COGNITIVE SCHEMATA AND THE DESIGN OF GRAPHICS DISPLAYS

Michael I. Mills
Université de Montréal

ABSTRACT

What makes an effective graphics display? This paper argues that the problem of practical image-making can be usefully related to some basic concerns of cognitive and perceptual theory. It discusses how properties of cognitive schemata -- the flexible way schemata get "composed" during comprehension -- can help us understand some issues in pictorial representation potentially of interest to the designers of graphics systems: the comprehension of cartoons, caricatures and pictorial metaphors. The paper goes on to show how these cognitive issues suggest ways that pictures, including computer-mediated images, might be effectively used to solve basic problems of graphic communication: how to show appearance, structure, organization and movement.

SOMMAIRE

Qu'est-ce qu'une image "efficace"? Cet article relie le problème du design d'images graphiques efficaces à certaines préoccupations théoriques en cognition et en perception. Il décrit comment certaines propriétés de nos schèmes cognitifs -- la souplesse avec laquelle ils sont "composés" lors de l'acte d'interprétation -- nous aident à comprendre les problématiques relatives à la représentation picturale: l'interprétation des cartoons, caricatures et métaphores picturales. À l'aide de ces notions de cognition, l'article suggère comment les images, surtout celles créées à l'aide de l'ordinateur, peuvent aider à résoudre des problèmes fondamentaux de la communication graphique: la représentation de la forme, de la structure, de l'organisation et du mouvement.

KEYWORDS: Cognition, Schemata, Representation, Graphics

"A picture worth a thousand words must first be a good picture." -- William Bowman

When graphics designer William Bowman said this, the kinds of pictures he had in mind were not the images of "high art" but pictures intended as vehicles for practical communication. But what constitutes a "good" picture? It is easier to find examples of "good" pictures than to say explicitly why they are good. In fact, although there is a substantial literature on the psychology of pictorial representation, few authors have tried to apply their theories to questions of practical image-making (but see Arnheim, 1974). My goal in the present paper is to apply some theoretical ideas relevant to the psychology of pictorial representation to the question of what constitutes a good picture: i.e., images which are effective conveyers of information.

The first part of the paper looks at what I feel are some important properties of how our cognitive systems function when we make and understand pictures. The second part of the paper asks what insights these cognitive issues can provide for the problem of designing effective visual displays: especially those whose goal is to harness the unique strengths of computer graphics technologies.

1.0. Cartoons, Caricatures and Mental Schemata

A good place to begin our discussion of the psychology of pictorial representation is with cartoons and caricatures. Cartoons and caricatures are worth examining for two reasons. First, they are pictures which do not aim at accuracy in portrayal yet are often quite easily recognized--sometimes even more easily than photographs of the real thing. This raises a relevant question for the designer of graphics displays: namely, will the most realistic pictures necessarily be most effective? A second pertinent issue raised by caricature is how we can use pictures to think metaphorically-i.e., how imagery can reveal novel likenesses between unlike things. Interesting examples of visual metaphor are portrait caricatures which fuse normally distinct schemata (e.g., animal and human faces) in a single image so that we perceive unexpected and sometimes informative similiarities. What implications does the metaphorical use of pictures have for graphics communication?

1.1. Realism and Recognition

Let me begin with question of whether the most realistic pictures are always the most informative. In a well-known experiment Ryan and Schwartz (1956) compared people's ability to recognize objects in one of four modes of representation: photographs; shaded drawings; accurate line-drawings traced from photographs; and cartoons or "caricatures" of objects. (See Figure 1). People were shown these pictures at very brief exposure times and asked to identify some aspect of the displayed object. In the case of Figure 1, for example, subjects had to describe the relative positions of the fingers. Surprisingly, the experiment showed that the cartoons were the most quickly identified. Outline tracings took the longest time to recognize while photographs and shaded drawings required about the same time and were in between the two extremes.

