McGILL UNIVERSITY

On Fitting Cartoon-Strips to Descriptions:
A Study of Some Relations between Perception and Language

By

Michael Irwin Mills

Dissertation directed by Albert S. Bregman
Professor of Psychology

A Dissertation submitted to the Faculty of Graduate Studies
of McGill University in partial satisfaction of the
requirements for the degree of Doctor of Philosophy
in Communications

November 1979

ABSTRACT

Three experiments were performed to examine some of the creative aspects of the interactions between perception and language. People were asked to (1) interpret cartoon-strips whose elements are geometric shapes as analogies for meaningful events, (2) evaluate the closeness of analogy between a cartoon-strip and a set of alternative descriptions and (3) use their imaginations to find analogies between cartoons and verbal descriptions. The results of the experiments were used to argue against two commonly-held views: (1) that visual interpretation can be reduced to "pattern classification" and (2) that an adequate model of the interface between language and perception can be based on the task of "sentence-picture verification." Both views ignore the potential role of imaginative processes. The present experiments were designed to give imaginative processes a chance to operate. They challenged viewers to go beyond the surface features of visual and verbal messages in order to show that many underlying concepts or "schemata" interact during the interpretation process. Such concepts were found to include not only peoples' prototypical knowledge of the conceptual relations underlying events, but also "transformations," i.e., their knowledge of how the surface appearance of events is affected by changes in their own movements.

SOMMAIRE

Trois expériences étudient le rôle des processus imaginatifs dans les relations entre la perception et le langage. Les sujets sont demandés de: (1) interpréter des bandes dessinées abstraites en tant que "analogies" pour des classes d'événement familières, (2) juger le degré de correspondance entre une bande dessinée et une série d'interprétations verbales et (3) utiliser l'imagination afin de trouver une analogie entre un dessin et une description verbale. Les résultats de ces expériences remettent en question deux notions traditionnelles des modèles de la perception et de la compréhension: (1) que la perception visuelle soit un processus de "classification" et (2) qu'un modèle adéquat de l'interface entre le langage et la perception puisse se baser sur la "vérification" comme paradigme de recherche expérimentale. Ces deux approches négligent le rôle possible des processus imaginatifs. La présente recherche privilégie les processus imaginatifs en utilisant des tâches expérimentales où des sujets doivent "imaginer" afin de découvrir des relations analogiques entre des images et des mots malgré des différences initiales du sens. Ces expériences soulignent ainsi l'importance de considérer la représentation du sens comme le résultat d'un processus interactif où divers concepts ou schèmes se modifient flexiblement en assimilant des messages particuliers.

Acknowledgements

I owe a special debt to Albert S. Bregman, my adviser, for his guidance and support throughout the twists and turns this thesis has taken since its beginnings. His critical eye helped me to look beyond the surface layers of my initial questions to a deeper grasp of the issues. I am also grateful to the members of my committee, Irwin Gopnik and Myrna Gopnik, for their wisdom, encouragement, and especially for some much needed hand-holding from time-to-time.

I have received help in various ways from a number of people and institutions. I would like to thank my friends and colleagues Jim Taylor and Paul Licker for being available when I felt the need to discuss and to hear some fresh points of view. I have learned much from their insights. Paul was kind enough to give me access to his students at the University of Ottawa for Experiment 2. Special thanks go to Nicole Leduc for helping me to work through my ideas and for putting up with the lows as gracefully as with the highs. Additional thanks go to my colleagues at the Université de Montréal, and especially to Annie Méar, for lightening my load in several ways, both concrete and subtle, during the writing of this thesis. Deep appreciation to Monique Pepin for doing such a careful job in typing this thesis.

I am grateful to the Université de Montréal for providing support and a leave of absence enabling me to work full-time on this thesis during 1976-1977. This research was also supported by a grant to Albert S. Bregman and Irwin Gopnik from the Humanities and Social Sciences Grants Committee of McGill University.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Figure		Page

LIST OF TABLES

Preface

Images and words are two great pathways into the human mind. They are powerful media through which human beings can communicate about their inner and outer worlds. How people derive meaning from words is of obvious interest to communications theorists. It has been the object of intense study by scholars in many disciplines ranging from philosophy and linguistics to recent efforts by workers in the new field of cognitive science to develop formal psychological models of linguistic comprehension. How people derive meaning from pictures has received less attention than language. However, the study of visual perception -- how the brain forms meaningful representations of the environment from the patterns of light energy falling on the eye -- has long been of interest to psychologists and, in recent years, has been studied by researchers in the area of artificial intelligence whose goal is to someday enable machines to "see."

Because "perception" and "language" are each extremely complex processes, researchers have tended to concentrate their efforts on either one or the other treating each as if it was a separate, independent subsystem of the human mind. While such a division of labor is probably necessary to some extent, it carries certain dangers. It obscures the fact that an important part of everyday communication involves describing what one sees to others. And, more importantly, it neglects the fact that words, when applied to perceptual events, can alter the meaning, not only of what is seen, but under certain conditions can change the meaning of the words themselves.

Perception and language, then, far from being separate, independent subsystems of the mind, are intimately linked in everyday communication. Yet little is known about the nature of this link. Moreover, one might argue that lack of knowledge about the nature of the interface between perception and language may impair the development of an adequate theory of either process. Can one hope to achieve a complete theory of language without taking into account how the meanings of words are tied to perceptual knowledge? Is an adequate theory of visual perception possible without considering how seeing relates to speaking and listening?

In contrast to the practice of keeping them separate, the purpose of the present thesis is to contribute to our knowledge of how perception and language interact. Part of this contribution has been to devise a research paradigm which allows one to study a wider range of interactions between language and perception than have been the focus of most past scientific research on this problem. Through the use of less restrictive experimental tasks and a richer set of verbal and visual stimuli, compared to previous research, this thesis has been able to examine some of the more creative, problem-solving aspects of the mental processes which are called upon when someone compares a perceptual event with a verbal description.

A word about the format of the thesis. An advantage of working in a new field such as communications is that one can draw upon useful ideas and theories from other disciplines without necessarily adhering to a particular discipline's traditional, and sometimes rigid, constraints on modes of investigation and style of presentation. In this thesis, I have made use of this freedom. I have drawn upon recent ideas, mainly from the areas of cognitive psychology, perception and artificial intelligence in exploring specific questions about language and perception. Some of these questions, I felt, were sufficiently novel to justify a more interpretive approach than, say, a typical thesis in experimental psychology. Thus, in Chapter 1 an "experiment" is presented in which people are asked to generate interpretations for cartoon-strips whose characters are abstract geometric shapes; they are asked to describe these cartoons as "analogies" for familiar event classes. I then use specific interpretations of the various cartoons as a means of probing the basis of the perceived analogy. The experiment also serves as a vehicle for reviewing various theories of perception and meaning and for focussing on basic issues. This interpretive approach paid off in suggesting other questions which could be investigated using more controlled methods. Hence, in Chapter 2 an experiment is described in which people were asked to rank order sets of verbal descriptions as to their relative "goodness-of-fit" to an abstract cartoon. This experiment enabled me to develop some hypotheses about the cognitive strategies underlying peoples' intuitions about the closeness of fit across symbolic modes. Finally, in Chapter 3 an experiment is reported which narrows down the problem even further; it investigates, in a controlled setting, the psychological reality of some of the component mental operations that were hypothesized in earlier chapters as playing a critical role in fitting a visual event to a verbal description.

In sum, the aim of the thesis has been to deepen our understanding of the link between language and perception. It has done this by showing more clearly how interactions across symbolic modes can be usefully studied as a kind of "problem-solving" where people are often called upon to go beyond the information given in either a visual event or a verbal message in order to find a meaning which fits them both.

INTRODUCTION

The purpose of this introduction is first to give a brief definition of analogical thinking and secondly, to indicate why the problem of how people find analogies across symbolic modes is worth studying.

A. What is Analogical Thinking?

Analogical thinking can be broadly defined as the ability to discover how two things, ordinarily considered to be different, are alike in some previously unnoticed way. Examples of analogical thinking are not hard to find. They abound in everyday language and can be observed in the solving of problems of many kinds, the development of scientific theories and in the play of children. Here are some examples: models (the blood system is like a water pump, with arteries and veins like tubes...); metaphors ("man is a wolf"); and pretending (using a cereal bowl as a soldier's helmet in a game). Although the best examples of analogical thinking may come from scientists, artists and children, there is no reason to believe that the ability to recognize novel resemblances is a specialized skill found only in a small subset of the population. On the contrary, the widespread use of analogical thinking in everyday communication suggests that the ability to discover new resemblances is a universal cognitive skill.

Despite its apparent universality, however, the ability to create and comprehend novel resemblances is not well-understood and has received relatively little attention from those engaged in the scientific study of human symbolic processes.[1] Why should this be so? Perhaps part of the reason is that, traditionally, students of thought, language and perception have never felt quite at ease with processes such as imagining, imitating and pretending. These processes entail a kind of playfulness in manipulating symbolic representations that seems opposed to the logical, literal and rule-governed tasks which are the focus of most research in language comprehension, concept-formation, problem-solving and pattern recognition. For many, it is as if analogical comprehension requires a temporary suspension of the normal operation of the human classification system. Instead of doing its proper job of sorting things into their logically distinct categories, it is as if the classification system is playing with its own concepts -- hopping about from one to the other, merrily seeking correspondences between things that don't belong together: bowls and helmuts, hearts and pumps, human faces and cloud formations, and the like.

B. <u>Analogical Thinking Mistrusted</u>

While acknowledging that this type of mixing and matching can serve useful purposes in artistic creation and perhaps even in more "serious" pursuits such as problem-solving or building scientific

---

[1]While there has been comparatively little research into the psychological processes involved in analogical comprehension, its importance, especially in figurative language, has long been recognized (and sometimes feared) by philosophers (Cf. Langer, 1957; Black, 1962; Turbayne, 1970) and literary critics (Cf. Wheelwright, 1962; Brown, 1962; Hester, 1967).

theories, few psychologists or psycholinguists have been willing to devote themselves to extensive studies of the mental processes underlying analogical of metaphorical comprehension.[1] Rather, the typical response toward the problem of analogical comprehension has been either to treat it as a departure from the normal modes of speaking, thinking and perceiving or else to ignore it entirely, delegating the responsibility for its analysis to philosophy or literary criticism.

A good example of mistrust toward analogical comprehension can be seen in the attempt of transformational linguistics to deal with figurative language. In their semantic theory, Katz and Fodor (1963) proposed that word meaning could be represented in a lexicon where each entry would contain the grammatical category of a word, a set of distinctive features and selection rules restricting the contexts in which a word could appear. The problem with this scheme was that it disallowed many forms of figurative language because they violated the restriction rules. For instance, the verb "drinks" is marked [+verb] with the constraint that it can only take a subject [+animate] . Consequently, expressions such as "my car [-animate] drinks gasoline," while easily understood by most people, would be

---

[1] There are a number of exceptions, however. For example, Gardner (1973) has recently proposed a framework in which imagining, imitating and pretending play a central role in the young child's emerging symbolic skills. Pollio, Barlow, Fine and Pollio (1977) have examined the role of figurative language in educational and psychotherapeutic contexts. See also the work of Perkins (1978) on metaphorical perception. One of the most ambitious attempts to bring analogical comprehension into the mainstream of psychological research is the work of Verbrugge (1974) and Verbrugge and McCarrell (1977). Verbrugge's work will be discussed shortly.

rejected by the selection rules. In an effort to keep figurative language as part of semantic theory, some linguists have proposed "special rules" for handling "figurative" as opposed to "normal" uses of language. These special rules would be used to temporarily loosen the selection restrictions or to block incompatible feature sets (Cf. Thomas, 1969; Bickerton, 1969). This "dual process" approach to the problem of figurative language has been recently criticized on the basis of parsimony, the argument being that there should be a single, unified theory of linguistic comprehension -- not one for "normal" comprehension and a second for instances of "figurative" language. See Verbrugge (1974) for a discussion of this issue.

C. Is the Star Above the Plus?

If figurative language has been mistrusted as a deviant mode of communication in linguistic theory, imaginative processes in general have been totally ignored in certain important areas of cognitive psychology and psycholinguistics. A flagrant example comes from the work of Clark and Chase (1972) and Clark, Carpenter and Just (1973) whose broad aim is to understand the nature of the interface between language and perception. How language and perception are related is indeed an important question and the experimental work of these authors has contributed to our understanding of some aspects of this problem. They have suggested specific hypotheses concerning the nature of the internal code serving to link language and perceptual experience and have proposed detailed models of the process by which people compare various kinds of pictorial and linguistic inputs. Yet one must question the usefulness and generality of their approach for two main reasons: the limited range of stimulus materials used in their experiments and, more importantly, the fact that their experimental task precludes any opportunity of observing the potentially important role imaginative processes might play in theories of how language and perception are related.

In a typical experiment of Clark, *et al*, the subject is shown a display consisting of a sentence such as "The star is above the plus" followed by a picture of, say, a star below a plus sign, e.g. $\begin{matrix}+\\ *\end{matrix}$. The subject's task is to read the sentence, look at the picture and to "verify" it; that is, to say whether the sentence is true or false of the picture. The amount of time taken to respond is recorded and is used to make inferences about the component mental operations assumed to make-up the encoding and comparison processes. One problem with this task is that it is based on verbal and pictorial stimuli which constrain subjects to an extremely narrow range of meanings. In particular, stimuli such as those cited above do not specify the kinds of dynamic, changing (often causal) relations among objects which are an essential part of our experience of objects and events as mediated by language and as encountered in direct perception.

A more serious problem in my view, however, is that the use of the sentence-picture verification paradigm forces the subject into an overly narrow mode of decision-making. Verification requires the comparison of the internal representations of the sentence and the picture. An exact match will result in a "true" response being generated. If a mismatch occurs at early stages of comparison, extra operations are carried out which involve the changing of "truth indices" until the picture is determined to be either true or false of the sentence.

A model of how people compare sentences with pictures based on an overly narrow verification task also risks being overly narrow. For example, pictures which only partially matched the internal representation of the sentence would always be rejected as false. Moreover, the comprehender is not given a chance in this task to deal with mismatches in more interesting ways -- i.e., he is not allowed to mentally transform the internal representation of the picture to see how it could match the sentence given certain assumptions (for example, by mentally rotating the entire display 180°).

In sum, the verification paradigm, which has dominated much of the research on how people compare visual and verbal events, provides no role for imaginative process. The task obliterates the opportunity to observe the active, restructuring and reshaping of mental representations which are the hallmark of an analogical and metaphorical thinking.

D. Words and Images Interact

There is a need to conceive of the problem of how language interacts with perceptual experience in a more flexible way than allowed by the task of sentence-picture verification. In fact, outside the psychological laboratory it is difficult to find examples where one is called upon to merely examine the truth or falsity of a picture. More often, verbal descriptions are applied to pictures in specific contexts -- scientific, educational, artistic -- in such a way as to call upon processes other than judgements of truth or falsity. For a given type of visual display, a short list of some of the processes which verbal labels might trigger include: directing the viewer's attention to a single element among many ("look at Aunt Fanny's hat"); highlighting common properties ("find all the oval-shaped objects in the picture"); pointing out relations ("building B is taller than C but shorter than A"); providing explanations for depicted actions ("the man is smiling because the package he is holding contains a diamond"); helping one to perceive formal structural relations in art ("the two embracing figures are fit into a regularly shaped block").

Note that the direction of influence is not always one-way, from label to picture. Sometimes applying old words to familiar objects in new ways can affect the meaning of both. Imagine a painting of a vicious looking wolf captioned "mom." Not only might the label cause one to see the wolf as possessing certain human traits, but also the picture might modify the concept "mom" making it seem more wolf-like. Here we seem to be back in the realm of metaphorical comprehension, but across symbolic modes. Other examples of how words and images can interact come from modern art. Consider Mondrian's "Broadway Boogie-Woogie," a painting which shows rows and columns of brightly colored squares against a pale background. The caption refers at once to a place, a style of music and a way of life. It provides the viewer with a set of concepts which fuse with the formal properties of the painting transforming it into a kind of visual metaphor: The forms and colors take on the pulsing rhythms of jazz and the flashy brashness of times square. But note that this fusion is not just one-way. The caption provides a schema for processing the painting but, by the same token, the straight rows and alternating squares of different colors -- the formal qualities of the work -- provide a schema which modifies our concepts of Broadway and Boogie-Woogie. The painting induces us to think of these concepts in novel ways.

Now the bare fact that language influences perception is hardly news to psychologists or, for that matter, to linguists (e.g., Whorf, 1956). Any introductory psychology text will provide illustrations of how an ambiguous drawing can be perceived in either of two ways depending on "set" -- that is, providing someone with the name of one of the two objects prior to exposure of the drawing. An example would be using a label to bias someone's perception of Wittgenstein's well-known "rabbit-or-duck" drawing. Moreover, Clark <u>et al</u> have shown that it is not only language that can influence perception, but vice-versa. They have demonstrated, for example, that people prefer some spatial encodings of visual displays over others and that this will be reflected in their verbal descriptions: e.g., spontaneously describing the vertical spatial relation between two objects as A is "above" B, even though B is "below" A is also possible.

But while the problem of how language and perception are related has not been entirely neglected in cognitive psychology, there is still a gap to fill. There is a need for an approach that is more flexible than the precise, but overly narrow model of Clark and his colleagues, but at the same time promises to give a more complete account of the processes involved than the shop-worn demonstrations of the effects of language on ambiguous drawings. Ideally, such an approach would allow for the generation and testing of specific hypotheses, yet be open enough to include the role of imaginative processes in the study of how language interacts with perception.

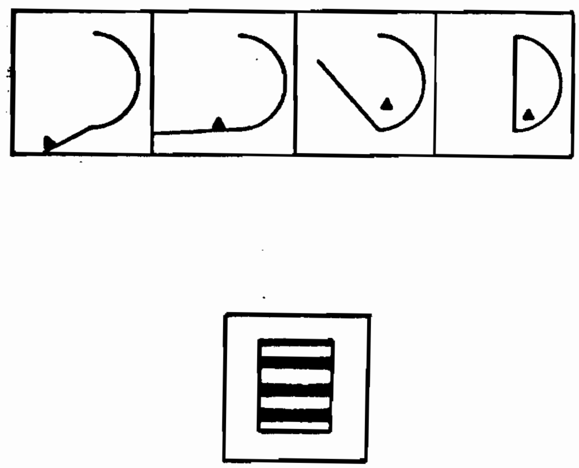
E. Cartoon-Strip Analogies

I have argued that past research on the relations between perception and language have neglected a very important area; namely, the role that imaginative processes might play when linguistic descriptions are applied to perceptual events. In the present thesis, I wish to introduce an approach where imaginative processes are the main concern. Moreover, my aim is to use a richer set of verbal and visual stimuli as the basis for experimentation and theory-building than the limited kinds of materials used by Clark *et al*. Ideally, these stimuli should specify dynamic, changing relations. The reason for insisting on the dynamic dimension is that we exist in a world of changing, causally related events and our visual and linguistic concepts reflect this fact.

An approach which could be adapted to meet these requirements was suggested to me by a recent set of experiments by Verbrugge (1974) on the comprehension of analogies. In one of his experiments, Verbrugge had people watch a series of slides containing phrases such as "An empty prison cell is like..." Each phrase would be followed by either of two animated cartoon sequences projected onto a screen. One pair of cartoons from the experiment is shown in comic-strip form in Figure 0.0. Each cartoon in the Figure was designed to focus subjects' attention on a particular set of relations concerning the topic of the phrase. For example, a subject viewing the phrase, "An empty prison cell is like..." and then seeing the cartoon in Figure 0.0(a), according to Verbrugge, would have his attention drawn to a specific kind of abstract relation based on his tacit knowledge of prison cells. This relation might be paraphrased as follows: "Just as the open form in the cartoon ensnares the triangle into its cavity by closing its movable segment, so an empty prison cell is a kind of cavity which could ensnare a prisoner by shutting its doors." Had the subject seen (b) in Figure 0.0, however, he might have experienced a very different set of relations about prison cells; i.e., the fact that one sees an alternating pattern of bars when looking through them.

To test whether different cartoons actually caused different relations to be predicated of the same verbal topic, Verbrugge used a type of memory task known as prompted recall. In one condition, a subject trying to recall the verbal topic he had seen earlier (e.g. "an empty prison cell...") would be shown a <u>relevant</u> prompt; i.e., a

(a)



(b)

Figure 0.0. - Cartoons from Verbrugge's experiment on the comprehension of analogy. See text for explanation.

Source: Verbrugge (1974)

cartoon depicting the same kinds of changing relations as the cartoon he had seen with the phrase. However, although the dynamic relations <u>among</u> the elements were the same as in the original, the specific kinds of visual elements were different in the prompt cartoon. For example, in the relevant prompt for the topic "an empty prison cell..." a circle would be seen being entrapped by a rectangular form instead of a triangle being ensnarred by an oval object as in Figure 0.0(a). In the <u>irrelevant</u> prompt condition, a subject also having seen the topic "an empty prison cell..." paired with (a) in Figure 0.0, would receive a cartoon prompt. But this time, the prompt would show an event or scene similiar to (b) in Figure 0.0 -- that is, showing a pattern of alternating bars. An irrelevant prompt, therefore, specified the <u>other</u> type of relation which had been used originally to guide another group of subjects' comprehension of the verbal topic. Verbrugge found, as predicted, better recall for the relevant than for the irrelevant prompts.

