all the key features of analogy to make their appearances. In fact I would go further and claim: Not only does the Copycat domain allow all the central features of analogy to emerge, but they emerge in a more crystal-clear way than in any other domain I've yet come across, precisely *because* of its stripped-down-ness. Paradoxically, Copycat's conceptual richness and beauty emanate directly from its apparent impoverishedness, just as the richness of the "ideal gas" metaphor emanates from its absolute simplicity. Time will tell if this limb I am out on is solid.

Post Post Scriptum.

In retrospect, it seems that this *P.S.* probably ought to have been a chapter on its own. I did not dream that it would grow to this size; I merely wanted to let my readers know what sorts of issues I am working on currently —and I discovered that sketching that out takes a good deal of time. The original column was disappointingly coolly received. I hope that this more complete explanation of the driving forces behind my research projects will awaken more enthusiasm.

Below I give our "answers" to the analogy problems given in the *P.S.* Each problem merits a much longer discussion, but life is short.

Page 579:

1. *dab* (chosen over rival *cac*)
2. *dba* (hands down, over *cbb*)
3. hard to decide between *pdt* (ugh!) and *pcu* (yuk!)
4. *pxqxsx,* of course—not *pxqxry* or *pxqxsy*

5. too obvious to need any comment
6. hmm . . . maybe *aaabbbddd,* maybe *aaabbbddk,* maybe *aaabbbccl,* maybe even *aaabbbddl*
7. *trqp,* but maybe *srqo* (definitely not *srqq*)
8. *tptqtrtt* is pretty, but so is *spsqsrst*
9. *abcdeabcdab,* of course
10. *bcdacdabc* (it's a figure-ground problem—*bcd* is *a,* in "code")
11. *acg* is way better than *acf*
12. you know. . .

Page 585:

1. *pqr,* a far more insightful answer than *pbc*

2*a*. *nws* is the only reasonable answer

2*b*. *uuuuu* (not *vvvvv,* despite the answer to 2*a*)

3. *aBc dEf pQr* goes to *aBc pQr dEf*
4. *qabcxyzq* is incisive, but has a strong rival in *abcqxyz*
5. *zabczdefzstuz*—certainly better than *zabcdefzstuz*
6. *eeeffghhiii,* and yes, that *g* in the middle is the whole point

7*a*. *dcbabcd*—not hard

7*b*. *abbbc,* based on seeing 3-1-3 go to 1-3-1

7*c*. *pfr*—a daringly abstract vision of "inside-out-ness"

When a program can do analogies *like this,* I'll be impressed!!!

Post Post Post Scriptum.

After I'd completed the *P.S.* and *P.P.S.,* I ran into Richard Feynman at a conference. I reminded him of my lecture at Caltech three years earlier; his somewhat vague recollection of it was that it was "silly". I took that as a charitable way of saying that he hadn't seen any point to it. Which made me think that maybe his "village-idiot" stance was due to genuine puzzlement, and not just an act.

I then told him, with a certain amount of trepidation, that in my new book I had humorously referred to his blunt way of answering all my analogy problems as "village-idiotic" a few times. Would this offend him? "Oh, no!" he said. "A while back, *Omni* magazine interviewed me, and on their cover they advertised it as an interview with the 'world's smartest man'. I think it's good to counterbalance that—so now you're calling me a village idiot. That's fine. I think my mother would agree with you more than with *Omni.*"