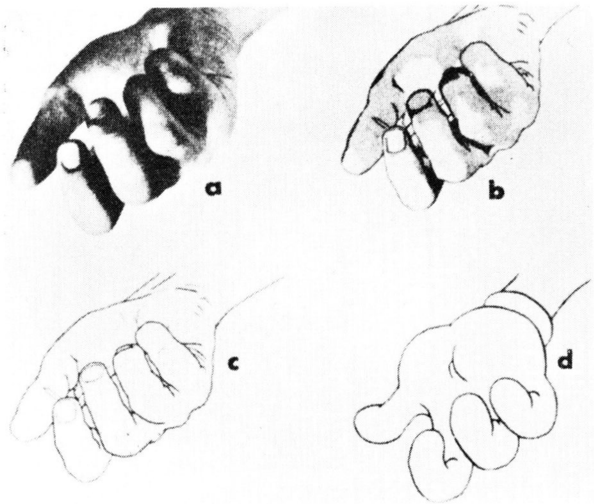


Figure 1. Source: Hochberg (1972)

How is it that the cartoon-hand--a kind of caricature--is more easily recognized than the photographic hand? After all, the photograph faithfully records the sheaf of light rays projected from the real object onto a two dimensional surface while the cartoon is obviously a "distortion" of the information in the light. One provocative explanation, advanced by Hochberg (1972), is that the cartoon is more readily identified because it comes closest to how the brain encodes and remembers what a hand looks like in the first place: the cartoon is closer to what Hochberg calls the hand's "canonical form." By canonical form I believe Hochberg means that the brain encodes its knowledge of appearances not by storing in a mechanical way the exact details of every object we perceive from every possible viewpoint, under all changes of lighting, etc.

Instead the brain is more economical: It forms concepts or "schemata" as they are sometimes called and stores these. In simplest terms, a concept can be defined as a kind of idealized or stereotypical representation which encodes the "regular" or "invariant" aspects of experience. For example, our concept of what a hand looks like might be a kind of schematized or idealized hand which records only the characteristic structural features of hands in general--not a mental photograph of the hand of any specific individual.

Recognizing an object in a particular situation, then, according to this view, would involve fitting it to its canonical or idealized representation in our brains. Now because we are constantly moving about in the visual world, objects never present themselves to our sensory surfaces in exactly the same way twice. This means that there will always be some mismatch between the canonical form or mental schema for an object and the real-world object itself. Discovering the rules by which the brain reconciles these mismatches is one of the great challenges for perceptual theory (c.f., Bregman, 1977; 1979; Minsky, 1975).

So how can all this help explain why cartoons might be more easily recognized than photographs? The explanation is simply that the closer an external representation matches its schematic form in our brains, the easier the fitting process referred to above should be. Now a photograph of a real world object will always yield a representation of an object of a specific instance--i.e., it will record information about particular traits and idiocyncratic detail. A cartoon, on the other hand, like that Figure 1, although itself a specific object, is a representation which purposefully tries to omit extraneous detail and exaggerates characteristic structural features. In other words, the cartoon strives to be more "schema-like." To the extent that cartoons succeed in mirroring the "schematizing" tendencies of our minds, the less time they should take to recognize compared with photographs because there will be fewer mismatches to reconcile between the mental concept and the specific instance.

What techniques can the cartoonist employ to facilitate the recognition of canonical form? Hochberg has described some of these with reference to the cartoon-hand in Figure 1. Compared with the detailed and "accurate" photographs and drawings, the contours of the cartoon-hand are simplified--smooth curves substituting for complex and irregular ones. The intersecting lines have been drawn at right angles thus helping the viewer to discover which surfaces overlap and which are continuous. Finally, the distance between the contours representing edges of surfaces have been exaggerated to help clarify which regions are meant to be "finger" and which are meant to be interpreted as "space." (Hochberg, 1972, p. 74). To summarize: as opposed to the photograph, the cartoon employs techniques of simplification, smoothing and exaggeration which make it easier for our brains to fit the drawing to a mental schema--to see it as a canonical hand which has undergone a "bent finger" transformation.

1.2. Metaphoric Images: The Fusion of Schemata

Figure 2 shows one of the most effective political cartoons ever made: Phillipon's famous drawing "Les Poires" (The Pears), appearing during the 1830's, which lampooned the bourgeois King Louis-Phillipe as a pear. The four drawings in the figure were part of Phillipon's defense against the accusation that the caricature insulted the King. He challenged his accusers to show which step in the series of transformations from "face" to "pear" constituted the crime. The impact of this

Figure 2. Phillipon's "Les Poires"

caricature depends not only in pointing out the visual analogy between the shape of Louis-Philippe's head and the shape of a pear, but also in knowing that in french argot "pear" means "fathead" (See Wechsler, 1982 for a detailed look at how "The Pear" and its off-shoots were used as a devastating political weapon leading up to the second french revolution).