The difference in performance between subjects receiving relevant and irrelevant cartoon prompts was taken by Verbrugge as indirect evidence of the power of cartoons to converge subjects' interpretations of the verbal topic onto different kinds of relations. Further, he claimed that these relations are "abstract." Otherwise, how is one to explain the ability of the relevant prompts to remind subjects of the original cartoons? Remember that the original and prompt cartoons used different kinds of visual elements: what they had in common were the relations <u>among</u> elements. Thus, although one can argue that the basis of resemblance may be a physical pattern, it must be a higher-order or "abstract" pattern of relations among elements, not limited to simple physical features of individual objects.

The main point of the experiment was to argue that our knowledge of concepts such as "prison cell" affords a comprehender a potentially infinite variety of features and relations depending on the context in which it is placed. The experiment shows as well that this context can be a visual event. That is, one set of relations occurs when prison cell is juxtaposed with the "entrapment cartoon" and another set with the cartoon depicting "looking through bars," and there may be a potentially unbounded set of relations and properties of prison cells that could be evoked by pairing the concept with other kinds of visual events.

Based on this and similar experiments using verbal analogies, Verbrugge argues against theories of analogical comprehension which hold that concepts can be defined by a finite list of criterial features or verbal associates and which reduce the comprehension of analogies and metaphors to the search for, and pairing of, features or verbal associates shared by the two concepts being compared. (Cf. Osborn and Ehninger, 1962, for an example of such a model).

F. The Present Thesis

Verbrugge's experiment is useful in three respects. First, it challenges traditional views of comprehension (metaphors included) which do not take into account the interactive nature of our concepts

-- i.e., that different and sometimes new meanings can emerge by juxtaposing superficially different concepts. Second, the experiment shows that the search for novel likenesses can happen across linguistic and pictorial modes. Third, the use of cartoon sequences emphasizes the importance of the temporal, dynamic component of events. But the demonstration that language and perception can interact in the search for resemblance, while important, is only a beginning. One must probe more deeply into the nature of these interactions? If they are not well characterized by the search and pairing of common associates or features, what else is there? Are there alternative theories of comprehension better suited to the problem of imaginative processes? If so, how might these alternatives help us better understand the kinds of processes activated when someone seeks to fit a visual event to a verbal description?

The purpose of this thesis is to propose some answers to these questions. The key assumption is this: Finding a fit between a verbal description and a visual event can be usefully thought of as a kind of problem-solving. Thinking of this task as a type of problem-solving has certain advantages. First, is that we already know something about it. There is a vast literature in cognitive psychology and artificial intelligence on how people solve various kinds of problems -- including certain kinds of analogies (Cf. Newell and Simon, 1972; Evans, 1968). But more importantly, problem-solving does not have to be thought of as applying narrowly to specific domains. On the contrary, several authors in recent years have stressed that the kinds of issues with which problem-solving deals are quite general and can be used to characterize processes which traditionally go by different names -- perception, memory, thinking, learning -- but which share deeper similarities. These authors would argue, for example, that such apparently diverse activities as perceiving a room, hitting a baseball, understanding a sentence, planning a vacation, all involve aspects of problem-solving. Each of these activities requires using general frameworks or structures which guide the interpretation of, or adaptation to particular situations or events. These general cognitive frameworks have been given different names in the literature: "schemata" (Piaget, 1947; Bartlett, 1932); "frames" (Minsky, 1975); "ideals" (Bregman, 1977) or just plain concepts.

Although the names are different, these terms all refer to the same beast and are meant to deal with the same kind of problem. Frames, schemata, ideals, and the like, are meant to convey certain views about how our knowledge is packaged. Roughly, the claim is that our knowledge is packaged in "stereotypes" or "formulas" which capture important regularities or generalities about objects, situations and actions. Without such "formulas" life would be difficult indeed because each novel scene or experience would have to be confronted anew, dealt with from scratch. Instead, when we encounter a new scene or novel situation, we have a "formula" all ready to go telling us what to expect or how to behave.

But one should not think of these formulas or schemata as rigid, unyielding structures. There will always be some mismatch between the formula and the specific aspects of the situation it confronts. If these frameworks are too rigid, our ability to adapt to constantly changing situations would not be as smooth and efficient as it is most of the time. Aware of this, the authors mentioned above (notably Bregman and Minsky) stress that our formulas for dealing with the world must be extremely flexible. And an important part of being flexible means knowing what to do -- how to change the formula -- in case of a mismatch between what was expected and what was actually encountered.

A simple example from everyday life would be entering someone's house for the first time. Although you might never have been inside this particular house before, your "typical house" formula guides your expectations as to what lies behind the door: a hallway of a certain size and shape, various kinds of objects and their likely locations. Once inside, however, the "typical house" formula must be modified to fit the particular circumstances. Perhaps there is a sunken living room right off the hallway. Encountering this mismatch should not cause one to freeze in one's tracks. Instead, one smoothly negotiates the stairs to the living room. The smoothness of the behavior was only possible however because the "typical house" formula was flexible enough to deal with an initial mismatch; it was able to modify itself by calling upon another kind of formula -- "sunken living room" -- which describes the deviation from the norm.

This simple example illustrates why formulas or concepts are intimately related to problem-solving: they not only capture important generalities but are also necessary for dealing with mismatches or problems that arise when trying to accomplish some particular goal.

It might seem that this brief excursion into schemata, formulas and problem-solving has taken me far afield from the original goal of studying imaginative processes. But I claim otherwise. I believe that current notions of problem-solving provide a rich set of concepts for exploring the nature of imaginative processes and their role in how people discover resemblances between linguistic and perceptual events. This is not the place to defend this argument in detail. That is the purpose of the chapters which follow. Instead, let me offer a preview which illustrates how the study of abstract cartoon-strips as a kind of problem-solving can further one's understanding of the nature of the interface between language and perception.

G. Cartoon-Strips as Problems: A Preview

In everyday communication, perception and action, cognitive frameworks combine and modulate each other so smoothly and rapidly that we are seldom aware that these processes involve aspects of problem-solving. This is especially true of visual interpretation. Therefore, in order to see how visual interpretation involves problem-solving it is necessary to challenge it; to slow it down somewhat by making it do more work than usual. To accomplish this, the medium of abstract cartoon strips is very useful. It provides us with a kind of artificial mini-world which can be manipulated and controlled in order to reveal more clearly the effects of schemata, strategies and assumptions in the interpretation process.

One kind of manipulation that cartoon-strips make possible, for example, is to leave out or render ambiguous certain kinds of information necessary to perceive events as meaningful. One can then observe what kinds of strategies subjects will call upon to deal with the missing or ambiguous information. Consider the cartoon-strip in Figure 0.1. This drawing shows a half-circle and a blob-shaped object in the first frame. In the frames that follow, these objects are depicted as undergoing various changes of state. Assuming that an observer's task is to see how the cartoon could be an "analogy" for a meaningful event, this drawing presents him with somewhat of a problem. Usually changes of state in a meaningful event have a causal source or agent. In this cartoon, however, although the state-change is depicted (i.e., the correlation of the gradual disappearance of the blob and the darkening of the semi-circle), the causal factor is not specified; it is ambiguous. The problem the observer faces in order to see the drawing as representing a coherent event is to resolve this ambiguity -- to find a formula or schema (or structure of schemata) which accounts for the shapes and the changes they undergo.

In a pilot study, one subject described the drawing as "water emptying down the drain in a swimming pool." This solution is

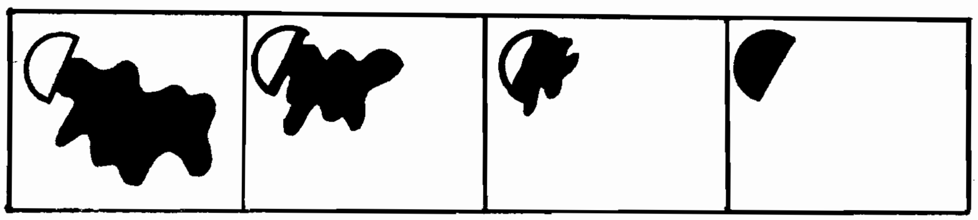


Figure 0.1. - A cartoon-strip problem from Experiment 1.

interesting, not only because it solves the problem of the missing causal agent, but also because it assimilates the visible features of the cartoon in a somewhat unusual yet highly efficient manner. For example, in this interpretation, the white area surrounding the shapes is not seen as "empty space," but is assigned the role of a "solid surface" -- a "pool bottom" whose edges presumably extend beyond the borders imposed by the cartoon frames. If the white space is assigned the role of a "pool bottom," the contour which outlines the half-circle is then constrained: it must be seen as belonging to the solid pool surface -- not as defining the semi-circular shape of a solid object. In other words, the half-circle is transformed into an "empty space" -- a hole -- in a solid object into which water flows. Note that the causal factor in this interpretation is "gravity:" a concept which neatly accounts for the initially ambiguous change-of-state of the visible shapes, but which has, itself, no corresponding visible element in the cartoon.

This example shows that fitting a cartoon strip to an interpretation can challenge someone to go beyond the information given: to make assumptions and to call upon unseen concepts which provide explanations about the underlying causes of a surface event. And because concepts are diverse and can combine flexibly to account for the visual input, many alternative solutions are possible.

In Chapter 1 an experiment is described in which people were asked to find interpretations (solutions) for a number of abstract cartoons such as that in Figure 0.1. The purpose of the experiment is to generate a rich kind of data base conducive to generating hypotheses about the nature of the interpretation process. My strategy in Chapter 1 in exploring the interpretation process will be to imagine a hypothetical cartoon understanding device and ask: What kinds of rules, knowledge and processes would it need in order to accept an abstract cartoon strip as input and generate the kinds of interpretations that people do? Obviously, I am not going to discuss the "hardware" necessary to build a cartoon understander, but rather use the idea of a device as a convenient way of structuring a discussion of the issues concerning resemblances, schemata, concepts, and so on. Chapter 1 will also serve as an historical introduction to later chapters in that past research is brought to bear on specific problems raised by the data including: (1) The problem of multiple interpretations of the same cartoon; (2) The problem of "metaphorical" descriptions and (3) How the linguistic and perceptual systems are linked in this task.

While finding an interpretation for an abstract cartoon strip is already a kind of analogical reasoning or problem solving, the viewer's conceptual system can be challenged even further. For example, instead of just having to come up with a meaning for a cartoon, a person can be asked to rank order a number of possible solutions (interpretations) as to how closely their meaning fits a cartoon-strip. Consider the cartoon in Figure 0.2. Which of the following three phrases would be ranked best, next-best or worst fit to the cartoon?

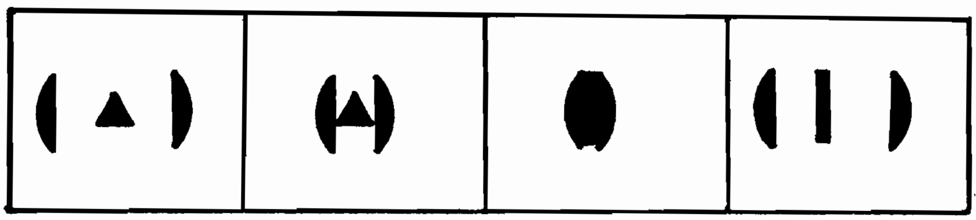


Figure 0.2. - A cartoon-strip from Experiment 2.

(a) ironing a shirt

(b) slipping on a banana peel

(c) minting a coin

In Chapter 2 an experiment is described in which people are asked to compute the "goodness-of-fit" between a cartoon and sets of verbal descriptions such as in the example just given. The purpose of the experiment is first to determine the strength of agreement across subjects in their goodness rankings and secondly to develop some hypotheses as to the strategies used by the conceptual system in computing the relative closeness of analogy across modes.

One of the hallmarks of problem-solving is that, given repeated experience with the same type of problem, people often get better and faster at finding solutions. Why should this be so? One reason is that they may discover a rule, procedure or strategy which they believe can be applied to new problems in order to arrive at solutions more efficiently. Discovering such a rule constitutes the activation of a formula which can be readily called upon for dealing with novel but similar problems.

The experiment reported in Chapter 3 is designed to show that this general feature of learning-by-doing is applicable in the context of fitting cartoon events to verbal descriptions. The basic idea is this: if over a number of trials in which a subject tries to figure out how to eliminate a mismatch between the typical meaning of a cartoon and the typical meaning of a phrase, he discovers that a particular kind of mental operation succeeds in eliminating or accounting for the mismatch, then the procedure or operation itself should become a formula or schema which will generalize to new problems.

The purpose of experiment 3 is to demonstrate this by inducing in subjects a "set" or "expectation" concerning how their assumed viewing angle must change in order to match a cartoon to a verbal description. An expectation about assumed viewing angle can make a cartoon more difficult to fit to a verbal description under certain conditions. I therefore demonstrate the existence of a bias about an observed viewing angle by making it interfere with the interpretation of a later cartoon.

This preview has suggested that the medium of abstract cartoon-strips can provide a useful means of exploring how people use their imaginations to solve various kinds of problems: finding meanings for cartoon events and fitting cartoons to verbal descriptions. In concluding this introduction, I would like to emphasize that although I refer to these tasks as involving "imaginative processes," I believe that the kinds of knowledge evoked by the experiments to be described below are general; i.e., are the same processes people draw upon in everyday life when communicating about what they see.

Having provided the preview of what is to come, let us now turn to the first problem: How people derive meanings for abstract cartoon-strips.

CHAPTER I

THE COMPREHENSION OF CARTOON-STRIP ANALOGIES

A. Experiment 1 -- Introduction

Please look at the cartoon-strip in Figure 1.0. What kind of event does it remind you of? How could the drawing be an "analogy" for a familiar situation or everyday happening? Most likely you are able to come up with at least one interpretation of the cartoon, perhaps seeing it as a person ducking behind a rock or as someone diving into water. One would guess, in fact, that most adults possess the cognitive competence to find a meaning for an "abstract" cartoon-strip such as Figure 1.0. The main goal of this chapter is to develop some hypotheses about the human conceptual system's ability to perceive these visual patterns as representing familiar event classes. What basic cognitive building blocks are necessary for making sense of abstract cartoon sequences? Exploring this question will be a useful prior step to examining the problem, in later chapters, of how language and perception interact when someone attempts to find an analogy between a visual event and a verbal interpretation.

What is the best way to learn about how people derive meanings for abstract cartoon-strips? Because so little is known about the problems of visual meaning, especially in the case of complex, changing events such as those depicted in cartoon-strips, a rigorous, experimental approach is premature. Instead, what is needed

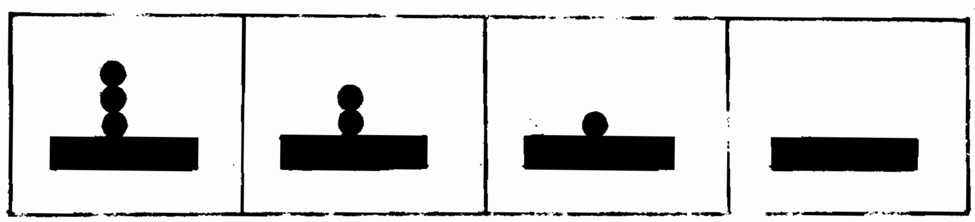


Figure 1.0. - An abstract cartoon-strip from Experiment 1.

at this early stage of knowledge is a more speculative approach that promises to generate data conducive to theory building.

My strategy will be to present a series of cartoon-strips of the type shown in Figure 1.0 to a group of people and ask them to write down the kinds of events and objects the drawings remind them of. One can conceive of such an experiment as posing an information-processing problem for the theorist: Given a set of cartoon patterns as input to an information-processing system and a set of event descriptions as output, what kinds of knowledge and processes must the system possess in order to produce the transformations? Moreover, one can use the set of input-output transformations as "data" with which to confront specific theories of perception and meaning. That is, one can ask if some theories are more useful than others in providing insights into the mental operations underlying the interpretation of given cartoons. By hinging a discussion of the strengths and weaknesses of particular theories onto specific examples from the data, Chapter 1 will serve, not only to examine the mental processes underlying the comprehension of cartoon-strips, but also as a vehicle for reviewing pertinent research on basic issues: the problem of resemblance, the nature of visual interpretation, and so on.

In addition to the basic objective of learning about how people interpret cartoon-strips as familiar event patterns, there are specific questions likely to be raised by the interpretations which a theory of cartoon comprehension should confront. As a minimum we can identify the following problems:

(1) Although the cartoons are designed by a single individual (myself) based on my own tacit knowledge of how to generate optical configurations which specify various kinds of events, they will be interpreted by a random sample of subjects from a university population. Thus, we can expect a certain amount of variation in the interpretations across subjects for any given cartoon. How might one account for the production of multiple interpretations of the same input? Do different processing strategies lead to different meanings being assigned to the same input?

(2) In a sense, these minimally specified cartoons are like projective tests such as Rorscharch ink-blots. The regularities and patterns in the cartoons may remind some subjects of ordinary everyday events, something someone might actually observe in the world (e.g., stepping on a bug). Other subjects, however, might use the same visual patterns "metaphorically." That is, they may feel less constrained to see the cartoons as familiar events of everyday perception; instead, they may project onto the features of a cartoon so-called "abstract" ideas (e.g., "democracy," "power," etc..). What account can we give of how subjects derive metaphorical interpretations?

(3) Finally, the output of the comprehension process will be a linguistic description. What implications does the experiment have for theories of linguistic comprehension? How are the linguistic and perceptual systems connected in the understanding of the

cartoons?

B. Method

Materials. Ten cartoon sequences were designed to represent a range of visual events (see Appendix B). Starting with the left-most frame, each cartoon depicts four "frozen" moments of what was meant to represent a continuous event. Comprehension of the cartoons therefore required subjects to understand that the drawings were meant as a spatial translation of an unbroken temporal event where adjacent frames represent some of the states in that temporal event. Understanding this correspondence is, of course, the basis of our ability to comprehend ordinary comic-strips.

Why choose cartoon-strips as the medium for representing dynamic events? After all, events in the real world unfold through time; our experience of them therefore has a strong phenomenal component which one might call its "rhythm" or "timing" which is lost by translation into the cartoon-strip format. If timing is so important, why not make movies of real-life events which would preserve information about time and motion? The advantages of using a small set of simple, abstract forms versus real objects have already been discussed in the Introduction. Sacrificing the phenomenal experience of "timing," however, is less easily justified. For example, it would be quite difficult to represent certain classes of events by sampling only four frames. Think of an event in which an object falls to earth in a spiralling motion. Thus, not every class of continuous, dynamic event is specifiable directly by the sequence of states in the "sampled" format of cartoon-strips. Also, the lack of a specific timing component in the cartoons introduces a certain amount of ambiguity in the choice of interpretation: Is an event happening in a matter of seconds, minutes, hours? As we shall see later on, the meaning of an event can change as a function of one's assumptions about the missing timing parameter.

Despite the importance of "timing" in understanding visual events, there are several good reasons for choosing carton-strips over movies. First is that making a large number of animated cartoons is cost and labor intensive. Second, an animated film would exclude the possibility of conducting certain kinds of experiments. For example, a slide of a cartoon-strip can be displayed for a desired period of time. How long a subject spends looking at a cartoon -- trying to make sense of it -- can be measured and will be an important variable in a later experiment involving the comparison of sentences and cartoon events. If the events are animated movies whose duration is on the order of one or two seconds, the possibility of using such latency measures is lost. Third, as we have argued before, the use of cartoon-strips will make it easier to generate hypotheses about what characteristics of the visual input influenced a subject's interpretation.

The geometric shapes which make up the set of elements found in the cartoons include different sizes of circles, triangles, rectangles and half-circles. In most drawings, the figures are solid black forms. In a few cases, only the outlines of the shapes are shown in order to depict certain kinds of relations: e.g., "containment." It was attempted as far as possible to generate events that differed from each other in order to have a wide range of event classes represented. The cartoons were drawn with black felt pen on white bond paper. They were then photographed and mounted as 35 mm slides.[1]

Subjects. The subjects were 35 undergraduate students from the University of Montreal, aged 18-26. They participated in the experiment to satisfy a course requirement.

Procedure. Subjects were told that they would see a series of cartoon strips showing geometric froms undergoing various kinds of changes (full instructions are given in Appendix A). They were instructed to think about each cartoon in order to see how it could be an analogy for a familiar event from everyday experience. An example of how to read the cartoons was provided as well as a sample interpretation. Numbered answer sheets were provided such that subjects could write down more than one analogy if they so wished.

The slides were projected by a Kodak Carousel projector onto a large screen about ten feet away. The ambient light in the room was adjusted so that the slides were clearly visible from all parts of the room while at the same time subjects could see their

---

[1] All the cartoon sequences used in this and later experiments conform to the format described here.

answer sheets. Subjects viewed the slides as a group in a large classroom. Each slide was projected for about three minutes before the appearance of the next slide.

C. The Nature of the Perceived Analogy

The most general result of the experiment is that the task of finding an analogy between an "abstract" cartoon event and familiar events of everyday experience is easily performed by an average group of university students. All 35 subjects were able to derive at least one analogy based on each of the drawings; most were able to generate two or three. Appendix B shows the first interpretation generated by each subject for each of the ten cartoons. One can conclude then that the skill required to perceive meaning in the cartoons is not a special gift of the highly imaginative or artistic, but is within the cognitive competence of university undergraduates.