Historical context aside, "The Pear" is an excellent example of metaphorical thinking through pictures: It uses a single image to draw our attention to novel similarities between things that usually occupy separate boxes in our mental filing systems. What account can be given of the cognitive operations involved here? Is there something special in our ability to make connections between faces and pears? Is pictorial metaphor magical -- the mysterious domain of artists and poets -- or are such metaphors constructed by using the same cognitive building blocks we use in ordinary perceiving and understanding? The answer to this question could influence our beliefs about the potential of visual metaphors to convey ideas in graphics systems.

I believe that the mental processes involved in everyday cognition and so-called metaphorical thinking are more alike than one might expect. To see this, though, will require going somewhat more deeply into the nature of cognitive schemata. A way to do this is to ask if our cognitive schemata function differently in our interpretation of "The Pear" -- a visual metaphor -- than they do in the way we interpret the cartoon-hand in Figure 1.

Earlier, I argued that the cartoon-hand facilitated recognition by omitting extraneous detail and emphasizing characteristic features, thereby making the cartoon a closer fit to the canonical or idealized hand in our minds. Let's look now a bit more closely as at the fitting process. Even though the cartoon hand has been rendered more "schematic", it is still not (nor can it ever be) a perfect match to the schematic hand in our brains. It is unlikely, for example, that our mental schema for hand specifies that the fingers are bent-- at least not the exact way as they are in the cartoon. In order to match the cartoon input to the schema in our brains, therefore, this difference between the mental schema and the cartoon hand must be identified, and then reconciled. I believe that the conceptual system accomplishes this reconciliation by a process of "composition" (Bregman, 1977). That is, it searches its repertoire of concepts for one that, when applied to or "composed" with the mental schema for hand -- the canonical form -- would yield a hand which matches the input-- i.e., accounts for the precise deviation that has been identified. One such concept is the notion "bent". "Bent" is an independent psychological unit -- a modifier -- which, when applied to the canonical hand schema, transforms it so that it matches the way the fingers are "instantiated" in the drawing. Actually, "bent" is only one of the hundreds of concepts which are probably "composed" together to eliminate mismatches between the internal schema and the external input.

I would suggest that our comprehension of "The Pear" can also be seen as involving the "composition" of schemata. Just as "bent" is a schema which can be used by the conceptual system to modify the canonical hand shape so that it will "fit" the input, so can the concept "pear-shape" serve to adjust an ideal or canonical head-shape to match the input in Figure 2. In other words, "bent" and "pear-shape" are both independent psychological units -- i.e., schemas, concepts, frames, or whatever you want to call them -- which can be composed with other psychological entities (fingers, head-shapes) to transform them in specified ways. (The importance of the notion of "composition" to cognitive and perceptual theory has been elaborated in Bregman, 1977 and applied to the problem of metaphor in pictorial comprehension in Mills, 1980, 1981c).

We are left then with the following question. If understanding the cartoon-hand and the "The Pear" both involve a process of composition of schemata, why do we see only "The Pear" as an example of visual metaphor? One answer might be that it is simply the non-typicality of the schemata chosen for composition that tips us off to Phillipon's metaphorical intent in "The Pear." In everyday vision, for example, certain compositions seem more typical than others: i.e., schemas which "typically" act as modifiers such as "bent," "twisted," "thin," "upside-down," "shadow," "top-view," whose job is to transform the properties of schemas they get composed with. What Phillipon has done is to use the "pear-stape" schema to modify a schema to which it is not normally applied -- the "face-shape." Therefore, it is simply the non-typicality of the composition which yields the metaphor.

Clearly, an answer based on "non-typicality" isn't satisfactory to explain the metaphor. For example, we could, if we wished, apply the "pear shape" schema to any concept in our cognitive cupboards. We could, for example, make pictures of pear-shaped automobiles, televisions, lamps, etc. The results may be novel, untypical, even amusing, but not necessarily metaphorical.