In the following sections we will probe the nature of this ability to discover resemblances between cartoons and familiar event classes in light of existing theories of perception and meaning. Rather than attempt to analyse every interpretation for each drawing, our strategy will be to draw upon those interpretations which best illustrate specific theoretical issues concerning the comprehension process.

Let us begin our inquiry by asking what traditional theories of perception say about how people recognize similarities among events. Traditionally, psychologists say that we perceive or respond to events or objects as similar because they belong to the same class or category. The process of assigning events or objects to one category rather than another is often called "pattern recognition." A prevalent view of pattern recognition holds that classes or concepts are defined in terms of shared lists of distinctive physical "features" or "attributes," i.e., those features common to all members of the class. Although theories of pattern recognition are usually concerned with explaining the perception of similarity among objects, the same reasoning would presumably extend to dynamic events. For example, just as our concept of "chair" might contain a set of criterial features (seat, back, legs) so a dynamic event concept such as "chase" could be reduced to some set of physical features common to all instances of chasing.

An early example of a pattern recognition system based on the idea of sorting objects into categories according to "distinctive features" is the Pandemonium model proposed by Selfridge (1959) and adopted by Neisser (1967). This system not only operates according to distinctive features but incorporates the idea that such features might be organized hierarchically. Let us look briefly at an example of how this system works in recognizing letters of the alphabet.

Upon being shown a particular letter, say "A" the Pandemonium system possess a number of detectors or "demons" which operate independently and in parallel in analyzing different features of the pattern. Demons are organized in layers such that the most basic layer responds to simple features in the pattern (lines, angles, curves); the next stage of analysis consists of "cognitive" demons which are on the lookout for the particular combination of features indicating the presence of a particular letter. In the case of the letter A, the cognitive demon would be on the lookout for the set of simple features (/,\,—) in the pattern. To the extent that it finds the particular feature set it is looking for, the cognitive demon begins to respond or send out signals along its neural pathways. Finally, a "decision" demon chooses among the cognitive demons the one which is firing or responding most strongly, and thus the pattern is classified.

The main attraction of such a model of pattern recognition is that it shows how, with a relatively small set of physical features and knowledge of relations, a system could be devised to sort a large number of items into their appropriate classes. A second attraction of such a model is that it seems to be supported by research showing the existence of neural mechanisms which appear to detect specific types of features in a sensory pattern -- edges, movement, etc... It has been shown, however, that even for relatively simple patterns like letters of the alphabet, a Pandemonium scheme runs into great difficulty, especially when one includes handwritten samples. (Cf. Lindsay and Norman, 1976). Yet the notion that concepts or "equivalence classes" are based on sets of simple physical features is still widespread.

Can a view of pattern recognition based on distinctive physical features account for the ability of subjects in the present experiment to classify the drawings as familiar event patterns? Consider the cartoon in Figure 1.1. This cartoon reminded a number of subjects of the event "soldiers forming ranks." Clearly, a search for physical features common to the concept "soldier" and the class of geometric shapes in the drawing (dots) will come up empty-handed (except perhaps for the buttons on a soldier's uniform). One might argue, however, that there is a common physical feature shared by the cartoon and the idea of "forming ranks;" i.e., the notion of "convergence" or a "coming together of similar elements." Yet note that the concept of "convergence," while involving physical elements, cannot be defined in terms of physical features of single, isolated elements. If one insists on a classification of the kind of event depicted in Figure 1.1 based on physical variables alone, one must base the classification on the kinds of "higher-order" variables which Gibson (1966, 1977) talks about; Gibson would attempt to describe the pattern, not in terms of combinations of simple physical elements, but rather as complex <u>relations</u> among elements. Furthermore, according to Gibson, such relations are often time-varying or dynamic. Verbrugge has emphasized that it is changing spatial relations over time which are critical to understanding dynamic events through cartoon-strips. A proof he suggests would be to show that a single frame from the figure would not be sufficient to specify a convergence. A further test would be to replace the dots by another set of shapes, say squares, and see if one still perceived the cartoon as specifying a convergence. If so, this would provide additional support for the view that it is higher-order or "formless"

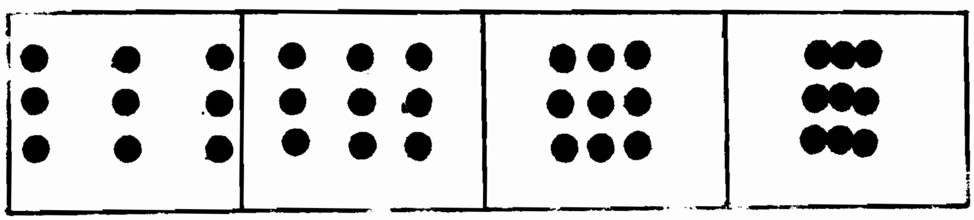


Figure 1.1 - A cartoon analogy from Experiment 1.

dynamic relations which are critical to the percept; not features of single elements.

While the higher-order abstract variables of Gibson seem better suited than distinctive physical attributes of objects as a basis for resemblance, there are still problems in reducing the explanation of the perception of resemblance to physical variables alone. This is because other "unseen" concepts seem to be playing a role in shaping the interpretations. Consider, for example, one subject's interpretation of the cartoon in Figure 1.1 as "sheep squeezing together to get through a narrow passage." In this interpretation, not only are the elements described as "squeezing together" or "converging" but other concepts are introduced into the interpretation to "explain" the convergence. That is, the subject has added an element to his interpretation -- "a narrow passage" -- which is not explicitly represented in the drawing. Narrow passage is an inference which makes sense of the coming together of the dots. The narrow passage then is an essential part of this subject's interpretation of the drawing which provides an explanation or theory which accounts for the changing relations among the visible elements of the pattern.

There are other unseen elements entering into the interpretation. For example, there is an implicit "vantage point" from which the subject imagines he is viewing the event. In the case of most of the interpretations of Figure 1.1, the assumed vantage point is from above, a "bird's eye view."

Assumed vantage point seems to be playing a role in the interpretations for many of the cartoons. A particularly good example can be seen in connection with the cartoon in Figure 1.2. One subject described this drawing as "a man diving into the sea" while a second interpreted it as "three people entering an elevator." These two interpretations illustrate the difference between what Arnheim (1974) has called "vertical space" and "horizontal space" in how someone chooses to interpret the flat pictorial plane of a two-dimensional drawing. In interpreting the cartoon as "diving" the subject is imagining the event as it might be seen from a side-view. Here the three dots are assumed to be arrayed along the vertical axis of the picture plane from top to bottom, the horizontal dimension distinguishing between right and left. This is an interpretation consistent with "vertical space." In describing the cartoon as representing "entering an elevator," on the other hand, the subject would appear to be imagining the event as it might look from a top or "bird's eye" view; i.e., in terms of "horizontal space" in which the flat picture plane is used to represent the directions of the compass (north-south, east-west). Figure 1.3 shows the difference between vertical and horizontal space diagrammatically.

Note also that choosing a particular vantage point can constrain the nature of the other concepts which shape the interpretation. For example, if the assumed vantage point is a side-view, then the disappearance of dots, one-by-one from the adjacent frames can be attri-

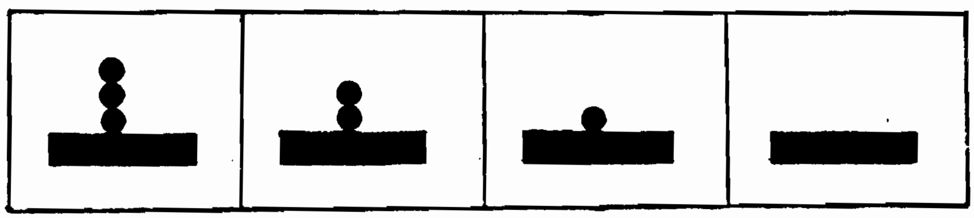


Figure 1.2. - A cartoon analogy from Experiment 1 interpreted as: (a) "A man diving into the sea" and (b) "Three people entering an elevator."

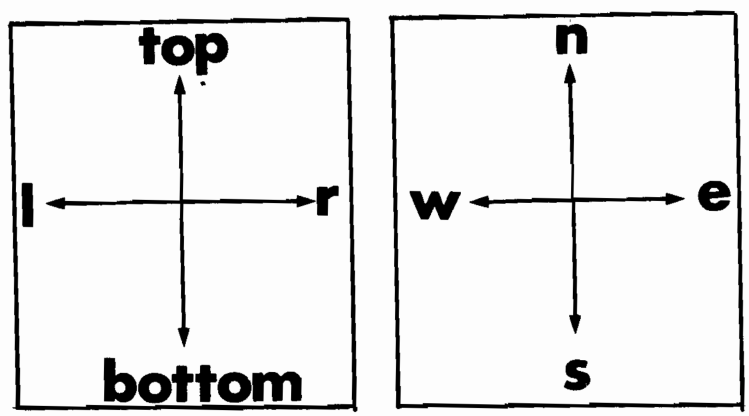


Figure 1.3. - Two ways of interpreting a two-dimensional picture plane: (a) vertical space; and (b) horizontal space.

Source: Arnheim (1964)

buted to the "pull of gravity" -- a concept which is part of our general knowledge of diving into water. However, if the imagined viewing angle is from above, then "gravity" is less easily inserted as a causal factor into the interpretation. Seen from above, the ground plane of the picture would be assigned the role of a "surface" upon which the dots would be perceived to be resting. In such a case, the downward disappearance of the dots could not be attributed to gravity pulling the objects into a mass, but another type of explanation must be sought. The "entering" solution attributes the disappearance to the autonomous movement of the dots as they pass beneath an occluding object.

In sum, these examples indicate the presence of certain unseen or viewer-supplied assumptions which interact with the visible elements in the cartoons to shape their meanings. The experiment seems to be telling us that a theory of perception able to explain the discovery of resemblance should consider possible contributions on the part of the perceiver which go beyond classification based on shared physical features or even the higher-order "formless" invariants of Gibson.

D. Perception as the Forming of a "Description"

Recently, a number of researchers dissatisfied with the "common features" approach to pattern recognition have put forward a a view of perception which seems better suited to deal with the kinds of issues raised by the interpretation of cartoon-strip analogies.

These alternatives are attractive because they attempt to take into account the possible contributions of the perceiver in the creation of meaning. An added benefit of this alternative approach is that by taking into account the contributions of the perceiver, one begins to have a clearer understanding of how the process of "seeing" is interlocked with "thinking" and "behaving" -- a subject about which traditional approaches to perception have had little to say.

One recent theory of perception which provides a useful alternative framework for examining the problem of resemblance is that of Bregman (1977). According to Bregman, a main difficulty in reducing the problem of understanding to classification is that classification itself depends on the ability of the perceptual system to describe, not only what is old or familiar in a particular situation or scene, but also what is new. Bregman has used the display in Figure 1.4 to illustrate this point. Looking at this display, the pattern of light falling on the retina gives rise to an experience of recognizable, familiar forms: A, B and C. We are able to classify these forms as familiar because we have internal representations or concepts for them. A concept can be temporarily defined as a pattern in the brain which represents a regularity or familiar aspect of experience. Bregman refers to these patterns as "ideals." While ideals represent separate, simple aspects of experience in the brain, a problem arises because they can often be in the world in complex ways, as the display shows. The problem for the perceptual system is to disentangle the sensory array in order to recover the ideal or familiar aspects which are in the scene. To do this however the percep-

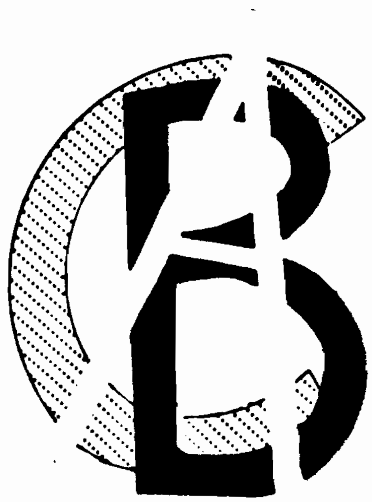


Figure 1.4. - The composition of three ideal forms.

Source: Bregman (1977)

tual system must know more than the distinctive feature set of the different forms in the scene; it must also possess "theories" or "concepts" about how such forms can be distorted or changed depending on a number of factors including the point-of-view of the observer, being partially hidden by other forms, and so on. These possible changes or distortions are what can be novel about any given perceptual experience. For example, in the scene one can only see the B if we assume that it has been "rotated" and is "behind" another object which partially overlaps it. Moreover, the overlapping object cannot be just any form but must be an ideal A whose shape, orientation and color "explain" what parts of the B are deleted. Remove one of these letters from the theory of the scene and the recognition of the other letters becomes very difficult. Recognition therefore depends on knowledge of how the two ideal forms, A and B, can interact in given situations. In other words, the recognition process, in examining the retinal image, must know about possible transformations which could account for a change or distortion of an ideal form. Thus pattern recognition depends not only on identifying and classifying familiar forms but also on knowing about permissible transformations explaining how forms may affect each other in particular contexts.

The output of the perceptual process according to this view is not just a classification of the objects present in a scene. Instead, it seems more accurate to think of the product of perception as a "description" which is formed by ideals and transformations. A description then is a mental representation not only of the old aspects of a scene but also of the particular way they are present in a given situation.

An important benefit of defining perception as "description-formation" is that it helps one understand what perception is good for. Perception enables one to adapt's one's behavior to new problems and situations. For example, in opening my refrigerator to get some butter, I notice that the butter is behind the milk and sitting on top of a loaf of bread. The scene contains several familiar elements -- the butter, the bread, the milk -- as well as the familiar relations "behind" and "on-top-of." I immediately push aside the milk, grasp the butter and remove it from atop the loaf of bread. In this example, my behavioral stream meshes smoothly with the requirements of the situation. But I would not have been able to perform such a simple task if my perceptual system was restricted to recognizing only the old aspects of the scene; i.e., labeling the objects and relations as things I have seen before. Rather, my ability to perform the task depends upon understanding the <u>exact</u> nature of the relations among the objects in this particular situation. Fortunately, this is what my perceptual system provides -- a "description" which serves as a map or model of the new situation including my own role in it. One sees from this example that descriptions serve to control actions, to generate expectations and enable one to adapt to new situations.

A good example of the usefulness of conceiving of perception as the forming of a description comes, not from research with people, but with machines. Researchers in the area of artificial intelligence called "scene analysis" have been trying to program computers to perform tasks requiring the intelligent use of information derived from perception of the environment -- i.e., where a "description" of the environment could control the machines behavior (Cf. Winston, 1975; Waltz, 1975; Guzman, 1968). As a first step toward this goal, they have been trying to get computers to perform simpler feats such as recognizing two dimensional line drawings in terms of 3-dimensional objects: blocks, wedges, pyramids. Although people can understand such scenes with ease, programming a computer to do so is extremely difficult because it requires making explicit the knowledge and processes involved.

An early program which could recognize two-dimensional drawings as representing 3-D objects is that of Roberts.[1] Roberts' program is an example of a perceptual system using what we have been calling ideals and transformations to form a description. In this program ideals are abstract definitions, based on the laws of projective geometry, of the three classes of object the system expects to find depicted in the picture (wedges, prisms, cuboids). These definitions are "ideal" because they define the class of object without being constrained to any particular size or point of view. The program also has transformational rules which allow it to evaluate whether a given picture fragment can be related to one of its abstract models through a change in viewing angle or distance. In this

---

[1]See Boden (1977) for a review of Roberts' and other scene-analysis programs which pays particular attention to the epistemological issues raised by such programs.

way the program can evaluate the goodness-of-fit between its description of an input scene and one of its stored models. Thus the program exhibits certain properties of intelligent seeing; it can "imagine" or construct an indefinitely large number of possible descriptions of an input scene knowing how parts of line segments would be deleted as a function of changes in perspective or distance. Such knowledge is essential if the program is to able to relate new scenes to old; for example, if it has stored a description of an object as seen from a top-view, then it should be able to relate this description to a new description of the object as seen from a side-view.

More recent programs have been grappling with the analysis of scenes depicting several "overlapping" objects such as shown in Figure 1.5. Interpreting such complex scenes requires providing the program knowledge about occlusion; i.e., what happens to the outline contours of one object when a second object is between it and the observer. Guzman (1968) has developed a number of heuristic rules for grouping overlapping figures into separate objects based on an analysis of the kinds of intersections found in the scene. Some kinds of intersections indicate that the regions bordering an intersection can be part of a single object; others indicate that an interrupted line can be due to an occluding second object. The possibilities for interpretation of one intersection are used by this program to limit the possibilities for interpretation of other intersections on the same line. The important point raised by the program is that,

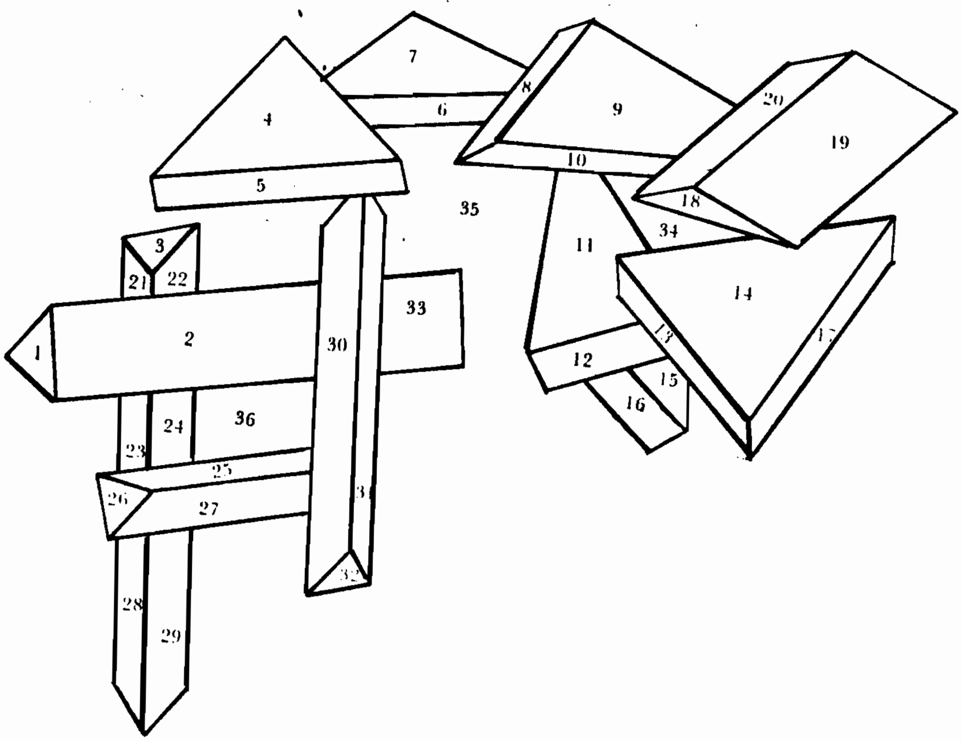


Figure 1.5. - How many separate objects? A problem
for visual scene analysis.

Source: Guzman (1968)

as with the disentangling of the intermingled letters in Figure 1.4, interpreting a scene often depends on a description of the entire scene; that is assessing the interdependencies among local features (intersections) before being able to derive an interpretation of its separate contents.

The above examples were intended to show that the problem of seeing a pattern as a familiar event or object often depends on the perceptual system's having a "theory" about how a familiar pattern could be present in a scene in a novel or complicated way. The concepts out of which these theories are built we have called ideals after Bregman (1977) in order to underscore the notion that they are the brain's assumption that a complex input may often be formed out of simple, separate aspects which have undergone transformations in a particular situation. The particular theory which accounts for a given scene or situation we have called a description.[1]

In the last two sections I have asked if the problem of seeing abstract cartoon-strips as representing familiar events could be reduced to a process of pattern classification: e.g. matching a

---

[1]A similar view of perception as the accounting for a sensory input by a "description" has been put forward by Rock (1975). Ideals bear a strong family resemblance to a recent theory of knowledge representation advanced by Minsky (1975). Instead of ideals, Minsky's unit of knowledge is a "frame" which he conceives as a data-structure encoding prototypical knowledge of objects and situations. Encountering a new situation or problem involves fitting it to its proper frame. Frames often have to undergo changes when fit to particular situations just as ideals are said to undergo transformations in particular contexts.

visual input to an internal representation based on a shared set of distinctive physical features. I have argued instead for an alternative view, based on a theory of perception advanced by Bregman, whereby interpreting a cartoon would be caracterized as a process of "description" in which viewer-supplied assumptions or "ideals" interact with the visible features of a cartoon in shaping its meaning. In the next section, I ask what kinds of basic ideals, rules or assumptions a cartoon comprehender draws upon during the description process and how these ideals are organized.

E. Levels of Description in Cartoon Comprehension

The examples I have used to introduce the description forming process have been restricted to relatively simple displays of familiar letters or simple arrangements of blocks. The ideals and transformations in these examples concerned such notions as "assumed viewpoint," "surface," "occlusion," and the like. But these concepts are not sufficient to account for the kinds of interpretation people are able to derive for the cartoon drawings. How can the notion of "description" be extended to the more complicated case of finding resemblances for cartoon analogies?

To help us appreciate the kinds of knowledge necessary for understanding a cartoon-strip analogy, let us imagine that the description process is hierarchically organized; i.e., that it takes place in a series of stages such that the results of processing at the earlier stages are available for analysis at the later stages until the final output is produced in the form of a meaningful event description.

We will see later on that there are problems with a purely hierarchical model of the description process, but for now it is worthwhile trying to sketch the various stages. Consider the cartoon in Figure 1.6. Suppose that this cartoon is input to a hypothetical cartoon comprehender -- an imaginary information-processing device for generating meaningful event descriptions of the drawings. What is the sequence of operations which will occur and what kinds of knowledge are called upon at each stage?

The first type of knowledge our cartoon-understanding device must be provided with are some rules concerning the "syntax" of cartoon-strips as a communication medium. It must know, for example, that adjacent frames represent a spatial translation of time such that spatial units denote moments in time. Part of this rule states that the temporal beginning of the event is the left-most frame and that the frames are scanned from left-to-right in succession. A second type of knowledge concerns the ability to describe the contents _within_ each frame by a structural description. Beginning with the first frame, a structural description would define the types of objects shown; their shape, size, location, orientation and also the spatial relations holding among them. This means that our cartoon-strip comprehender's visual knowledge would have to possess extremely sophisticated scene analysis programs in order for it to understand such relations as "sitting-on," "overlap," and "left-of."

Another type of knowledge the comprehender would likely possess is that the rectangular frame is itself not usually a structu-

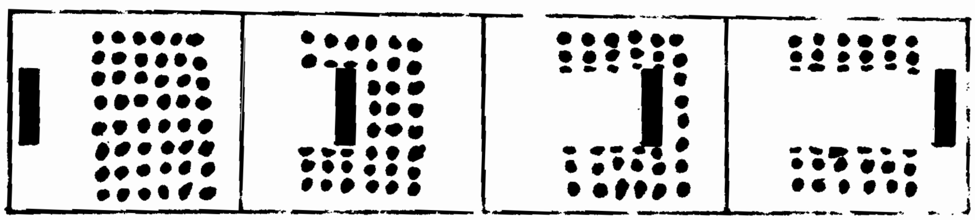


Figure 1.6. - Input to a hypothetical cartoon understanding device.

ral element in the picture but outlines a field of view. (An exception would be when a cartoonist draws an element within the frame as hiding part of the frame's border. Then the contour of the frame is seen as the edge of a solid surface which forms a space or "window" whose edge is covered by an occluding object. However, the fact that such drawings are viewed as "funny" suggests that they violate an accepted rule.)

Knowledge of the syntax of cartoon strips and rules for generating within-frame structural descriptions are prerequisites for the next stage of analysis: a description of the differences between the successive frames. Such an analysis would compare the structural descriptions of two successive frames and might ask such questions as: Are there any visual elements deleted from frame 2 that were present in frame 1? Have any new elements been added? Have identical elements changed position, orientation? This last question presumes possession of a sub-rule which states that the same or similar shape appearing in adjacent frames is not necessarily two different objects but may be a single object shown at two different times. In Figure 1.6, for example, the cartoon comprehender would note two basic changes between frames 1 and 2: (a) The rectangle has changed position from the far left of the frame in 1 to the left-of-center of frame 2, and (b) part of the array of dots has been deleted.

Once a description of the between frame changes has been generated, the recognition system might ask itself if there is any way to account for the changes it has found. In so doing it could call upon certain kinds of "change concepts" or ideals. Some of these change concepts might be very primitive. For example, "movement" is a change concept which would account for the appearance of the same shape appearing in different spatial positions in two successive frames. In Figure 1.6, for example, we infer that the rectangle has moved through a series of intermediate positions from its starting location in frame 1 to its final location in frame 2. Our cartoon recognizer could even be supplied with a rule for inferring "accelerated movement" by noticing that the distance an object has traveled between frames 3 and 4 is greater than the distance traveled between frames 1 and 2 -- assuming that each frame represents an equal temporal interval. Usually, however, the cartoonist's graphic device for indicating acceleration in speed of an object within a single frame is to draw a few sweeping lines emmanating from the object.

To summarize: we can characterize the early stages of processing a cartoon-strip as an attempt to account for the visual elements and changes in the drawing in terms of a few basic ideals: object, space, time and movement. These ideals, of course, are the same fundamental concepts which underlie our perceptual descriptions of the everyday world.

Beyond accounting for changes happening to a single object, we can identify a next stage whereby the changes in one object can be attributed to changes in another object. In other words, the cartoon comprehender will have to know how to recognize causal relations.

Although he worked with moving geometric shapes, not cartoon-strips, the Belgian psychologist Michotte (1963) has studied the perceptual conditions giving rise to the impression of physical causality. As a Gestalt psychologist Michotte was interested in showing that perceived causality was not based on learned associations between events but, as with the perception of stroboscopic motion or the constancies of depth, size and shape, could be attributed to innate organizational processes in the brain. To support his claim he had people describe their impressions of the motions of "meaningless" geometric forms whose spatial and temporal parameters he could precisely control. He found that even slight changes in the spatial or temporal relations between the two movements could significantly change the causal impression. Thus a brief pause inserted between the time an object A is seen touching object B, and B's subsequent departure, can impair the impression of a causal "impact."

For Michotte, the fact that only precise spatio-temporal conditions gave rise to the causal effect was evidence that innate mechanisms were governing the perceptual experience. We can use Michotte's findings, not to support the hypothesis of innate field forces in the brain, but in terms of description-formation. Just as visual programs use such cues as intersections in order to generate descriptions of 2-D drawings in terms of 3-D objects, so must the perceptual system be able to use spatial and temporal discontinuities to activate "theories" about the causes of movement.[1] Moreover, despite the loss of "continuous" movement, subjects' interpretations of cartoon-strip analogies indicate that they too use features of the visual display to activate theories about causal relations.

Kinetic movement, where the motion of object B can be attributed to the transmission of physical energy from object A, is only one kind of change concept found in subjects' interpretations of the drawings. There are many others. For example, a different kind of change concept is needed to make sense out of Figure 1.6. Here one needs to make sense of two simultaneously occurring changes: (1) the displacement of the vertical bar and (2) the successive deletion of part of the array of dots. Instead of seeing these changes as two independent events, a single causal paradigm can be invoked which describes the changes as related. That is, one sees the disapperance of the dots as <u>caused by</u> the displacement of the bar. Most of the interpretations in Appendix B for this cartoon embody such a causal paradigm. Three examples are: "mowing a lawn," "a bulldozer plowing snow," and "erasing a dirty blackboard." While these interpretations use different kinds of objects as their referents, all three seem to be derived from the same "deep" causal schema -- the deletion of part of an array of elements is attributed to the displacement of the rectangle. An underlying rule which the conceptual system seems to be

---

[1]Some of these implicit theories of causal movements have been made explicit recently by Sylvia Weir (1978) who has written a computer simulation of the kinds of causal percepts studies by Michotte. Her work will be discussed in more detail in section J.

using in interpreting this cartoon and others is that correlated changes are causally related in some way.

Michotte's experiments also reveal that the perception of physical causality is often closely intertwined with knowledge of social and psychological causality. An interesting example is the way some of his subjects interpreted the "triggering" sequence in which B's speed after impact is twice A's speed prior to impact. In accounting for the increase in velocity, some people introduced an explanatory concept into their descriptions, not from the domain of physics, but social relations -- e.g., "A's approach frightened B, and B ran away." Thus, a description in terms of physical causality can be assimilated by a higher level of description of the event in terms of psychological causes and effects.

An early experiment which probed the perceptual bases of social causality was that of Heider and Simmel (1944). In their study, subjects watched an animated movie whose characters were a large triangle, a small triangle and a circle moving in-and-out of a "door" of a large rectangle. Just as Michotte's subjects were asked to make sense out of the motions of geometric shapes in terms of the laws of physics, subjects in Heider and Simmel's experiment were told to understand the movement patterns in terms of meaningful social events. The authors were interested in the rules by which subjects mapped motion patterns onto social and psychological schemata. For example, one of the questions they asked concerned the rules for determining which geometric shapes are "agents" and which belong to the class of "inanimate object." One rule which they used can be paraphrased as follows: "Shapes which move under their own power are 'agents' while shapes whose movements are always under the control of another shape are 'objects.'" Thus, subjects in the experiment described the moving segment of the rectangle as a "door" because it moved only when in contact with one of the other shapes, say the small triangle. Because the triangle had been perceived throughout the film as capable of self-initiated movement, it was assigned the role of an animate agent while the rectangle's segment was described as a "door," an inanimate object.

Now the problem of identifying the agent in an event is not restricted to the perception of social events; "agent" is an abstract concept which underlies our knowledge of events in general. We can therefore interpret Heider and Simmel's study as an early attempt to define a heuristic rule, based on motion cues, for "parsing" perceptual events into their underlying concepts.

It is interesting to speculate that the rules for mapping movement patterns onto underlying event structures are acquired in developmental stages. There is some evidence for this from an early study by Piaget (1929) on the changing criteria which children apply to their judgements of whether or not objects are alive and endowed with consciousness. Piaget found that for the youngest children, any object involved in some action is considered alive, whether or not it moves. At a second stage, only objects which move are alive. Thus a bicycle at this stage is alive; a sofa is not. At a third stage, the criterion is whether or not an object is capable of autonomous movement as opposed to being moved by an outside force; a cloud is alive, but not a bicycle. It is only toward the late stages of cognitive development that children use "formal" criteria for restricting their attributions of life to animals and plants.

For Piaget, young children's willingness to classify inanimate objects as alive was symptomatic of immature reasoning about the real causes of movement as well as indicating a lack of formal criteria for defining living systems. However, there is another way to interpret these results more in the spirit of the present thesis. Is it possible that children's "mistakes" in judging animacy derive, not so much from a confusion about what is or is not alive, but rather from their misinterpreting the true goals of the experiment? Suppose, for example, that part of what children learn from their perceptual encounters with the world is a set of abstract descriptions of typical movement properties of different classes of object. One such description might be that "living things generate their own movements." Other descriptions might be more complex dealing with causal relations between movements; e.g., "hitting." If the child possesses such movement descriptions, and they are abstract, then they could be used imaginatively: for example, during play, as instructions for how to simulate a "fight" or a "chase" using any pair of inanimate objects as "actors" -- even a couple of sticks. Moreover, this same abstract knowledge about movement could be used analogically to answer questions in an experiment. If, instead of "Is a cloud alive?" a child interprets an experimenter's question as "How could a cloud be like a living thing?" then he could use his powers of analogical reasoning to find a partial fit between the movements of clouds and the movements of people. In terms of abstract dynamic relations, a man strolling through a meadow does share some features with a cloud floating about in the sky.

Note that even adults can be induced to attribute animacy to non-living things under certain conditions. A good example is Norman McLaren's film "A Chairy Tale." In the movie a chair is seen as dodging and darting about in order to avoid being sat on by a person. By combining his knowledge of the rules of animate movement with his knowledge of the techniques of film animation, McLaren convinces us, in the context of a movie, that a chair can have goals, be stubborn and vindictive in its dealings with people. Arnheim (1964) has pointed out that in purely visual terms there may be no hard-and-fast cut off point in making judgements about animacy. Rather there is a "scale of complexity" -- a continuum whereby complex movement patterns may be judged more life-like than simple movements, but the boundaries between "living" and "non-living" motion are fuzzy:

> ...first the difference between what moves and what does not move. Second, flexible movement, which involves internal change, is at a higher level of complexity than the mere displacement of rigid objects or parts of objects. Third, an object that mobilizes its own power and determines its own course is higher than one that is moved and steered -- that is passively submits to being pushed, pulled, repelled, attracted by an external agent. Fourth, among the 'active' objects there is a distinction between those that move merely on internal impulse and others whose behavior is

> influenced by external centers of reference. Within this later group there is lower-level behavior, which requires direct contact by the outer agent (e.g., object B's 'taking off' when touched by A), and higher level behavior, which involves response to the object of reference across some distance in space (e.g., A seen as moving 'toward' B, or B escapes while A is approaching) (p. 401).

Taken together, the experiments of Michotte, Piaget, Heider and Simmel as well as our ability to delight in a Norman McLaren movie, suggest that part of what learning contributes to perception is a set of "parsing" rules based on various kinds of perceived change for assigning objects to a number of abstract concepts which underlie events. A number of researchers in artificial intelligence and cognitive science, interested in the problem of the representation of meaning, have suggested what kinds of concepts might constitute an abstract event schema. The parts of an event have been referred to by various names: "cases" (Schank, 1973; Fillmore, 1971; Norman and Rumelhart, 1975), "frame-slots" (Minsky, 1975), "IMPS" (important parts) (Winograd, 1975). Minsky has suggested that a good way to think of the parts of an event schema is as a series of questions a comprehender tries to answer about an event in order to understand it:

> "What caused it? (agent)"
>
> "What was the purpose? (intention)"
>
> "What are the consequences? (side-effects)"
>
> "Who does it affect? (recipient)"
>
> "How is it done? (instrument)" (p. 246).

These concepts do not exhaust all the possible components of an event structure, but they do suggest some of the concepts that are necessary if an event is to perceived as "well-formed" or meaningful. One way then to characterize the task facing someone in the present experiment is as the attempt to find a meaning for a drawing which satisfies a minimum set of underlying conceptual relations making-up an ideal or prototypical event structure. But because the cartoon-strips are often designed to be ambiguous or incomplete, fitting a cartoon to a well-formed event concept may not always be straightforward; it may require a kind of "problem-solving." The next section takes a closer look at the process of fitting a cartoon to an event schema and gives some examples of the kinds of strategies people use in generating alternative interpretations for the same cartoon.

F. Fitting a Cartoon to an Event Schema

Thus far we have portrayed the attempt to comprehend a cartoon-strip analogy as basically a hierarchical or "bottom-up" process. That is, we have imagined that it occurs in a sequence of stages, each stage analyzing the input for certain kinds of information and then passing this information along to the next stage of analysis. Figure 1.7 summarizes these stages diagrammatically.

While a hierarchical model of the description process helps us to see some of the different kinds of knowledge required to understand cartoon-strips, it is not yet adequate. This is because there are really two complementary demands being made on our cartoon comprehender. On the one hand, the visual features and elements in the

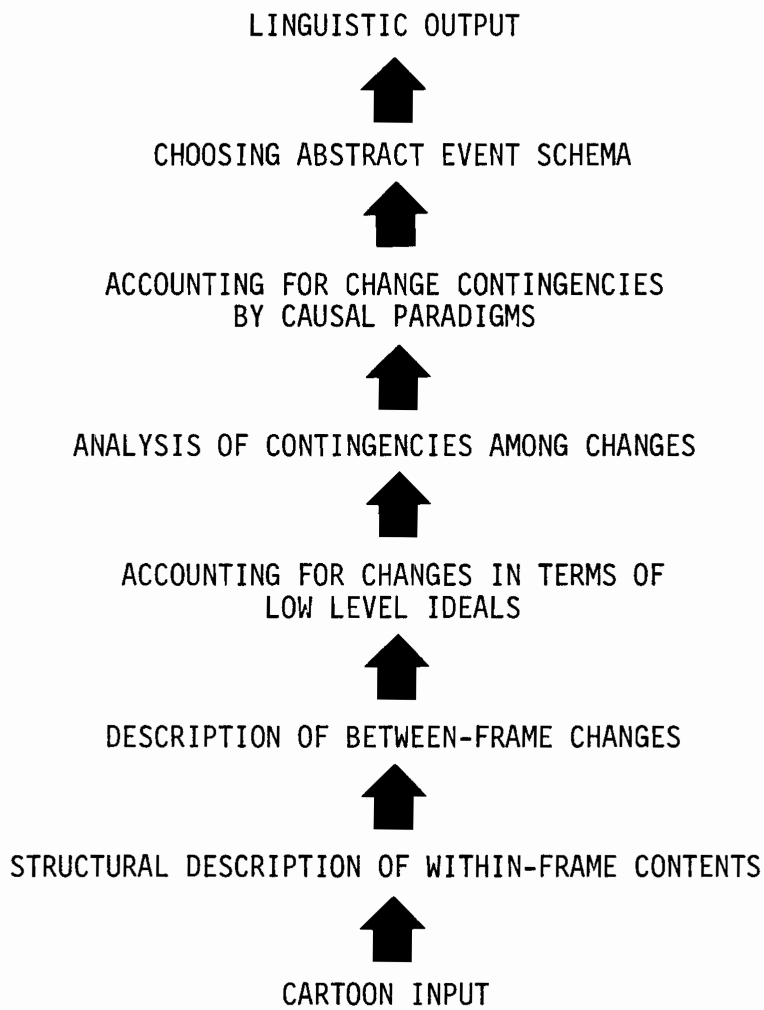


Figure 1.7. - Sequence of stagesin cartoon description.

drawing are demanding to be accounted for; i.e., to be assigned some role in an event schema. On the other hand, the parts or slots of the event schema themselves are demanding to be filled. Hence, the process of understanding a cartoon analogy is not strictly hierarchical; rather it is driven both from the "bottom" (by the visual features in the drawing) and from the "top" (by conceptual knowledge of events).[1]

Note, however, that this "bottom-up/top-down" process does not always work smoothly. Sometimes, the comprehension process may have difficulty assigning a visual element to a role in an event schema or, correlatively, it may not be able to fill an empty slot in the top-level event structure right away. When this happens, the description process runs into trouble, but also gets more interesting because it forces the comprehender to deal with various kinds of problems. Once it is realized that understanding a cartoon can run into snags, it is easier to see how cartoon comprehension qualifies as a kind of problem-solving or theory proving. For example, a particular shape cue in a drawing may activate a hypothesis as to the kind of event being represented. But in testing this hypothesis against the "data" (the visual features of the cartoon), other features may be

---

[1]The terms "bottom-up" and "top-down" come from computer science and signify the way an underlying representation of an input is derived. Usually, the input is a phrase or sentence. A bottom-up analysis starts by examining the input, trying to categorize its various constituents; for example, asking what part of speech a given word belongs to. A "top-down" analysis, on the other hand, begins with a hypothesis about what it will find in the stimulus; that is, it might ask if there is a verb in the sentence. Roughly, bottom-up analyses are "data-driven" while top-down analyses are said to be "concept-driven."

seen as inconsistent with the initial hypothesis. At this point, the comprehender is faced with a decision. He can abandon his initial hypothesis and look for a new one; or he can seek an explanation which would account for the inconsistency. Seeking explanations may involve the use of transformations: mental operations which modify the description in various ways; e.g., imagining the event from a different viewpoint, adding "unseen" elements to the description, seeing the cartoon as the "result" of an earlier event, and so on.

Once we let the description process work top-down (guided by hypotheses) as well as bottom-up (guided by features) and admit the role of transformational strategies, we can see how there may be many ways to account for the visual elements in a given drawing. Any number of acceptable event descriptions may be generated depending on the goals and strategies of the problem-solver. This view of description as problem-solving can help us better understand two kinds of issues raised by subjects' interpretations: (1) how the same cartoon can be described as fitting different kinds of events; and (2) the problem of metaphorical interpretations; e.g., seeing a cartoon as an instance of a so-called "abstract" concept. In the following sections we give some specific examples of how the description process deals with these problems.

G. The Problem of Multiple Interpretations

Simple visual structures require complex solutions. Of all the cartoons, the one in Figure 1.8 probably has the simplest visual structure. But as we shall see, simple visual structures do not always mean simple solutions. In fact, the opposite may be true. Complex visual patterns which show several interacting elements provide more constraints on which roles they should be assigned to in the overall event structure. Simple patterns, on the other hand, with a single object undergoing a single transformation, may require more cognitive effort to satisfy the minimum conditions of the abstract event schema.

Figure 1.8 shows three black dots of increasing size followed by a completely blacked-out last frame. This is the input to the description process: the raw data to be explained. The first problem posed by the cartoon is how to account for the changing size of the dots. Several solutions are possible. A popular solution used by subjects in the experiment is illustrated by the interpretation "getting crushed by an oncoming car." Here, the increasing size of the dots is seen, not as an actual change in physical size, but as attributable to the movement of a single dot toward an observer; i.e., the dot paints an increasingly bigger image on the retinal surface of the observer or on the cartoon-frame "window", not because it is getting larger, but because it is getting closer.

Describing the dot as moving toward an observer solves one problem but raises another. What is causing the movement? Dots cannot usually move by themselves. By transforming the dot to a "car", the subject provides a way, not only of explaining the dot's movement, but also the final blacked-out frame. First, note that the "dot-as-

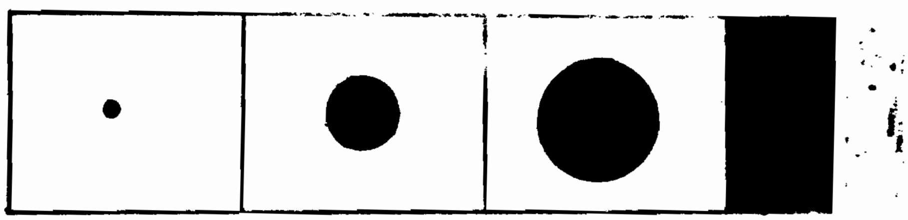


Figure 1.8. - Simple cartoons can require complex solutions

car" is assigned to the instrument slot in the event structure. Instruments are controlled by agents; thus, the subject has introduced an agent into the description (cars cannot usually drive themselves) although there is no agent physically represented in the drawing. Similarly, the recipient of the event, the observer, is unseen in the drawing but plays an important role in the description. Finally, the blacked-out last frame can be explained by the approaching-car schema -- it represents the moment just before impact when the car blocks the entire field of vision.