What makes "The Pear" qualify as metaphor is that the choice of schemas which get composed is not only unusual, but also that the similarities yielded are appropriate, not just visually, but also in terms of its semantic associations (the word in slang means "fathead.") The answer, then, to the question of whether or not metaphorical thinking is special is "yes and no." Visual metaphors do seem to be constructed from the same cognitive building blocks we use in "ordinary" thinking and perceiving in that both involve a process of composition. But metaphor seems to use the composition process, not just to mix unlikely schemas, but to do so in such a way as to ensure that the similarities produced will be appropriate or "principled." Unfortunately, how we are able to invent "principled" metaphors, whether visual or verbal, is poorly understood.

The moral for graphics systems would seem to be that while graphics technologies may provide the hardware for playing with visual schemata -- blending, transforming, etc. -- the ability to choose appropriate compositions-- to invent powerful and meaningful metaphors-- not only is poorly understood, but may be subject to strong individual differences. This does not mean that most people cannot comprehend successful metaphors. Invention is not comprehension. Nor does it necessarily mean that people could not be taught to improve their metaphoric competence. Or that graphics systems could not be used to do such teaching. These are empirical questions. After all, creative behaviour -- of which successful metaphor is an example -- can become, as Perkins (1979) has said, a "habit of invention": that is, creative thinking can itself become a "schema" or patterned way of behaving.

1.3. Metaphorical Fusion Across Words and Images

I have tried to show that the metaphorical use of imagery shares with ordinary everyday perception and understanding a basic mode of operation: the composition of schemas. Let me conclude this section by calling attention to a fact that graphics specialists should be sensitive to: that the metaphoric process does not have to occur within a single medium -- be purely verbal or visual. Metaphoric compositions can sometimes be triggered by applying words to images. Good examples are captions for abstract works of art. Think, for instance, of Mondrian's "Broadway Boogie-Woogie". The concepts evoked by the caption provide a novel set of categories with which to perceive the intersecting rows and columns of coloured rectangles. And, by the same token, the work's formal properties can change the way we normally think about "broadway" and "boogie-woogie."

You can experience for yourself the compositional process at work in the following example which comes from some recent work of mine on how people fit verbal descriptions to pictorial events (Mills, 1980). Please look at the Cartoon in Figure 3. In one of my experiments, someone described this drawing as "a non-resolvable problem you have to learn to live with." At first, this abstract cartoon may seem to have little to do with non-resolvable problems. But if you persist you will eventually find a path between the drawing and the description which may not only change the way you see the cartoon -- the meanings assigned to the abstract geometric elements -- but also the visual structure of the cartoon provides constraints which may change your usual way of defining what a "problem" is. I leave it to the reader to ferret out the connections.

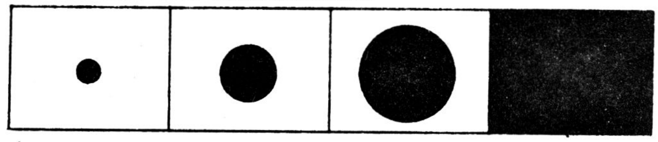


Figure 3. Source: Mills (1980)

2.0. Schemata and the Goals of Graphic Communication

In his splendid book called Graphic Communication, William Bowman has said that graphic design is concerned with the practical, as opposed to the purely personal and artistic, uses of visual imagery (although I suspect that even the most utilitarian of images, if successful, will respect aesthetic values). Bowman's book gives many examples of how the vocabulary of graphics (line, shape, texture, value, etc.) can be used to answer a set of basic questions one might ask about a given topic: WHAT, HOW, HOW MUCH, and WHERE. In this section I will examine some examples from Bowman in light of the issues raised earlier concerning the role of cognitive schemata in pictorial representation: specifically, I will ask what insights the earlier discussion of pictorial realism and visual metaphor can provide about the applied use of graphics, and suggest some ways that computer generated imagery could enhance the communication process. I will limit my discussion to two of the goals of graphics communication listed by Bowman: using images to communicate WHAT and HOW.