Seeing the dot as a "looming" object is not the only way to account for the size transformation shown in the drawing. Another possibility is to imagine the dot as a fixed point in space and that the observer is moving towards it. A number of interpretations chose this solution; e.g., "watching a planet get bigger as I descend in a spaceship." Here the dot (planet) has been assigned the object or destination role in the event schema while the unseen observer is the mobile agent. Note also that once the dot is transformed into a "planet," it activates a space-travel schema or script which imposes its own constraints; for example, it specifies that a special vehicle, a spaceship, be inserted into the description to explain how travel through space is possible. Thus, a more precise characterization of the underlying conceptualization of this cartoon would have to include the idea that it is really the spaceship doing the moving, under the control of an unseen agent.

While in the first two solutions, the dot is perceived as maintaining a constant size, although changing its distance relative to an observer, a third solution is to describe the dot as actually undergoing a size transformation. A number of solutions achieve this. Consider the example "an expanding circle caused by a pebble dropped in water." This interpretation is a good illustration of a particular transformational strategy -- imagining the visual input as part of a larger temporal context -- which remains unseen. In this interpretation the entire cartoon depicts an event which itself is a consequence of an earlier event. Specifically, the "expansion of the circle," the seen event, is caused by the prior dropping of the pebble, an unseen event. Again, an essential component of the description is an agent -- an instigating force -- which dropped the pebble causing the ripples which are described as forming an expanding circle. The agent, although critical to the meaning, has no explicit surface representation in either the verbal description or the pictorial sequence.

This interpretation is a good illustration that while invariant features of the drawing may be simple -- i.e., an expanding circle -- a good deal of knowledge of the world is often required to see it as an analogy for a meaningful event.

Changing the rules to fit the theory. In an earlier section we mentioned that knowledge of the "rules" or "syntax" of cartoon-strips is necessary to make sense of these drawings. While knowledge of these rules may be necessary for comprehension, some of them can apparently be changed if it suits the needs of the overall problem-solving strategy. An example of changing the rules can be seen in some of the interpretations for the cartoon in Figure 1.9. Most subjects described the half-circle in the drawing as a type of container into which the blob-shaped object is gradually "absorbed" or "sucked." Seeing the cartoon as a kind of "absorbing" respects the conventional left-to-right scanning of the adjacent frames. This was not always the case, however. Several interpretations describe the event as a kind of "spilling" as in "spilling ink from an inkwell." Now seeing the event as a kind of spilling requires changing some of the usual assumptions about how cartoon-strips represent time. This might have occurred as follows.

In trying to make sense of the drawing a subject begins by analyzing the visual features of the left-most frame. Unlike many of the cartoons where the contents of a single frame alone are meaningless (e.g., the need for dynamic relations), this first frame contains a blob-shaped figure which is highly representational -- that is, it looks like what a liquid might if spilled. It is highly probable then that the blob acts as a strong visual cue activating a "spilling" schema as the best way to account for the pattern.

The subject now has a theory but he also has a problem. If he respects the conventional left-to-right scanning direction, he finds the blob gradually disappearing into the container. This action is inconsistent with the idea of spilling. Most subjects get around this by simply imagining the first frame as the consequence of an unseen event which has caused the substance to be spilled. The semi-

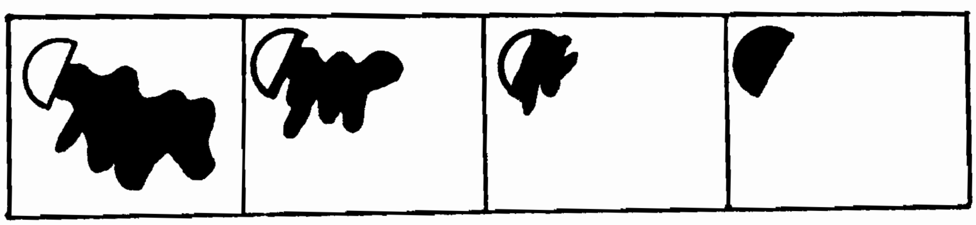


Figure 1.9. - Changing scanning direction changes the meaning.

circle then becomes an instrument (vacuum, cloth) for cleaning up the spill. The inconsistency is resolved. For other subjects however, the spilling schema is such a good fit to the first frame that they are reluctant to abandon it; instead, they seek a means of eliminating the inconsistent last three frames which will permit them to preserve the idea of spilling. The solution is simply to reverse the normal scanning order so that the beginning of the event is the right-most frame. Then describing the cartoon as liquid spilling from a container is quite compatible with the features of the drawing. Moreover, this solution is, in a way, more parsimonius, because it does not require the insertion of unseen agents or instruments into the description as do the "absorbing" or "sponging" interpretations.

"Timing" as an abstract property of events. "Timing" is an important part of our knowledge of invariant properties of certain event classes. Timing refers to our estimate of the typical duration of events. That knowledge of timing plays a role in the comprehension of the cartoons can be seen by comparing the interpretations for the drawing in Figure 1.10; (a) "a crowd leaving a stadium," and (b) "opening a shower tap."

These two interpretations account for the changes in the drawing in very different ways. In (a) the conceptual structure of the event describes the dots (the crowd) as animate agents moving away from an inanimate object, the stadium. In (b) the dots are seen as inanimate particles of water whose movement is seen as caused by the action of an unseen agent. Although not explicit in the surface interpretations, each implies very different time values for the event.

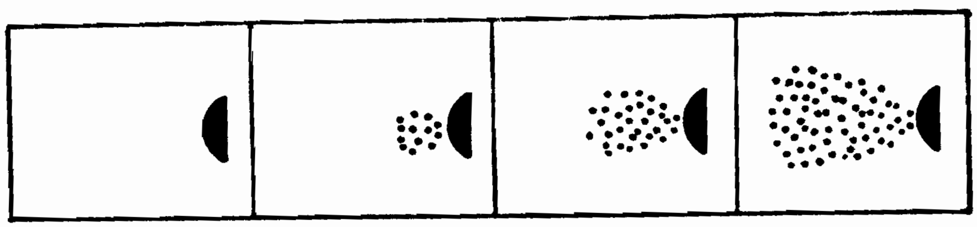


Figure 1.10. - The scale of "timing" changes depending on whether the cartoon is seen as: (a) "opening a shower tap" (seconds); or (b) "a crowd leaving a stadium" (minutes).

For example, our prototypical knowledge of the time it takes for crowds to leave stadiums is on the order of 10 minutes to an hour. Opening a shower tap, on the other hand, occurs in a matter of seconds. The timing value ascribed to each event then is partly constrained by whether or not the dots are seen as human agents or inanimate particles.

Note that what is important in this example is the existence of a constraint between "timing" and "agency." The order in which they can be applied may vary, but the constraint is constant. For example, one could lessen the probability that the dots would be described as animate agents by suggesting to a subject a timing value for the cartoon; i.e., by telling him "this event takes place in two seconds." Similarly, suggesting that the dots represent "people" could bias the inferred duration of the total event. Again, a conceptual system which can apply the constraints in any order must be very flexible.

H. The Problem of Metaphorical Interpretations

So far the examples have emphasized how multiple interpretations of the same cartoon could be explained by acknowledging the role of transformations in the description process. Different transformations can account for the same "data" in different ways. Some of these transformations involve assumptions about vantage point, movement, unseen causal agents, timing, and even assumptions about the conventions of scanning cartoon-strips.

Earlier we pointed out that some of the interpretations go beyond a description of the drawings as familiar visual events -- something one might actually observe -- to describing the cartoons figuratively or metaphorically. Examples of metaphorical interpretations found in Appendix B are as follows: cartoon 3 ("a symbol of evolving domination"); cartoon 4 ("a non-resolvable problem you have to live with"); cartoon 5 ("assimilation of the minority by the masses"); cartoon 6 ("clarification of fuzzy ideas").

A first reaction to these interpretations is that they are qualitatively different than the others. They seem more creative; the correspondences strike us as deeper or more abstract than those which merely describe the cartoons as analogies for concrete, observable events. For example, finding a fit between cartoon 3 and the concept of "evolving domination" seems less direct, more unexpected and therefore more imaginative than describing the pattern as "crushing a mushroom with a rock." But are there really important qualitative differences between the so-called "metaphorical" and "analogical" kinds of interpretation? Do the metaphorical interpretations call upon processes and knowledge which differ fundamentally from those we have been discussing all along? Or, are they consistent with the view of descriptions we have been developing? I believe that the difference between the metaphorical and concrete interpretations is more apparent than real; more a difference of degree than of kind. But in order to see how this is so, we must expand our notion of the description process to consider its role in human memory.

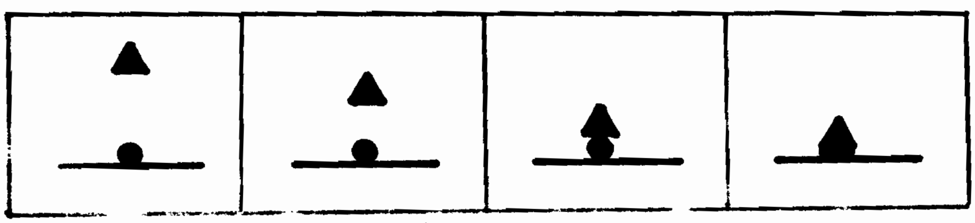


Figure 1.11. - A cartoon analogy from Experiment 1 interpreted as (a) "Crushing a mushroom with a rock"; and (b) "A symbol of evolving domination."

The role of descriptions as memory units has been the focus of several recent articles by Norman & Bobrow (1979, 1976) and Bobrow & Norman (1975). The way these authors view the role of descriptions in memory meshes nicely with our earlier discussion of perception as the building of descriptions. At that time, we argued that an intelligent perceptual system should be able to generate many different descriptions from a single underlying event or situation; different descriptions would arise from changes in the activity of the observer or in from changes in the environment. Norman and Bobrow take this idea one step further. They reason that if perception can generate multiple descriptions of an event, then a collection of such descriptions could be organized into a memory record associated with the event. Imagine, for example, that one witnesses a policeman chasing a robber. One might generate two descriptions for this event. When considered from the point of view of the policeman, the event is described as one of "chasing." When attention switches to the robber, the event is described as "fleeing." The later retrieval of the original underlying event could be prompted by encountering a similar event described as either "fleeing" or "chasing."

In addition, these authors point out that not only can descriptions of an underlying event represent differences in perspective or viewpoint, but descriptions may also vary along a continuum from specific to general. Consider, for example, cartoon 3 shown in Figure 1.11. We can construct a number of descriptions for this cartoon and order them along a scale from specific to general as follows:

1. "A black disc against a white background is resting on a surface represented by a horizontal black line. A black triangle, about the same size as the disc, is posed slightly above the disc. The black triangle moves downward making contact with the top of the disc. The disc's shape buckles under the weight of the triangle. The shape has changed from round to eliptical.

2. A mass causes a change in shape of a sphere by application of physical pressure.

3. Object X causes the destruction of Object Y by crushing it with a superior weight.

4. X controls Y by application of force."

Clearly, we have not exhausted the different levels of description possible for this particular cartoon. The example does illustrate, however, that it is possible to successively transform an initial description of a cartoon to abstract levels. We believe that the ability to transform an initially specific description to a desired level of generality is the key step in the derivation of metaphorical interpretations. It assumes, in agreement with Bobrow and Norman, that finding a resemblance between an event and an item from memory, is often a cyclic, recursive process where descriptions can be modified until they meet certain criteria; one of these criteria might be that an initially specific description be generalized before it can be used to guide a memory search for a given concept.

Let us illustrate how a generalized description might have been used in creating a metaphorical interpretation by examining a particular example -- that of the interpretation "a symbol of evolving domination" for the cartoon in Figure 1.11. The problem is to explain how a subject could find a path between a description of the cartoon to the abstract idea "evolving domination." The path is certainly not a direct associative link between the cartoon based on common physical features; the meaning of "domination" has nothing intrinsically to do with triangles or dots. The link must be established at a more abstract level.

Let us begin our search for a common path by looking up the definition of "domination" in the dictionary. Does the dictionary provide any clues about the possible basis of matching? Webster's New International contains the following entry for "dominant:"

> Syn. preponderant...that which outweighs every other thing of its kind in power, influence or force.

Notice the term "outweighs" in the definition. It reveals that the meaning of dominant shares a common feature with an event one might actually observe or experience , a heavy thing exerting superior force over a lighter thing.

The dictionary definition provides a strong clue as to how the match to the cartoon was obtained. Returning to our list of descriptions for cartoon 3, we find at level 4 the description "control by the application of weight or force." This is a close fit to the dictionary definition. If we can assume that our mental concept of "domination" also contains multiple levels of description, one of which being the general component "control by weight or force," then we see how the path between the concept "domination" and the cartoon could have been established on the basis of a match a deeper, more general level of description.

This example suggests that there is no fundamental, qualitative difference between metaphorical and concrete interpretations of the cartoons. Rather, the two types of interpretation reflect the degree of generality to which someone has processed an initial description of a cartoon in accordance with how they interpreted the experimental instructions. If one interprets the task as one of finding an analogy for a cartoon which is a concrete, physical event -- something one might observe -- a match will be sought which meets this criterion and which imposes certain specificity constraints on the description. Thus, the "rock crushing a mushroom" interpretation of Figure 1.11 assigns the shapes in the drawing to specific, physical referents. If, on the other hand, a subject sees the task as finding a meaning for the cartoon which is a non-physical event or abstract idea, then the cartoon is processed to more general level. For example, in the "symbol of evolving domination" interpretation, although the triangle and the circle may be assigned the role of dominating force and dominated entity, respectively, the description is more general than the "crushing" interpretation since the shapes are not assigned to specific, physical referents. Rather, in this case, it is the abstract dominating relation alone which is the critical component; the shapes are free to represent many kinds of specific entities: A wife and a husband, powerful and weak countries, and so on.

An interesting example of the ability to use different levels of abstraction in everyday communication has been given by

Arnheim (1969). It involves the use of descriptive gestures during conversation to illustrate various qualities of events -- both physical and non-physical. For example, one can describe a "head-on collision" of cars by gestural movements alone without having to show the cars themselves. The same gestures could be used to illustrate a non-physical event -- e.g. the "collision of opinions" during a debate. Peoples' ability to use gestures in the way -- to apply the same gestures to physical and non-physical referents -- hinges on an awareness of the structural resemblances which abstract ideas and concrete events have in common. And in order to utilize such structural resemblances to produce descriptive gestures or even to create works of art, our cognitive systems must be flexible; they must be able to use concepts to generate descriptions at different levels of abstraction depending on the particular goals of the problem-solving process.

Appendix B shows that the number of interpretations which describe the cartoons as non-physical events are few compared to the number of interpretations of the drawings as physical events. Most likely this is due to the effect of the instructions (See Appendix A) which specifically directed subjects to try and understand how a cartoon could be "an analogy for a familiar event." Also giving subjects a sample interpretation which described a physical event (e.g., an egg falling off a table) probably contributed to the biasing of most subjects toward generating non-metaphorical interpretations.

In the last four sections I have discussed some of the basic concepts people may use in forming descriptions of cartoon-strips. Far from being a straightforward classification process, I have suggested that interpreting a cartoon is a kind of problem-solving or theory-proving which proceeds "bottom-up" as well as "top-down;" that is, I have tried to show how a description of a cartoon in terms of low-level concepts such as "movement" can activate higher level concepts or theories as to the kind of event represented by a cartoon. Moreover, fitting a cartoon to an event concept was shown to involve various kinds of transformations on the initial mental description of a cartoon to meet certain criteria of the problem-solving process.

In the final two sections of this chapter I will consider the problem of how the linguistic and perceptual systems are linked in the interpretation process, asking what insights recent theories of semantic representation can provide into peoples' ability to comprehend visual events and describe their meanings in natural language.

I. Exploring the Interface between Language and Perception

Up to now we have considered the problem of how people understand cartoon-strip analogies as basically a problem of visual pattern-recognition. We have asked what kinds of processes and knowledge a pattern recognizer would have to possess in order to interpret a cartoon as a meaningful event. Our discussion, however, has not included the role of language despite the fact that the final product of the comprehension process is a linguistic string.

A complete account of how people are able to transform a perceptual experience into a linguistic description is not yet within the grasp of current theories of language, meaning and perception (But see Miller and Johnson-Laird, 1976 for a recent examination of the relations between semantics and perception). Despite our generally low level of knowledge concerning the interface between language and perception, it is still useful to ask what the present experiment can tell us about linguistic comprehension.

Mental image mediators and the problem of reference. One popular theoretical construct concerning linguistic comprehension in particular seems hard to reconcile with the ability of subjects to describe the cartoon stimuli as familiar events. That is the notion that mental images act as "mediators" between lexical items and the objects and events to which they refer (Cf. Osgood, Suci and Tannenbaum, 1957; Paivio, 1971).

The hypothesis that imagery can mediate the link between words and external events, and thus plays an important role in language processing and memory, can be traced to the early views of the British empiricists on language. For these philosophers, the main problem to be explained concerning language was that of reference -- how words can name things. This was thought to be a problem because, unlike "signs" which bear an isomorphic resemblance to the things they represent (e.g., a portrait) or natural "signals" used by certain animals (the "releasers" of the stickleback), words were seen as "symbolic" in nature; they do not have to look like the entities they come to signify nor are they innate cues. The configuration of letters a-p-p-l-e in no way looks like the fruit and thus, in principle, could be replaced by any other configuration of letters. Because of their arbitrary nature, words were said to function by convention; that is, because a language community agreed to let a given word stand for a given object. But invoking the notion of "convention" does not constitute an explanation of how words, being arbitrarily defined objects, unlike signs and signals, can nonetheless refer. What psychological mechanisms could perform this feat?

The empiricists' solution to this problem of reference hinged on two particular aspects of their epistemological doctrine. First, was that for the empiricist words did not refer directly to objects. Rather the real referents of words were said to be mental events which they called "ideas" (Locke, in the 1961 edition). Ideas were categorized as _simple_ ideas (e.g., cold, red, soft) which are elementary sensory qualities, _complex_ ideas formed by combining simple ideas (e.g., a particular person or object) and finally _abstract_ ideas (e.g., the notion of "roundness" or the concept of "animal"). Abstract ideas were supposed to be an idea containing the common attributes shared by all members of a class of simple or complex ideas.

Moreover, the empiricists distinguished between ideas and thoughts or memories. A memory or recollection of an object or event was considered to be merely a weaker, faded version of the idea originally derived through sensory experience (Hume, in the 1960 edition).

It is important to note that according to this view, there was no qualitative difference between an idea (the direct perceptual experience of an event) and its memory image. Both had the status of mental images -- the recollection or memory image was simply a weaker, less vivid impression than the original percept.

A second theoretical construct was necessary for the British philosophers to solve the problem of reference. This was the principle of association whereby events which occur together continuously in time and space cause their corresponding mental events to become associated in the mind; once associated, the occurrence of one idea can trigger the appearance of the other. The principle of associationism, coupled with the notion that mental images resemble the ideas of sensory experience, were all the empiricists needed to explain how words refer to both objects and events, yet at the same time hold arbitrary relations to those events. Hearing a word activates in the mind of the perceiver the memory images of those ideas which have become associated, through past experience, with the sound pattern of a given word. Thus anyone who has frequently been exposed to the object, cup, while simultaneously hearing the word "cup" would come to have the mental image of a cup each time the word was spoken. Note that in this view, it is the memory image of a cup which does the referring via direct, non-arbitrary resemblance to the object. The relation between the word "cup" and the memory image is, however, still arbitrary and could presumably be replaced by any other word by the same associative process.

That there were problems in the empiricist solution to the problem of language was recognized rather early, especially concerning abstract ideas. Berkeley, for example, questioned how an idea, which is specific and image-like, could represent an abstract notion like "triangularity" which presumably has to cover different varieties of triangle (equilateral, obtuse, etc.) which have incompatible attributes. But the empiricist solution was ultimately forsaken, not because of logical arguments against it, but because of methodological ones. Behavioristic doctrine decreed that scientific explanation in psychology should not depend upon postulating unobservable, mental entities such as mental images. Thus the problem of language was to be explained by the laws of learning alone as embodied in stimulus-response psychology.

In recent years, however, the mental image has made a spectacular comeback as a theoretical construct in psychology and its role has been thoroughly investigated in a variety of tasks concerning memory and language (Cf. Paivio, 1971). Interestingly, the contemporary approach has preserved the basic empiricist model intact: a memory image can serve as a mediator between a word -- to which it is arbitrarily connected via association -- and a referent to which it bears a non-arbitrary, isomorphic resemblance.

Can the mediationist view be extended to account for the ability of subjects to generate verbal descriptions for cartoon analogies? While mediation theory usually considers only the case of how words can refer to objects via mental images, presumably the theory can be "run backwards"; i.e., applied to the situation in which an object, event or a picture is to be labeled by a word or a description. This would work as follows: on being presented an object, say a cup or a picture of a cup, the sensory qualities of the object would activate a mental image of a cup. Consequently, any words which had been associatively linked to the mental image would also be retrieved; e.g., the word "cup."

What happens when we extend this model to the case of a to-be-labeled cartoon drawing? Remember that the objects which populate these cartoons do not resemble specific objects other than classes of geometric forms. One can see right away the problem this poses for mediation theory. The relation between a mental image and its referent is supposed to be one of resemblance. If this is the case, then the mental images activated by the presence of the geometric forms in a cartoon must be the images of geometric forms. Names which have become associated with the mental images of the various classes of shapes would also be retrieved. Thus, mediation theory has a way of explaining how the shapes in a cartoon could be given their appropriate labels. But can it explain more? Clearly, subjects' descriptions in the present experiment are not just lists of names of geometric objects. For example, the descriptions include prepositions denoting spatial relations: "on", "between", and so on. Of course, one might argue that such spatial relations are like objects in that they are realized as physical patterns and could therefore become associated to prepositional terms, but it is difficult to imagine what the memory image of "between" might be. Such relations seem to be inherently abstract, not reducible to simple physical features.

The problem for mediation theory becomes even more acute when one points out that subjects' descriptions not only label spatial relations, but also contain words which appear to label <u>changing</u> relations among objects: i.e., "wiping," "absorbing," "squeezing" and so forth. If mediation theory is to work in the case of cartoon-strips, it will have to show what a stored image of a transformation such as "wiping" could be like which is not based on a resemblance to the particular class of objects doing the wiping. This is because, as we have seen, the specification of a wiping transformation does not depend on particular objects but on abstract dynamic relations.

If these arguments are beginning to sound familiar, it should not be surprising. The problem with the theory of reference was always thought to be how conventionalized, symbolic entities -- words -- could come to refer in a nonarbitrary way to external events. Postulating the existence of mental images which could do the referring via natural resemblance to the external event was thought to be the way out of the dilemma. The gambit, however, only succeeded in moving the problem onto another, more fundamental level -- that of pattern recognition.

The problem of reference is not so much in explaining how words can become associated to mental concepts; the real problem is deciding what constitutes a resemblance between a concept and an external referent. In other words, the critical question is how concepts refer, not just how words refer. Seen in this way, one realizes that the problem of linguistic meaning is intimately connected to one's theory of pattern recognition; i.e., how the perceptual system is able to recognize resemblances between external events and internal representations. Mediation theory's implicit model of pattern recognition is that internal representations of events and objects are mental images -- faded, picture-like entities which mirror the physical appearance of the external objects to which they refer.

The comprehension of cartoon-strip analogies presents a serious challenge to this "mental picture" model just as it does to "feature" theories of pattern recognition. This is because a key to understanding the cartoons as meaningful events, as we have argued, is the ability of a subject to define a set of relations among the elements in a cartoon which assign them to roles in an event schema. The relations defined by these roles can be best conceived of as abstract concepts (agents causing changes affecting recipients, using instruments, etc..). It is difficult to see how the notion of an unanalyzed, percept-like mental image could embody knowledge of abstract event concepts or the procedures necessary for identifying them for a given cartoon input.

In sum, mediation theory does not get us very far in explaining how subjects can attach linguistic labels to these cartoons.

In fact, mediation theory grinds to a halt after generating the names of the geometric shapes in a given cartoon. What seems to be needed is a model in which the perceptual component is sensitive to the underlying conceptual structure of event classes and to which the linguistic component has access.

Abstract perceptual knowledge and linguistic comprehension. One recent approach to the problem of comprehension which comes closer to meeting the requirements mentioned above is that of Bransford and McCarrell (1974) and Bransford and Johnson (1973). Although their main concern is with linguistic comprehension, the way they conceive of the relation between perceptual and linguistic knowledge can help us clarify how abstract knowledge of events underlies subjects' competence in generating verbal descriptions for cartoon events.

As with mediation theory, Bransford and McCarrell postulate that perceptual knowledge can form the basis for understanding linguistic strings. But unlike the mediationist view that the nature of this knowledge is picture-like mental images, Bransford and McCarrell argue that the knowledge derived from perception is "abstract and relational in nature (and includes constraints on the roles entities may assume, abstract invariants characteristic of events, etc.) (p. 220)."

A key assumption then for Bransford and McCarrell is that an analysis of perceptual knowledge should not be restricted to the properties of isolated objects, but should include knowledge of abstract event classes:

> Consider events we might describe as walking. Such events can vary in many ways, involving different agents, different speeds (within some limited range), different compass directions, different terrains, etc. Despite all these different particulars, the event walking is specified by certain invariant information about an agent who does the walking, some surface on which it walks, about some structural support for the walking movement (i.e. appendages that alternatively move ahead of one another), and about speed (to differentiate it from running). There thus exist abstract invariances characterizing this class of events (p. 197).

Walking is an event requiring only a single, animate agent. Other events may have more complex conceptual structures. For example, events involving "absorbing" require at least two objects with the constraint that one object is an "instigating force" which causes a change in state of the second object. Furthermore, "absorbing" imposes constraints on the kinds of entities that can perform the absorbing (e.g., cloth, sponges, etc.) and what kinds of entities can get absorbed (milk, water, ink, etc.). Also, we do not usually think of "absorbing" as occurring spontaneously; normally some animate agent is involved who controls the instrument of absorption.

Bransford and McCarrell point out that although we continually make use of knowledge of abstract invariants to comprehend, we may not be aware of the underlying structure of events until faced with an odd or anomalous situation.

> Consider events involving the movement of some entity. All movements involve some instigating force responsible for them, and the nature of the object undergoing the movement will affect one's assumption about the instigating force involved. Thus, imagine that one is sitting at home and that there is a piece of paper on the living room table. Suddenly it flutters upwards a bit and falls to the floor. Such an

> event is not necessarily puzzling because one can easily think of possible forces that instigated the paper's movements; for example, a sudden breeze. But if one could not postulate a force that instigated the movement he might be puzzled, especially if the event recurred despite efforts to eliminate the suspected cause. Indeed, such circumstances (generally with cups and plates) often cause people to postulate such forces as poltergeists. The inability to isolate possible instigating forces can cause considerable activity designed to understand what is going on. But note that one's interpretation of an event is partly a function of the entity entering into it. Thus if one's pet bird flew off the table in a similar fashion one would not be likely to postulate forces like breezes or poltergeists, since a bird is understood to be capable of self-initiating acts. (p. 198).

Bransford and McCarrell's examples are further support for our earlier claim that the comprehension of events through perception is a process of fitting them to an appropriate event schema; a cognitive structure whose components include such variables as: "agent," "patient," "instrument," and so on. And, in agreement with our earlier examples of cartoon-strip comprehension as a kind of problem-solving, these authors point out that the comprehension of anomalous or novel events (e.g., a sheet of paper floating up from the floor) often involves a kind of thinking; for example, the ability to insert an "agent" into the description of a scene which accounts for the anomaly (e.g., the wind). But the use of abstract perceptual knowledge is not restricted to the domain of perceptual experience. Bransford and McCarrell argue that our conceptual knowledge of events and entities, alinguistic in origin, plays a critical role in understanding through the medium of language. Let us consider some additional examples from Bransford and McCarrell which illustrate this point.

Consider the sentence: "The man put the plane in the envelope." Most people at first have difficulty understanding this sentence until they realize that the sentence would make sense if the plane referred to was a toy. Understanding this sentence requires changing some assumptions about the nature of the entities participating in the event in order to satisfy the abstract relation: "X is contained in Y." In terms of our earlier discussion, we might say that understanding this sentence requires a transformation of the ideal sizes of two objects -- plane and evelope -- in order for the sentence to be consistent. Introducing another ideal element, the modifier "toy," into the description effects the appropriate size transformations which satisfy the constraints of the containing-relation.

This example shows that our alinguistic knowledge of entities and relations can interact in the comprehension of linguistic inputs. In the above sentence, it is our knowledge of the prototypical sizes of planes and envelopes which forces the additional cognitive processing necessary to satisfy the relation "X is contained in Y." However, such additional cognitive computation would be unecessary if different entities were placed in the same relation; e.g., "The man put the plane in the hangar." Our ideal knowledge of "hangar" already specifies that it is large enough to contain an ideal plane.

A second example from Collins and Quillian (1972), cited by Bransford and McCarrell, demonstrates the role of "agent" or "instigating force" as an abstract category underlying our comprehension of linguistic events involving changes of state. The question they ask is whether or not one must tacitly specify an instigating force to understand a sentence such as: "The policeman held up his hand and the cars stopped (p. 327)." As a test, Collins and Quillian suggest imagining an earthquake has just occurred and two cars which were parked on a hill began to roll down. Now upon hearing the sentence "The policeman held up his hand and the cars stopped" one is forced to search for a possible agent or instigating force which could have accomplished this feat. This is because the implicit instigating force, the drivers, have been eliminated from the sentence by the context "earthquake."

The purpose of these examples is to advance the view that linguistic comprehension involves using words and sentences as cues which guide the comprehender in specifying situations or conditions which permit various kinds of abstract relations to be established among the referents of a sentence. Often this grasping of relations requires "thinking;" i.e., performing transformations which introduce new elements (e.g., instigating forces) or change assumptions about the ideal properties of objects (e.g., size).

In this section I have argued that the link between a cartoon-strip and a verbal description is not a simple one. Attaching verbal labels to these drawings cannot be easily explained by the notion of implicit verbal responses to picture-like mental images. Instead, it was argued that both language and perception seem to be avenues into a complex conceptual structure which encodes peoples' prototypical knowledge of event classes and entities as well as containing procedures for dealing with snags in the comprehension process. What then would such a conceptual system look like? In the next section, I describe some recent efforts in the area of cognitive science to develop explicit, formal models of the human conceptual system based on the idea of semantic networks. Two examples of this approach will be considered in some detail: One concerns the comprehension of language; the other with the comprehension of causal visual events. After this review, we will ask if the interpretation of cartoon-strip analogies reveals any particular flaws in the structural network approach to comprehension which make it less plausible as a model of the interface between language and perception.

J. Structural Networks: The Beginning of an Adequate Inferface?

The idea that a common conceptual framework underlies both perception and language has been advocated recently by a number of psychologists (see especially Pylyshyn, 1973) and workers in artificial intelligence. Their main goal is to develop theoretical models of human symbolic processes in which the mental structures and operations involved in a variety of cognitive tasks can be represented in an explicit manner; that is, in a formal system using a finite number of elements, relations and processes. This strategy has sometimes been called the "computational approach" because its basic premise, borrowed from computer science, is that cognitive and perceptual processes are best characterized as symbol manipulating systems where (1) concepts, events and objects are represented by patterns of symbols; and (2) these symbol patterns can be manipulated in various ways: compared, stored, destroyed, created, etc. In a symbol manipulation system, the meaning (as distinct from the reference) of an object or event (represented by symbol patterns) is defined in terms of its relations to or its effects on other symbol patterns. A critical assumption of this approach is that the relations among symbol patterns can be organized hierarchically in a structural network; that is, the meaning of an object or event can be decomposed into other more basic symbol patterns.

How can a symbol manipulation system be used to represent the meaning of events encoded in both verbal and visual media? What kinds of conceptual relations underlying events can one capture in a structural network? Although the research on the representation of meaning is still in an embryonic state, and no one has actually developed a working model which can generate linguistic descriptions of complex, visual events, it is useful to ask how workers in this area envisage this task. While a detailed discussion of formal models of intelligence and meaning is beyond the scope of this thesis, it is possible to capture the flavor of the computational approach to these problems by giving some examples. We will then ask if the ability to describe and understand abstract cartoon-strips is compatible with the semantic network models of the interface between language and perception. In examining this question I will cite two examples from the literature on representation. The first describes how structural networks represent the meaning of a sentence; the second shows how the same kind of network has been used to simulate peoples' perception of causal visual events.

Let us begin with an example from language. This example has been used by Norman and Rumelhart (1975) to illustrate how the meaning of verbal utterances can be represented in a semantic network -- a directed graph with nodes linked to each other by labeled connections. Consider the following sentence describing a simple action: "Peter put the package on the table."[1] The goal is to be able to represent the meaning of the event described by the words in terms of a finite set of basic concepts underlying our comprehension of events. Moreover, according to Norman and Rumelhart, the conceptual structure resulting from hearing the sentence about "putting the package on the table" should be similar to the structure produced by seeing a visual event in which someone puts a package on a table.

Figure 1.12 shows a diagram which demonstrates how Norman and Rumelhart's structural network would represent the conceptual relations activated when someone hears the sentence "Peter put the package on the table." The top half of the diagram shows that the

---

[1] This illustration is intended only to give the flavor of the kinds of conceptual relationships semantic networks attempt to capture. Not all the concepts of notation are illustrated in the example. For a more detailed discussion of the use of semantic networks in developing models of memory, language and visual perception, see Norman and Rumelhart (1972). Roger Schank's (1975) project of "conceptual dependency" analysis is another attempt to use semantic network representations to capture the underlying meanings of everyday verbal communication.

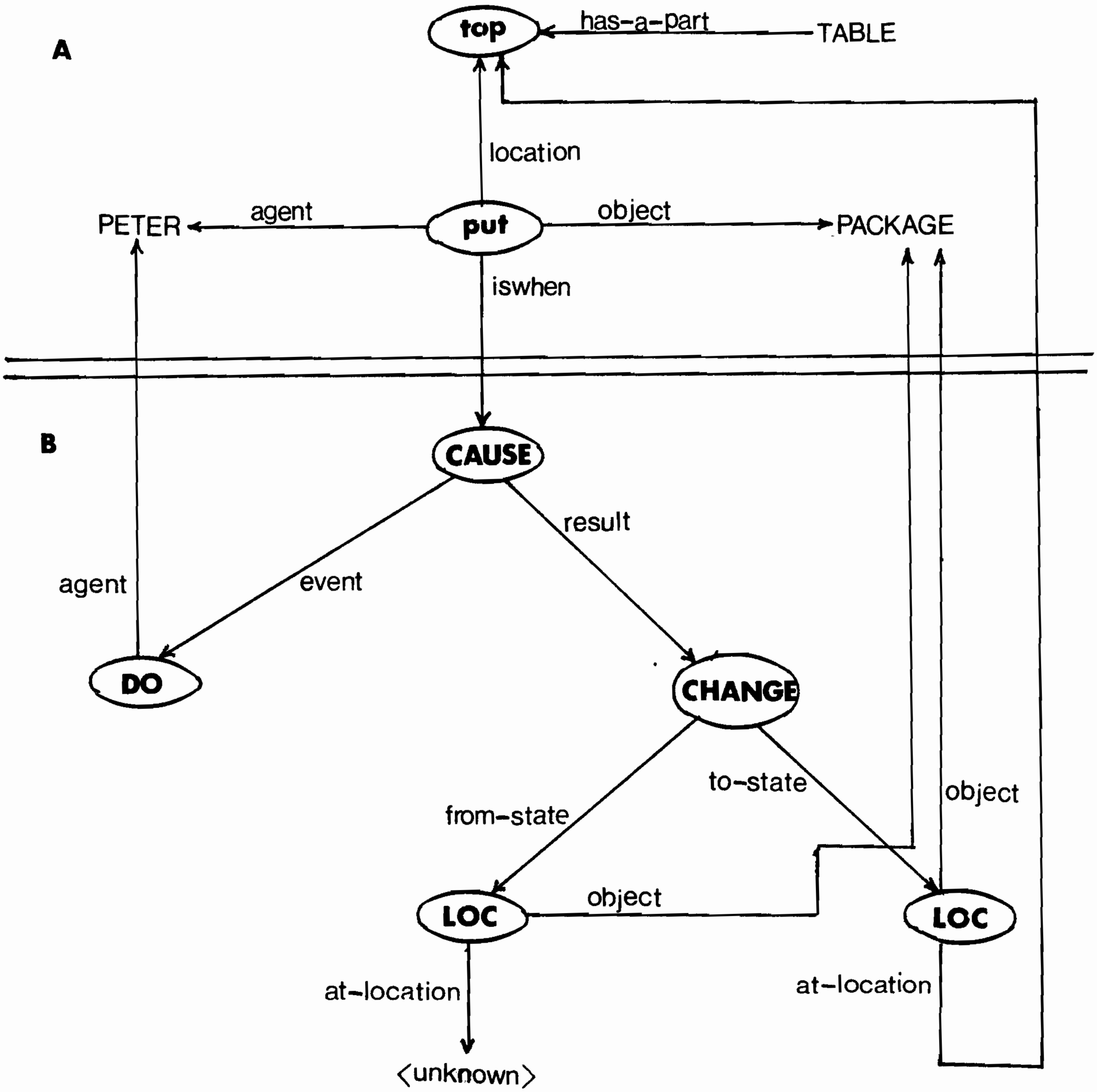


Figure 1.12. - (A) shows the basic representation of the sentence "Peter put the package on the table."
(B) shows the result after the sentence has been expanded into the underlying conceptual structure.

Adapted from: Norman and Rumelhart (1975) verb "put" is the central concept -- the basic action to which the other concepts in the sentence bear a particular relation. Stating this slightly differently, "put" is a schema which is defined by a minimum number of variables or "slots" representing our prototypical knowledge of the act of putting. The arrows leading from the oval with "put" written on it bear the names of these slots or relations. These arrows point to those concepts, which, in this particular sentence, fill these variables or slots.

While part (A) of the diagram shows the basic representation of this particular sentence, the meaning of the verb "put" must itself be defined somewhere in the system. One of the main assumptions of Norman and Rumelhart's representational system is that the meaning of all verbs can be defined by a small number of more primitive concepts which can be linked together in various combinations. Such basic concepts include the notions of state, location, change and causation. The arrow labeled "iswhen" leading from the "put" node connects the verb to its structural definition or schema in part (B) of the diagram. We can paraphrase the structural definition of "put" as follows: "Put" is when an agent performs some action (unspecified) which results in the change of location of some object from one location to another location. By decomposing the meaning of verbs into a hierarchical network of a small number of primitive components, Norman and Rumelhart claim to be able to show in a formal system how the meanings of verbs overlap or differ. For example, by looking at their respective structural definitions, one should be able to see how the verb "put" is related to the verb "throw."

From this example one can get the feel of how Norman and Rumelhart conceive of the human conceptual system: They see it as an "active structural network" which takes verbal inputs and generates "deep" conceptual representations of their meanings which are built out of a finite number of basic concepts. So far, this idea seems quite compatible with the view presented earlier that perception and understanding are processes of description-formation.

Can structural networks be extended to capture the meaning of complex visual events such as those portrayed in abstract cartoon-strips? One recent study in the area of artificial intelligence has begun to explore this problem. Sylvia Weir (1978) has used the computational approach in developing a computer simulation of how people interpret causally related visual events such as those studied by Michotte. Michotte's experiments, it will be remembered, concern peoples' ability to describe the motion patterns of geometric shapes as meaningful causal actions: "pushing," "triggering," "striking" and the like. Figure 1.13 shows six "still snapshots" or frames from a typical Michotte experiment. Human subjects in Michotte's study might have interpreted this pattern as various kinds of actions depending on such factors as the relative speeds of the objects, effects of instructions, etc. Since Weir's program is able to accept dynamic visual events (or descriptions of them) such as Figure 1.13 as input and classify them as belonging to one of several candidate actions for which we have verbal labels (e.g. "striking"), one must

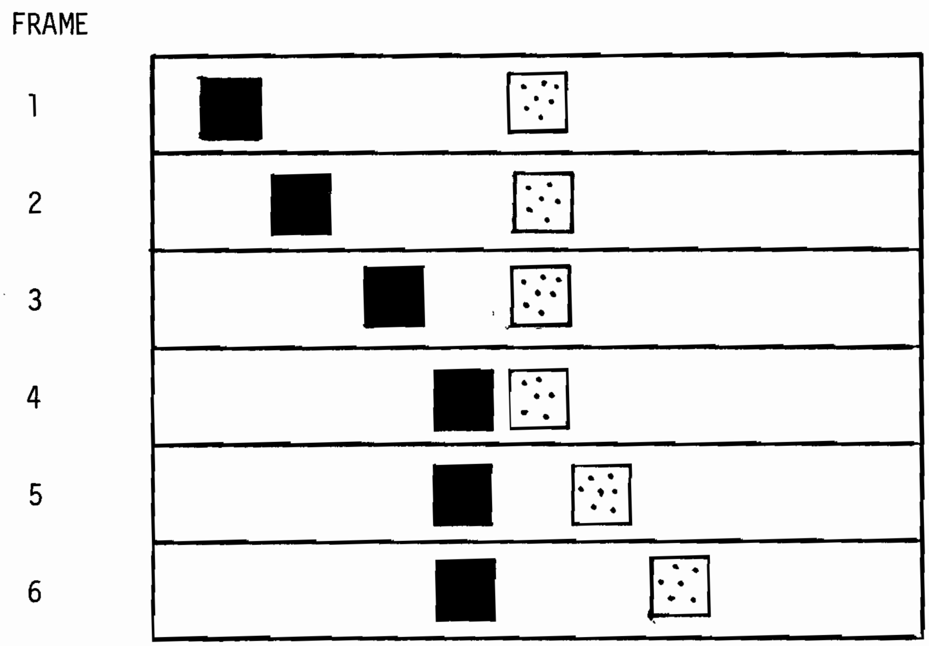


Figure 1.13. - Six frames from Michotte's experiment on the perception of causality

Adapted from Weir (1978)

ask if her system is a precursor of a model able to describe its perceptions of complex events -- even cartoon-strips?

The task facing Weir's program is not unlike that facing Norman and Rumelhart's language understanding system. The goal of both systems is to generate conceptual descriptions of input events and both make use of structural network representations of actions in accomplishing this task. There is a difference in the way these systems work, however. In the case of understanding a sentence, the system finds the name of the action first in the sentence (e.g. "put"), and only then activates the appropriate network of concepts defining its meaning. In Weir's program, on the other hand, the naming of the action is the final stage of the understanding process. The program first begins by analyzing the visual input, describing it in terms of primitive concepts such as "object," "location," "change," "movement," before activating particular candidate schemata as possible fits to the input.