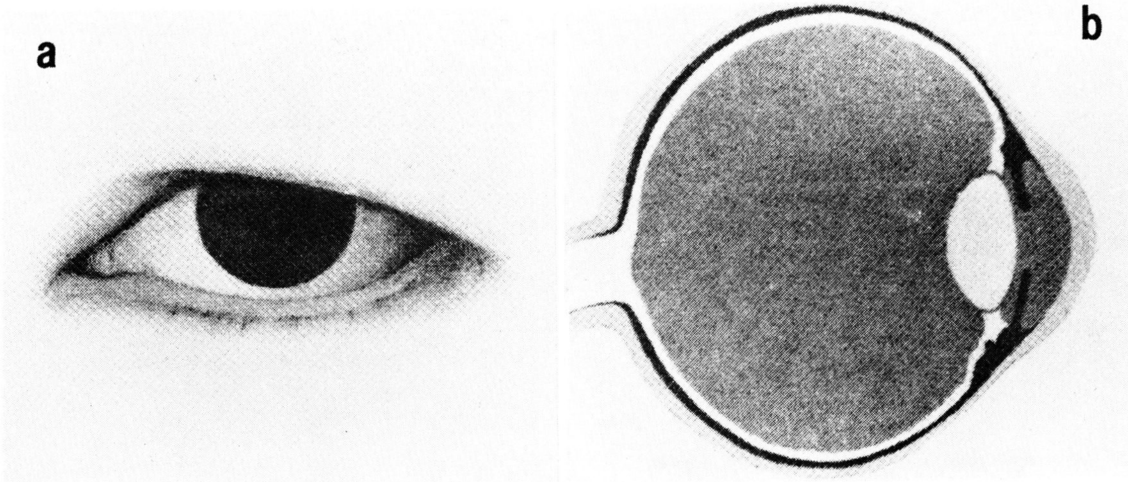
2.1. To Show WHAT

2.1.1. Representing Natural Appearance

Perhaps the first step in understanding what something is concerns knowing what it looks like: its natural appearance. There are many practical contexts in electronic or print media where one might want to use a picture to show someone what an object looks like: teleshopping, teaching, video games, and so on. The question arises: Will the most realistic and detailed pictures always be the most effective in these contexts? Our earlier discussion about the role of cognitive schemata in cartoon recognition warns that the answer to this question will not be simple--even for the representation of natural appearance. The answer will depend on the specific reason for using the picture--especially whether the goal is to convey generic structure (what is common to all members of the class as opposed to using a picture to provide information about the appearance of a specific instance -- a particular member of a class.

Consider, for example, a science lesson on the human eye--an example used by Bowman. If you wanted to use a picture to show the characteristic shape and appearance of the human eye, a full-colour, highly-detailed photograph may not necessarily be the most effective vehicle for doing so. A more simplified, cartoon-like representation which eliminates extraneous detail and uses shading and line to highlight important parts may indeed be more successful in revealing the characteristic appearance of the human eye. Of course, if the goal was to use a picture, not to show what eyes look like in general, but to show the specific details of the eye of a particular individual -- say in planning to perform surgery on it -- then the need for high-fidelity photographic detail may increase considerably.

So, the need for detail in portraying natural appearance can change depending on the task: especially, depending on whether the task is to evoke the generic structure as opposed to specific identification. Recently, William Treurniet, Paul Hearty and myself have begun to systematically investigate some of



c

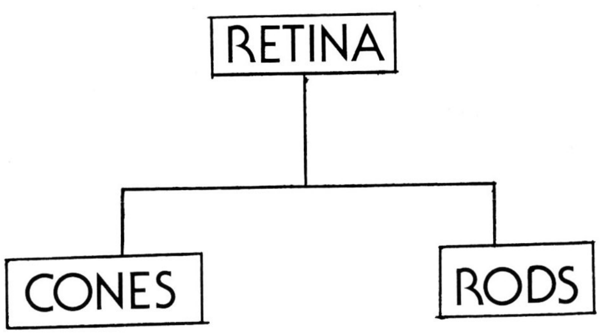


Figure 4. Adapted From Bowman (1968)

these issues in the context of Telidon pictures (Mills, 1981a). Another interesting applied area where issues of detail and realism arise is the problem of the identification of wanted criminals: What mode of representation will facilitate later recognition of a wanted criminal? A photographic mug shot or a simplified caricature? (see Baddeley, 1980 for a discussion).

2.1.2. From Outer Appearance to Hidden Structure

A deep understanding of what something is usually demands going beyond external appearances--what is normally visible--in order to reveal hidden structure. Continuing with our science lesson on the human eye from Bowman, to better understand the nature of the eye we must show more than its natural appearance. We need to represent its internal structure. What are its components parts? One traditional method of doing this is by using a cross-section diagram (Figure 4b). Note that even here questions of degree of realism are important. Thus, the cross section has omitted detail and has used line and shading to clarify the structural components. Here is a case where computer graphics techniques could enhance even further the effectiveness of the image beyond that possible in a text-book. A proper understanding of the cross section requires the viewer to perform a difficult mental transformation--i.e., it demands a complex mapping of the initial outer appearance of the realistic eye onto the representation of its internal structure: a transformation between Figure 4a and 4b. Performing such a mental transformation may be more or less easy to do--yet the success of the graphic depends on it. An animation sequence which actually portrayed the intermediate steps in the transformation could conceivably help in the mapping process. One would see the cross section slowly undergo a metamorphosis -- change shape to that of the realistic eye. In addition, it would be helpful to show explicitly in an animation sequence the simplification process--where realistic details are gradually eliminated ending in a view of the simplified cross section in Figure 4b.

2.1.3. Showing Organization: Graphic Metaphors

A third level of understanding what something is requires grasping explicitly the nature of the relations among its parts: its organization. And for this realistic images -- even those showing hidden internal structure -- will not suffice. To enhance the comprehension of relations graphically often requires tapping into the human cognitive system's ability to play with schemata--i.e., to do metaphoric thinking as discussed in the first part of this paper.

Using Bowman's example of the human eye, how could we visually portray the organization of the retina? Clearly, a realistic image would be of little help here. It is the relations which are important and they are abstract, not depending at all on what the physical structures of the retina look like. To show organization visually means finding a "schemata" whose properties are well-known and which can guide the viewer the grasping the appropriate relations. One well-known graphic metaphor which can be used is the inverted "tree" structure. (See Figure 4c). Tree diagrams are metaphors whose visible properties -- our conceptual understanding of which may stem from our earlier concrete experience with actual tree-like objects--helps us grasp the hierarchical organization of parts (even though the labels in the idea boxes are essential).

That interpreting tree diagrams involves metaphoric composition is amusingly brought home in Figure 5. The figure fuses our concepts of national stereotypes with the tree structure metaphor for organization thereby poking fun, not only at the metaphoric nature of tree diagrams, but also of our tendencies to stereotype national character.

2.2. To Show HOW

In addition to showing WHAT, a second typical problem for graphic communication described by Bowman is to use the language of graphics to show HOW: this means showing how something behaves, the movement of its parts or the chain of events in a system which constitute a process. The solving of HOW questions are especially challenging for the graphic designer since they require conveying notions of motion, causal relations and time despite the contraints of working with a still image. The challenge is to invent graphic metaphors and analogies to effectively communicate the sense of motion or dynamics in a still. Obviously, there is great potential here for computer graphics to enhance graphic metaphors for motion and process by techniques of animation.