To get a clear idea of how Weir's program uses structural descriptions to interpret causal visual events, let us work through an example with reference to the sequence shown in Figure 1.13. First, how is the program's knowledge of various kinds of action stored? Figure 1.14 shows a structural network representing the concept "push." Weir calls this an "action schema." Its organization is based on the same kind of network structure described earlier for the meaning of the verb "put." That is, the diagram shows a top level node "push" which is defined by a network of more basic components:

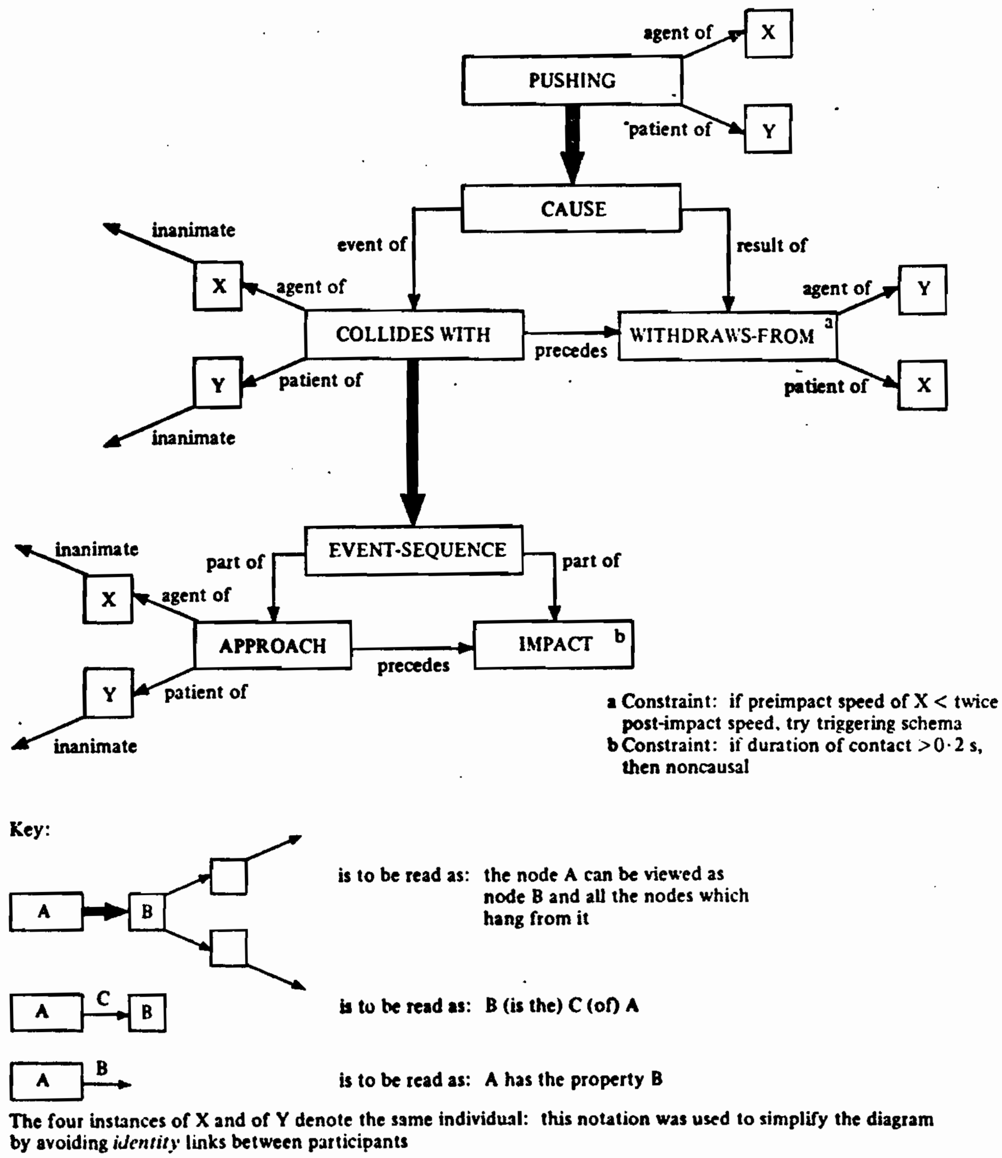


Figure 1.14. - A pushing-schema.

Source: Weir (1978)

"collides with," "approach," "impact," etc. The schema is meant to encode our prototypical knowledge of what "pushing" involves: i.e., "an agent causing the withdrawal of a patient from the agent because of a collision, etc.." The program contains similar structural descriptions for several other actions such as "triggering," and "striking."

An important aspect of Weir's program is the use of what she refers to as "procedures" or "demons." In the structural network shown in Figure 1.14, the lower level nodes labeled "approach" and "impact" are actually low-level descriptions of various kinds of change. For instance, the node "x approach y" is the generalized description for the case when one object is described as moving toward a second immobile object. Now these low level components in the network act as "demons" or "testing procedures." For example, let us assume that the program's analysis of the first three frames of Figure 1.13 produces the description "a black square is moving toward an immobile red square positioned at midscreen.[1] This description of the on-going sequence matches the general conceptual component "x approach y" found in the structural network. Once this match occurs, the "approach" component activates those nodes to which it is directly linked in the network. Thus, the "approach" demon alerts the "impact" demon to be on the lookout for a component in the stimulus description

---

[1] For purposes of brevity I have not described various details of Weir's program such as how it pairs squares in different frames as being the same object, and so on. The interested reader is referred to an excellent review by Boden (1977).

which matches its description. -- i.e., where an object A is "next-to" an object B.

At this point, the program would activate each of its stored action schemata containing "x approach y" as a component; the program, at this stage, expects an impact to occur but does not know what the result will be: Is the black square going to "push" the red square, "strike" it or perhaps "carry-it-along?" Once the expected adjacency occurs in frame 4, the next node up in the network -- "collideswith" -- is activated. This collision demon then modifies the program's expectations of what will occur next. Instead of expecting the black square to continue moving in frame 5, it expects the black square to come to a halt and is not surprised when it finally does so (as it might have been if the black square had stopped for no apparent reason). At the same time, the collision demon activates the "cause" node which in turn generates the expectation that the red square will begin to move away from mid-screen. When, in fact, this happens in frame 5, the program interprets the red square's movement as being caused by a collision whose agent is the black square. Finally, given that the constraints listed in Figure 1.14 have not been violated, the next node up in the network is activated and the sequence is classified as an instance of "pushing." If, on the other hand, a constraint is violated, the program can call upon a different schema to account for the mismatch. For example, if the speed of the red square after impact surpasses that which would be expected by physical causality alone, the extra velocity has to be explained somehow.

The program suggests trying a "triggering" schema. A "triggering" schema could make the inference that the "cause" of the red square's quick departure was a psychological, not a physical event: e.g., the red square was "frightened" by the approach of the black square. This type of explanation in fact was typical of Michotte's subjects in trying to explain the extra speed of the red square.

The purpose of citing the work of Weir and Norman and Rumelhart has been to see how some researchers in cognitive science view the interface between language and perception. These researchers, representatives of the computational approach to the study of cognitive and perceptual processes, view the human conceptual system as a vast structural network in which one's prototypical knowledge of objects, actions and events is represented by "schemata." The formal representation of these schemata consist of hierarchies of propositions which are built out of a finite set of primitive conceptual components. Understanding an input, whether perceptual or linguistic, is characterized by these authors as generating a symbolic representation which captures the underlying conceptual relations contained in the input.

Are structural networks good candidates for a model of the interface between language and perception? Could a system as such as Weir's, based on such networks, potentially be able to simulate the kinds of interpretations of cartoon-strip analogies seen in the current experiment? Clearly, the notion that knowledge is packaged in schemata which capture invariant properties of objects and events is compatible with our earlier discussion of perception and understanding as the accounting for an input by a structure of "ideals." But there was another notion introduced in our earlier discussion; that of transformations. Earlier, it was argued that the interpretation of a visual scene depends not only on knowing about the prototypical features of objects or event classes, but also about transformations; i.e., knowing how the prototype or ideal can undergo various kinds of change in specific contexts. An important source of such changes, it was suggested, is the perceptual activity of an observer as he moves about in the world. Moreover, we tried to show that knowledge of observer-induced transformations seemed to be playing a role in the kinds of assumptions people make in finding meanings for cartoon analogies (e.g., imagining the same cartoon from two points of view).

One might ask, therefore, in evaluating current semantic network models of the human conceptual system, if they embody (or could embody) knowledge of such transformations. One could at argue that Weir's program already uses, to some extent, the idea of "ideals" and "transformations." For example, she mentions that Michotte considered some of his causal sequences as more "typical" than others and that these typical sequences gave rise to particularly strong impressions of physical causality: The "pushing" schema, for example, is such a "typical" sequence. Weir suggests that the other sequences used by Michotte can be considered as deviations from this "ideal." For example, when a sequence differs from the "pushing-physical causality" prototype, as when the red square's speed after impact is twice that of the black square before impact, Weir's program can call upon another ideal -- that of "fright" -- to account for the sudden burst of speed. This changes the description to "triggering." Triggering thus can be thought of as a kind of "transformation" (in a broad sense) of an ideal "pushing" schema.

It is less easy to imagine, however, how semantic network representations of knowledge, such as those we have been discussing could make use of transformations, based on knowledge of the perceptual effects of an organism's own motor activity, in interpreting a linguistic or perceptual event. Yet such knowledge may be critically important. One thing especially clear from the interpretations of the cartoon-strips in the current experiment is that people seem to be drawing on their knowledge of appearances -- how things look under various conditions -- in finding meanings for cartoon events. And because things look differently at different distances, from different vantage points, or depending on whether or it is the observer or the object which is moving, the shapes in a given cartoon can fit many different underlying schemata. Think, once again, of the expanding dot in Figure 1.8 which can be interpreted as either "an object moving toward an immobile observer" or as "a mobile observer moving toward an immobile object." Can a structural network of the type envisaged by Norman and Rumelhart and Weir be equipped with transformational procedures or demons which could interact with other kinds of knowledge in the interpretation of events? This might be a goal for future research. The apparent lack of such perceptual knowledge would seem to be a major weakness of current attempts to use semantic networks to model the human conceptual system. In the next two chapters, experiments will be described which present evidence that the human conceptual system uses knowledge of perceptually-based transformations in fitting visual events to verbal descriptions.

K. Conclusion

This chapter began by asking: What is the process by which people are able to interpret abstract cartoon-strips as analogies for meaningful events? The answer to this question, while by no means complete, has turned out to be anything but simple. In fact, the main contribution of this chapter has been to reveal just how complex a model of how people comprehend events through the medium of abstract cartoon-strips will have to be. Simpler traditional theories which explain the ability to perceive events as similar based on the notion of "equivalence classes" proved to be a poor fit to the process by which subjects in the experiment derived meanings for the cartoons. An alternative view of the comprehension process was advanced whereby the meaning assigned to a cartoon is seen as the end product of a problem-solving process. Such a view seemed necessary to account for the kinds of concepts which appeared to underlie the subjects interpretations. For example, concepts were often found in subjects's interpretations of the cartoons. (i.e., "narrow passage") which had no visible counterpart in the cartoons themselves but served to explain changes undergone by the visible elements (e.g. convergence). Perhaps the most important finding of our analysis of the interpretations is that people use tacit knowledge of their own perceptual activity in fitting a cartoon to an event concept. For example, it was shown how someone's assumptions about vantage point can interact with other concepts in assigning a meaning to a given cartoon. Moreover, the lack of such transformational rules, procedures or strategies, which would allow the conceptual system to make use of knowledge of the effects of its own movements (and their interactions with other concepts), was found to be weakness of current models of the comprehension of linguistic and perceptual events based on semantic networks. Although quite simple, Experiment 1 has helped define at least some criteria for an adequate theory of cartoon-strip comprehension. Let us now turn to a slightly more complicated experiment where, instead of having to supply their own interpretations, people are provided with a set of alternative interpretations and asked to evaluate their relative goodness-of-fit to a cartoon.

CHAPTER II

COMPUTING THE CLOSENESS OF ANALOGY

A. <u>Experiment 2 -- Introduction</u>

In Chapter 1 we explored the mental processes underlying peoples' ability to find a resemblance between an abstract cartoon-strip and a familiar event class. But analogical thinking requires more than the ability to detect a new similarity between two events. Analogical thinking also demands the ability to compute the <u>degree</u> of similarity between a pair of things, events or ideas. In many kinds of problems, for example, several items may match a criterion to varying degrees of acceptability. The ability to compute the goodness-of-fit among a set of items is an essential part of analogical reasoning whether one is taking an IQ test, selecting the best caption for a photograph or choosing among competing scientific theories the one that best fits the data.

In this chapter we examine the problem of computing the degree of analogy in the context of how people discover resemblances between perceptual and linguistic events. The main question is as follows: Given a cartoon-strip analogy which depicts a certain type of event and three phrases describing three different events, what psychological processes underlie the ability to rank order the phrases as best, next-best and worst fit for the cartoon? What makes this problem interesting can best be illustrated by the following example.

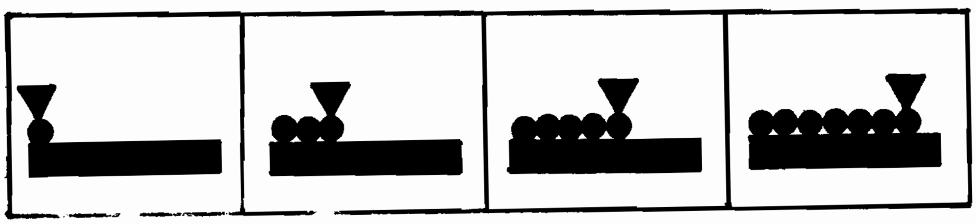


Figure 2.0. - A cartoon-analogy from Experiment 2

Consider the cartoon in Figure 2.0. Which phrase is the best, next-best and worst fit to the kind of event portrayed in the cartoon?

(a) spreading jam on bread

(b) crushing an insect

(c) infecting someone with a disease

Let us assume that most people agree that the preferred order is (a), (c) and (b). (This will be tested in the experiment). How can we explain the ability to rank order the resemblances? The problem is not only to explain the choice of the best-fitting phrase but also to account for our intuition that "infecting," while not as good a fit to the drawing as "spreading jam," is still judged to be a better fit than "sawing wood." Remember that there are no superficial resemblances between the physical features of the objects shown in the drawing and the referents mentioned in the phrases upon which to base our intuitions about the resemblances.

The purpose of this experiment is first of all to demonstrate that people behave non-randomly in making such goodness rankings for a wide range of cartoons and phrases. Such agreement will be used as further evidence against the commonly held view that our knowledge of event classes can be defined by a finite set of features which all the members of the class must possess. The experiment will show that our knowledge of event classes is flexible enough to permit "fuzzy" matches; i.e., although people can identify prototypical instances of specific event classes, they can also make fuzzy matches if given the opportunity to do so.[1] That is, people can perform operations on their mental descriptions of cartoons and phrases in order to find a match at deep levels of abstraction.

Once it is demonstrated that people are able to perform such goodness rankings across symbolic modes, our second objective is to put forth some hypotheses as to how they do it. What kinds of knowledge and processes are involved? In answering this question we will seek help from research in artificial intelligence on another kind of analogical reasoning we know something about -- the solving of geometric analogies. By exploring the points of resemblance between the geometric analogy task and the problem of computing the goodness-of-fit between cartoons and phrases, we hope to illustrate that some general principles of analogical thinking are common to both. Both these tasks involve the intelligent comparison of descriptions and the use of transformational strategies in assessing the degree of fit.

B. Method

Materials. Twenty cartoon-strip analogies of the same type used in Experiment 1 were designed to represent a wide range of visual

---

[1] The position that the underlying criteria of membership in a class are "continuous," not sharply bounded, has been advanced recently by several researchers in linguistics and cognitive psychology (Cf. Lakoff, 1972; Gopnik, 1976; Verbrugge, 1977). A different kind of objection to the traditional notion of logically bounded categories has been raised by Rosch (1975). She argues that many categories, both natural (e.g. colors, lines, shapes) and semantic (birds, furniture) are internally structured such that there are "clear cases" or prototypes in relation to which other members of a category are judged.

events. For each cartoon three phrases were constructed describing ordinary events. The three phrases were intended to represent three levels of "goodness" according to how closely they fit the kind of event depicted by the cartoon: a "best-fitting" phrase whose prototypical meaning closely matches the kind of event shown in the cartoon; a "next-best" phrase which, although differing in its prototypical meaning from the kind of event portrayed by the drawing, could conceivably match the drawing at a deeper level of abstraction; and finally a "worst-fitting" phrase for which a match even at a deeper level of abstraction would be difficult to achieve. Examples of best, next-best and worst-fitting phrases are provided in the introduction to this chapter.

Subjects. The subjects were 23 undergraduate students from the University of Ottawa. They participated in the experiment to satisfy a course requirement.

Procedure. Subjects were provided with a rating sheet containing a list of 20 sets of three phrases each. (Full instructions are given in Appendix A). The order of phrases within a set was mixed to insure that the best, next-best and worst phrases would not always appear in the same positions on the rating sheet. The same mixed order was given to all subjects. Subjects were told that for each group of phrases they would see a cartoon-strip showing geometric forms undergoing various kinds of changes. They were instructed to choose which phrase they considered to be the best, next-best and worst analogy for a given cartoon and to indicate their preferred rankings on the rating sheet. A practice example was provided.

The slides were projected by a Kodak Carousel projector onto a large screen about fifteen feet away. The ambient light in the room was adjusted so that the slides were clearly visible from all parts of the room while at the same time permitting subjects to see their rating sheets. Subjects viewed the slides as a group in a large classroom. Each slide was projected for about three minutes during which time subjects made their judgements.

C. Results

Table 1 shows each of the 20 cartoons and phrase sets judged in the experiment. The three phrases to the right of each cartoon are listed in their predicted order of preference. Next to each set of phrases is a contingency table showing the frequency with which each phrase was ranked best, next-best and worst fit to a given cartoon. Kendell's coefficient of concordance, W, was computed as a descriptive measure of the level of agreement for the 23 judges in ranking the phrases. Significance of W was tested using chi-square for 2 degrees of freedom (Edwards, 1954). Values of W for each of the 20 sets of judgements were found to be significant, $p < .01$. Table 1 shows that in five cases there is apparent lack of agreement as to which of two phrases is the better fit for a cartoon. In three cases (items 7, 13 and 19), the lack of preference is between best and next-best fit; in two cases (items 17 and 20) the lack of clear preference was between next-best and worst fit. Chi-square

TABLE 1

GOODNESS-OF-FIT RANKINGS OF PHRASE-SETS FOR TWENTY CARTOON ANALOGIES

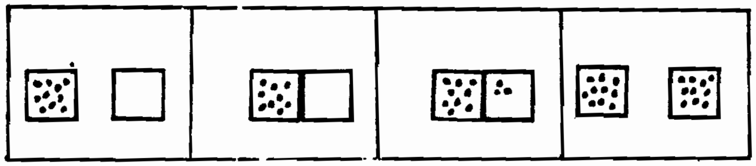
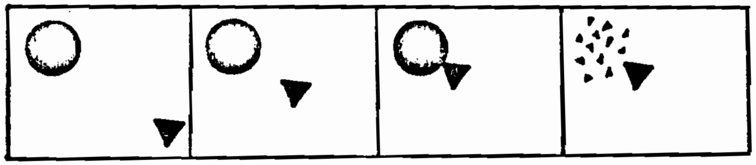
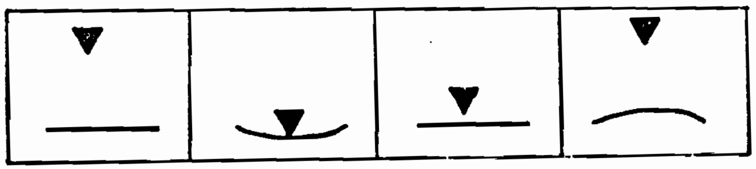
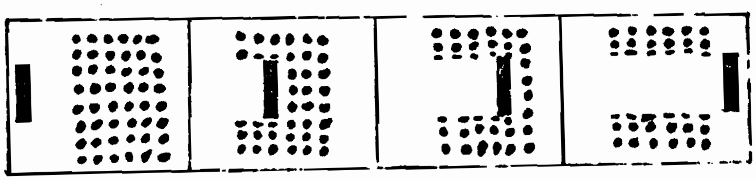
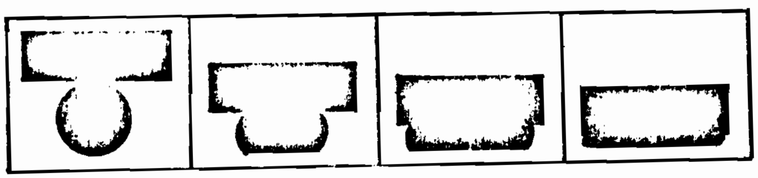
Item	Cartoon	Phrase-Set[a]	No. of Subjects Ranking Phrase 1st, 2nd or 3rd			W
			1st	2nd	3rd	
1.		infecting someone with a disease	23	0	0	1.00
		spreading jam on bread	0	23	0	
		sawing a plank of wood	0	0	23	
2.		puncturing a balloon	19	3	1	.839
		sitting on a hat	1	19	1	
		opening a door	1	2	21	
3.		bouncing on a trampoline	23	0	0	.830
		firing a gun	0	18	5	
		chasing a fox	0	5	18	
4.		wiping a dusty blackboard	23	0	0	.830
		extracting a tooth	0	19	4	
		dialing a phone number	0	4	19	
5.		sitting on a hat	22	1	0	.807
		lancing a boil	1	18	4	
		throwing a stone	0	4	19	