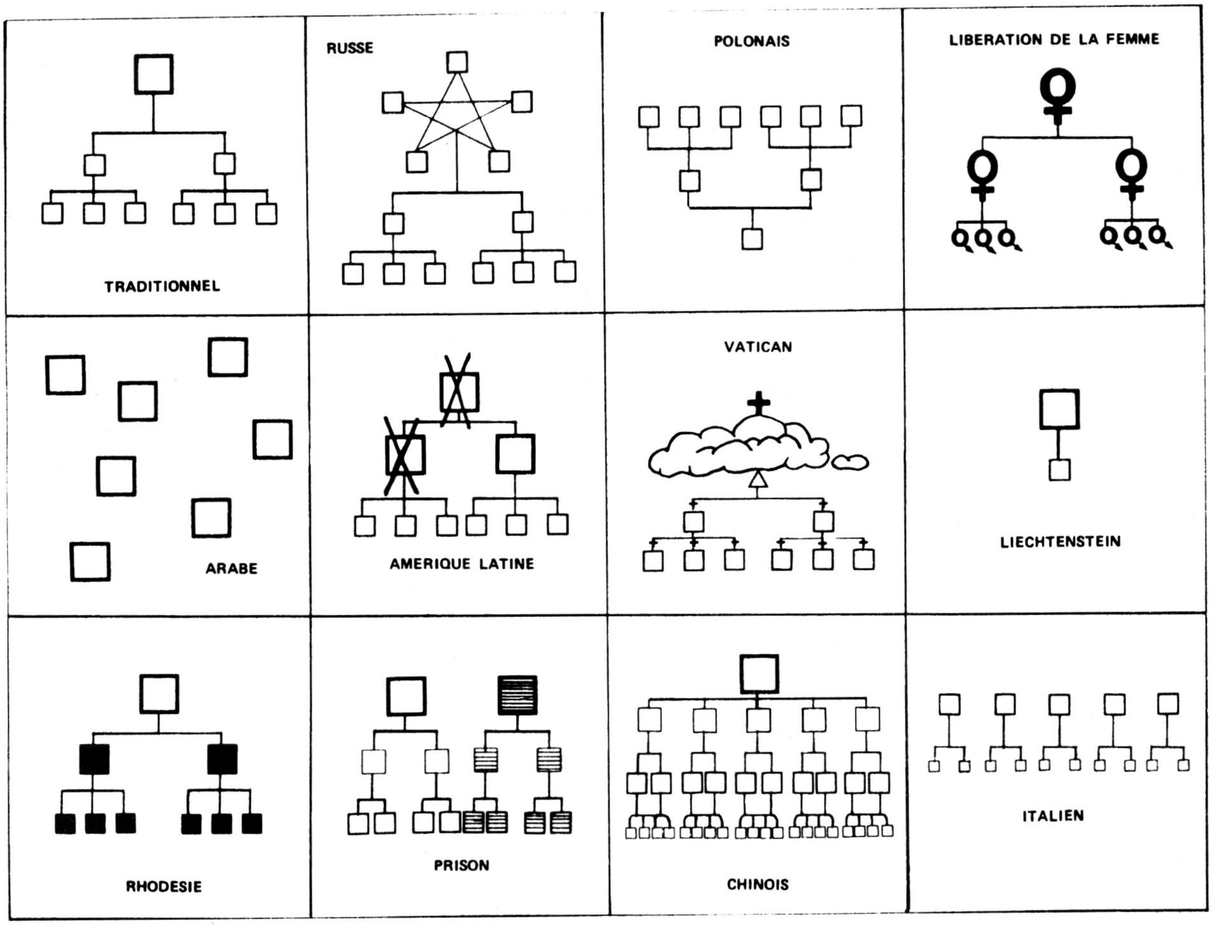


Figure 5. Source: (Unknown)

2.2.1. Graphic Metaphors for Movement

Let me consider in this final section one familiar graphic symbol conventionally used to convey movement in a still: the arrow. Bowman provides several examples of how the arrow, in combination with other graphic forms, can be used to show different varieties of motion path: "direct," "fixed," "circular," "compound," "obstructed" and so on. (See Figure 6).

I think it is important to point out the extent to which our ability to grasp these visual metaphors for movement depends on the process of metaphorical composition across words and images as mentioned earlier: i.e., where the kernel or stereotypical concepts initially evoked by the verbal and graphic elements get transformed so that they will fit together in a particular context.

That a process of composition across symbolic modes is essential to the comprehension of graphic metaphors can be seen by considering either the drawing or the verbal descriptors alone. Take, for example, the graphic in Figure 6 for the idea of "obstructed." By themselves, I submit that the meanings of dots and arrows in this figure are highly ambiguous. The graphic may indeed be worth 1,000 words, but there is good chance that without the verbal label, they won't be the same 1,000 words for the viewer as intended by the sender (Mills, 1981b). For example, are the dots "agents" --animate elements in the process of tracing a motion path described by the arrows, or are they stationary "objects" around which a single flow will divide? There are many other ways to interpret these abstract elements in the absence of the label. Similarly, in the absence of the graphic, the concepts evoked by the word "obstructed" are still not precise enough to convey the exact nature of the movement path around an obstruction: the meaning intended by the graphic figure. Taken together, however, the initial "fuzziness" of the meanings evoked by the label and the drawing is constrained. Graphic metaphors whose aim is to show HOW-- i.e., to show movement or process are good reminders of the complimentarity of words and images in the communication process --and that neither alone may be sufficient to constrain meaning.

Finally, it is obvious that graphic metaphors for movement could be improved--their meanings made even more precise--if the graphic arrows in Figure 6 were animated, perhaps even accompanied by sound cues of various kinds.

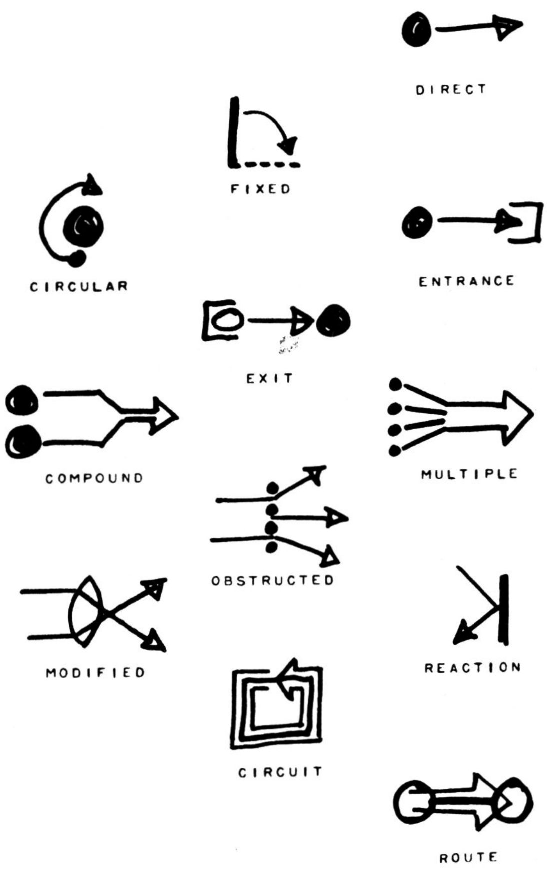


Figure 6. Source: Bowman (1968)

3.0. Summary and Conclusion

I began by asking, "What makes a good picture?" in the sense of being an effective conveyer of information through the language of imagery. Explicit predictive theories about the effectiveness of visual displays are still a long way off since students of visual imagery are still struggling to explain how pictures "represent." Nonetheless, it is still necessary to begin to build a bridge between theories of cognition and perception and the problem of practical image-making. I tried to show how some properties of cognitive schemata -- namely, the supple way in which they can get "composed" or "fused" during comprehension-- could help us to understand not only some interesting issues in the psychology of pictorial representation concerning the comprehension of cartoons, caricatures and visual metaphors but also how imagery -- including computer graphics-- could help satisfy practical goals of graphic communication: showing appearance, structure, organization and process.

References

Arnheim, R. Visual thinking. Berkeley: University of California Press, 1974.

Baddeley, A. Applied cognitive and cognitive applied psychology: The case of face recognition. In Theoretical issues in reading comprehension, Spiro, R.J., Bertram, C.B. and Brewer, W.F. (eds.), Laurence Erlbaum, ass., Hillsdale, N.J. 1980.

Bowman, W.J. Graphic communication. New York: John-Wiley Sons. 1968.

Bregman, A.S. Perception and behavior as compositions of ideals, Cognitive Psychology, 1977, 9, 250-292.

Bregman, A.S. Asking the "what for" question in auditory perception. In Kubovy, M. and Pomerantz, J.R. (eds.). Perceptual organization. Hillsdale, N.J.: Laurence Erlbaum, 1979.

Hochberg, J. The representation of things and people. In Art, perception and reality. Gombrich, E.H., Hochberg, J. and Black, M. Baltimore: The Johns-Hopkins University, 1972.

Mills, M.I. The 'less is more' theory of videotex pictures. Paper presented at the Working meeting on the social and psychological aspects of videotex. Montebello, Québec, 1981a.

Mills, M.I. Telidon behavioural research 3. A study in the human response to pictorial representations on Telidon. Department of Communications, Ottawa, 1981b.

Mills, M.I. The compositional process in fitting words to images. Paper presented at the Berkeley Cognitive Sciences Conference on Language and Mental Imagery. University of California at Berkeley, May 9, 1981c.

Mills, M.I. On fitting cartoon-strips to descriptions: A study of some relations between language and perception. Unpublished Ph.D. thesis, McGill University, 1980.

Perkins, D.N. Schemata, stereotype and invention in the arts. Paper presented at the conference on Perception, cognition and representation II Concordia University, March, 1979.

Ryan, T.A. and Schwartz, C.B. Speed of perception as a function of mode of representation, American Journal of Psychology, 1956, 69, 60-69.

Wechsler, J. Caricature, newspapers, and politics--Paris in the 1830's. Studies in visual communication, vol.7 (no.4), p. 2-30.