TABLE 1 (continued)

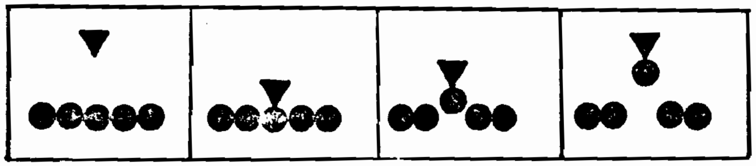
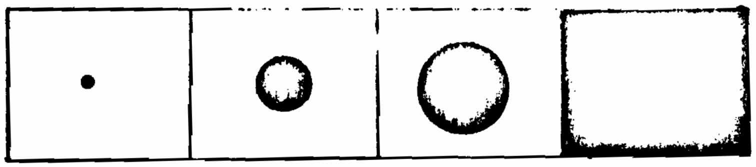
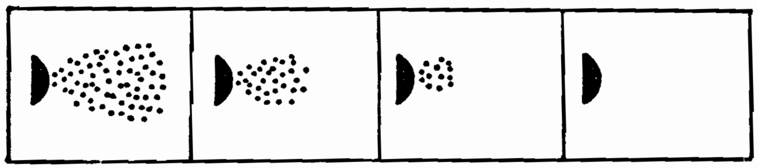
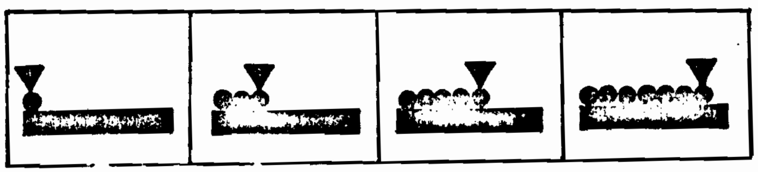
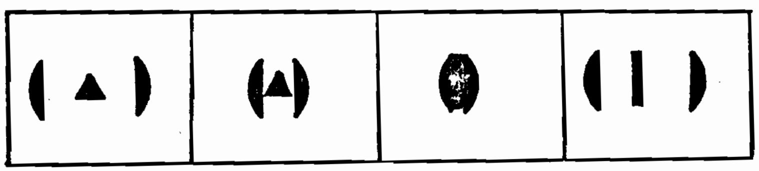
Item	Cartoon	Phrase-Set	No. of Subjects Ranking Phrase 1st, 2nd or 3rd: 1st	2nd	3rd	W
6.		extracting a tooth	23	0	0	.788
		wiping a dusty blackboard	0	16	7	
		twirling a baton	0	7	16	
7.		looking at a charging bull	14	9	0	.762
		falling off a cliff	9	14	0	
		tying one's shoelaces	0	0	23	
8.		vacuuming a dirty rug	22	1	0	.733
		diving into a still pond	1	15	7	
		opening a sealed envelope	0	7	16	
9.		spreading jam on bread	21	2	0	.731
		infecting someone with a disease	1	18	4	
		crushing an insect	1	3	19	
10.		minting a coin	19	3	1	.688
		ironing a shirt	3	19	1	
		slipping on a banana peel	1	1	21	

TABLE 1 (continued)

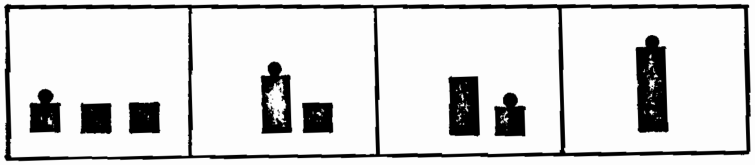
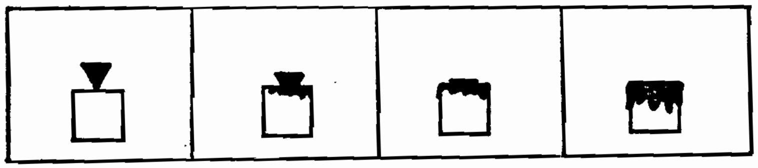
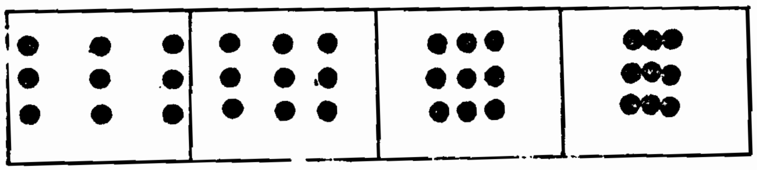
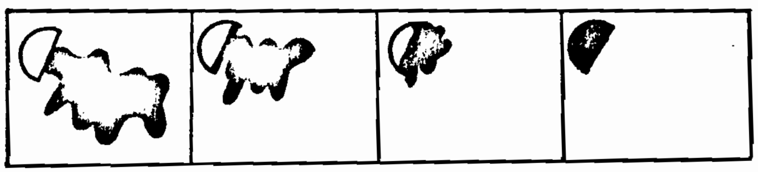
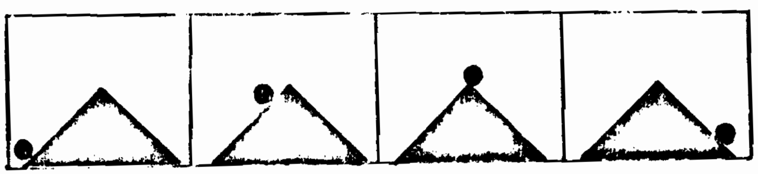
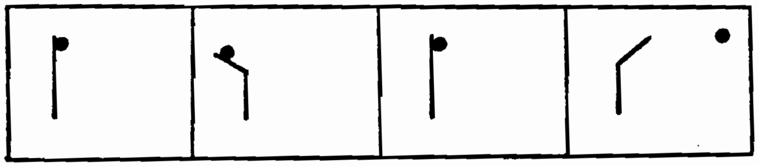
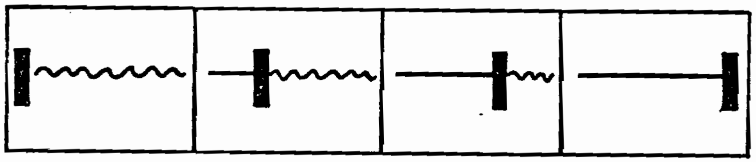
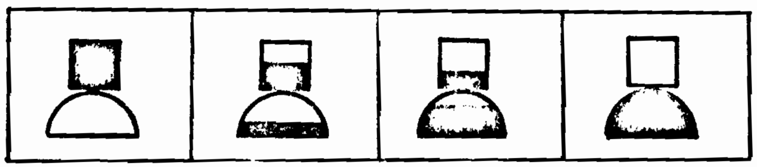
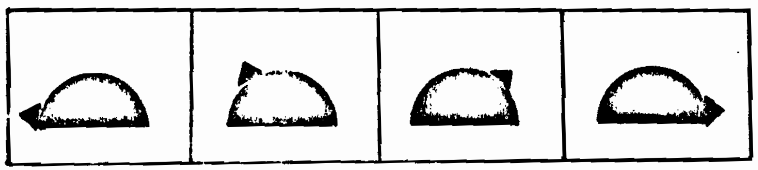
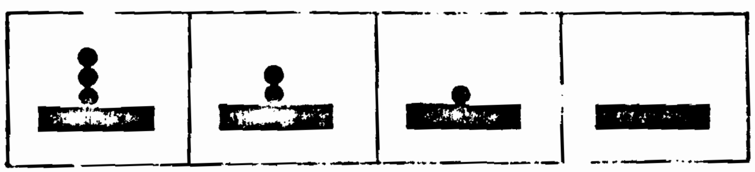
Item	Cartoon	Phrase-Set	No. of Subjects Ranking Phrase 1st, 2nd or 3rd			W
			1st	2nd	3rd	
11.		stacking a set of blocks	20	0	3	.682
		climbing a staircase	0	23	0	
		rowing a boat	2	0	21	
12.		melting butter on a baked potato	22	0	1	.682
		caressing a pussycat	0	18	5	
		jumping a hurdle	1	5	17	
13.		focussing a fuzzy image	12	10	2	.625
		tightening a loose screw	11	11	1	
		tossing a salad	0	2	21	
14.		sopping up spilt milk	20	2	1	.569
		pouring juice into a glass	1	17	5	
		changing a flat tire	2	4	17	
15.		sliding down a bannister	17	3	3	.521
		shaving a beard	1	17	5	
		wrapping a package	5	3	15	

TABLE 1 (continued)

Item	Cartoon	Phrase-Set	No. of Subjects Ranking Phrase 1st, 2nd and 3rd			W
			1st	2nd	3rd	
16.		firing a gun	16	7	0	.516
		bouncing on a trampoline	7	10	6	
		hopping down the street	0	6	17	
17.		ironing a shirt	20	2	1	.516
		minting a coin	0	12	11	
		closing a door	3	9	11	
18.		pouring juice into a glass	19	3	1	.510
		sipping coke through a straw	1	17	5	
		folding a bedspread	3	3	17	
19.		caressing a pussycat	8	11	4	.393
		throwing a switch	14	7	2	
		boiling some water	1	5	17	
20.		diving into a pool	16	6	1	.330
		sopping up spilt milk	1	11	11	
		digging a hole	6	6	11	

[a]Phrases listed in predicted order of preference

- Total number of subjects = 23

tests were performed on the appropriate cell frequencies for each of the five cases indicating lack of preference. None of the values of chi-square (df = 1) were statistically significant in any of the five cases.

D. Discussion

The results of the experiment are quite striking. Overall, subjects showed strong agreement in ranking the goodness-of-fit between phrases and cartoons -- despite the fact that their judgements could not have been based on superficial resemblances between the shapes of the objects in the drawings and the physical features of the objects referred to by the phrases. How then are we to explain the ability of subjects to compute the strength of analogy? The problem is one of accounting, not only for subjects' ability to recognize the best-fitting phrase, but also of explaining the consistency with which they identify second and third choices. In this section we will develop some hypotheses about the nature of this process. No attempt will be made to analyse each of the twenty sets of judgements. Instead, our discussion will focus on those examples which best illustrate specific theoretical issues.

While little is known about how people compute the degree of resemblance across symbolic modes, one kind of analogical reasoning which has received some attention in recent years is the "geometric analogies" problem. In this type of problem, commonly found on intelligence tests, people are asked to find structural similarities between sets of geometric patterns. Geometric analogies are of particular interest to the present thesis because the items to be judged are in the visual mode.

Our understanding of the mental processes underlying this type of task has been deepened by research in artificial intelligence whose goal has been to program a computer to solve geometric analogies. One of the earliest programs to accomplish this feat was that of Evans (1968). Evans' program, called ANALOGY, clearly illustrates a number of general principles of analogical thinking. For this reason a brief review of Evans' work can provide us with a useful starting point for thinking about the problem of how people evaluate the strength of analogies across symbolic modes.[1]

Evans' program is designed to solve analogies of the type shown in Figure 2.1. Such problems are of the form: "A is to B as C is to which one of the numbered patterns?" Successful solutions to such problems hinge on the program's ability to generate good descriptions of the patterns. To solve the problem in Figure 2.1 for example the program's first step is to parse the figures A and B into sub-figures and describe their relations in terms of such relational concepts as: ABOVE, INSIDE, LEFT, and so on. Once structural descriptions of the two figures have been obtained, the program searches for the minimum number of transformations that would change A to B. This includes describing, not only the changing spatial relations between

---

[1]My review is based on Evan's (1968) original article as well as two other reviews of his work in Winston (1977) and Boden (1977).

the objects, but also transformations which explain changes undergone by the objects themselves; e.g., ROTATE, REFLECT, SCALE CHANGE (size transformation). Figure 2.2 shows the structural descriptions for A, B and C while Figure 2.3 gives the rules which transform A into B. Once the transformational rules have been found for changing A to B, they are applied to pattern C. The result is a description of the "ideal" solution. Thus pattern 3 in Figure 2.1 would be chosen as the best fit because it matches closely the ideal description.

The program is a clear example of using descriptions and transformations to search for deeper resemblances underlying surface differences between entities. There are two features of the program worth noting because, as we shall see, they have their counterpart in the case of cartoon-analogies. First, is that the ability of the program to discover the best match depends on the search taking place at the appropriate level of generality. For example, if the transformational rules which describe the change from A to B specify that only squares can be rotated and that only dots can be deleted, then the rules could not be applied to C which contains neither dots nor squares. Fortunately, although the program first tries to seek correspondences between specific objects, it does not fail because of an initially overspecific description. If it fails at first to find an exact match, one of its strategies is to seek a fit at a more general level; e.g., instead of dots and squares, it might search for "a figure being deleted, a figure being shifted and a figure being rotated." A related point is that the greater number of elements and transforma-

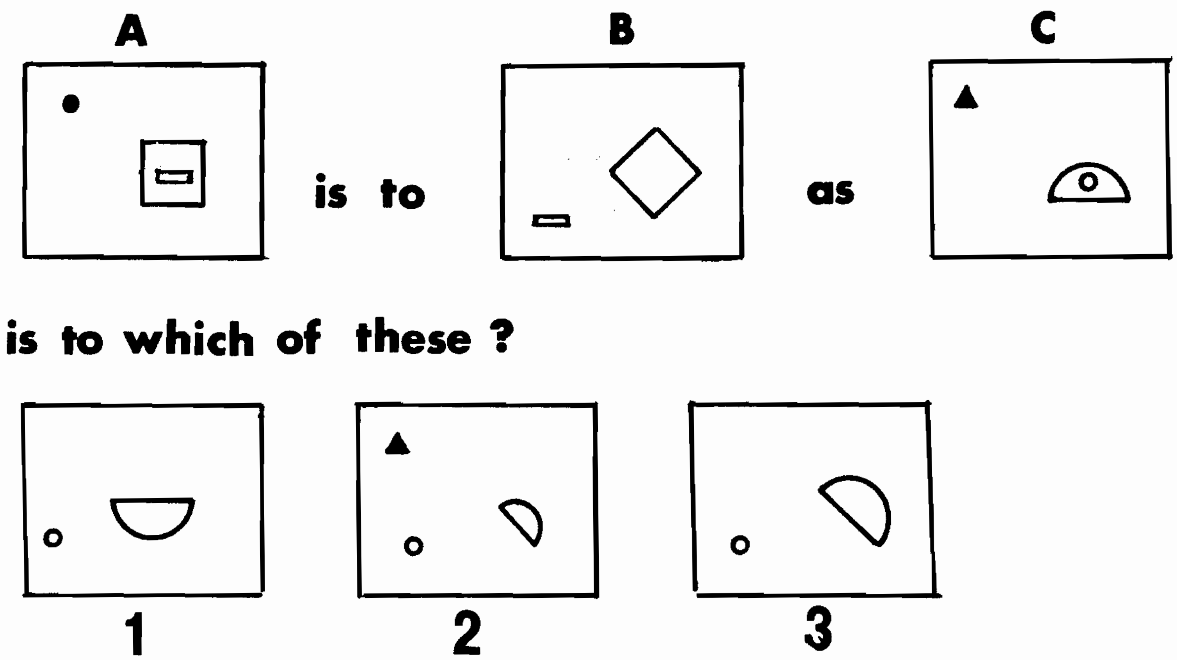


Figure 2.1. - A geometric analogy problem.

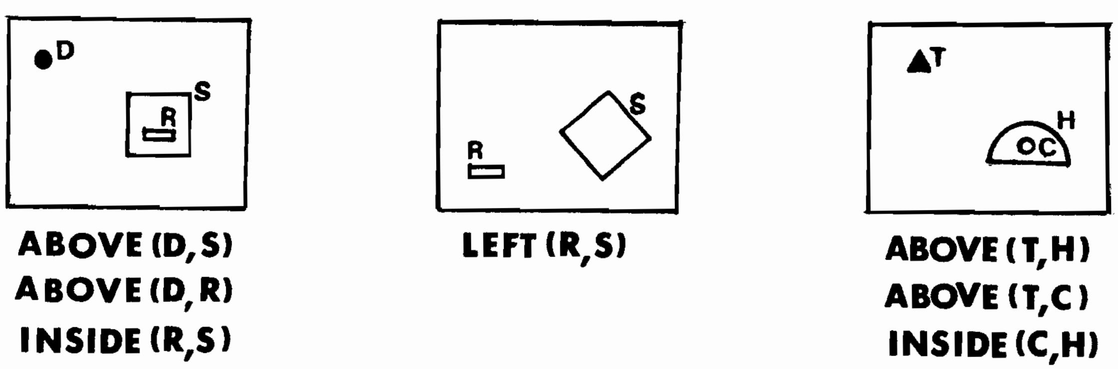


Figure 2.2. - Structural descriptions of the geometric figures.

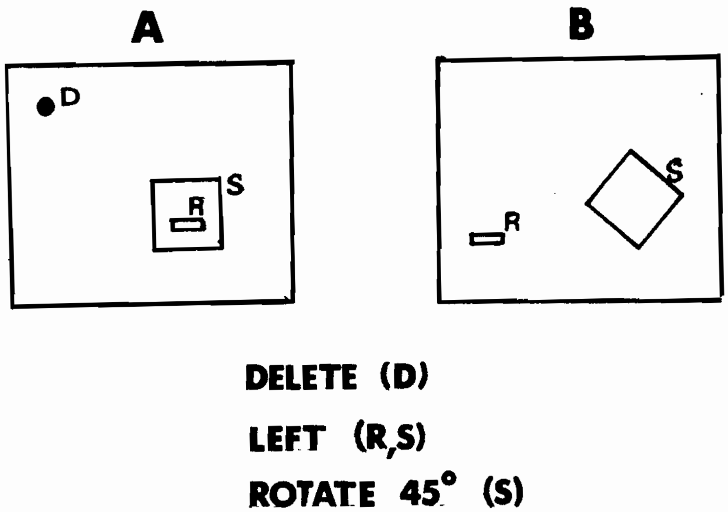


Figure 2.3. - Transformational rules changing A to B.

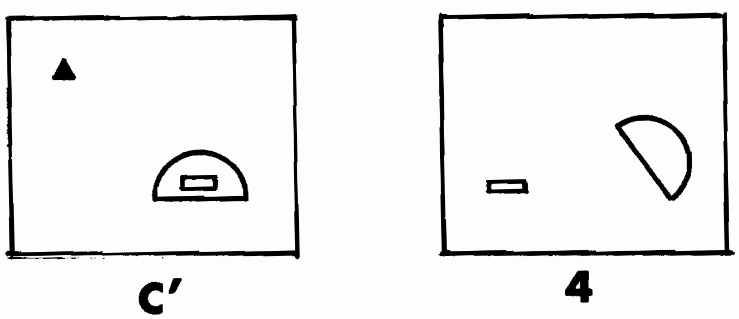


Figure 2.4. - C' and 4 are a closer match to (A,B) than (C,3).

tions that can be preserved, the stronger the analogy. Suppose, for example, that C' and pattern 4 in Figure 2.4 were added to the list of possible matches. In such a case, the pair (C',4), which contains a rectangle is a stronger analogy than (C,3) for the (A,B) pair.

Let us turn now to the more complicated problem of how people evaluate the closeness of analogy between cartoons and phrases. Figure 2.5 gives an example of the kind of problem subjects in the present experiment were asked to solve. One can see from Table 1 that there was perfect agreement in ranking the phrases shown in the Figure as best, next-best and worst fit. Let us trace the steps by which subjects might have arrived at these rankings. In so doing, we will explore similarities between the nature of the present task and the solving of geometric analogies.

Just as with the geometric analogy problem, the first step in the present task is to generate descriptions for the items to be compared. Let us assume that subjects' typically begin by trying to understand the cartoon. The process by which one generates a description for a cartoon analogy was the subject of Chapter 1. All we need to point out here is that the description will be that of an event, not just a description of the objects and spatial relations in the drawing. That is, the representation of the cartoon will assign the geometric elements to roles in an event schema; the representations of the figures in geometric analogies stop at the level of structural descriptions of objects and relations. For example, the left-hand square in the various frames would be assigned the role of

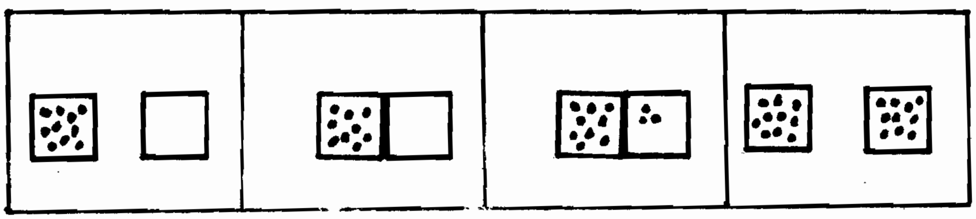


Figure 2.5. - Three interpretations for Cartoon 1 (Table 1) ranked as best (infecting someone with a disease); next-best (spreading jam on bread); and worst (sawing a plank of wood).

"agent," the right-hand square the role of "recipient" and the dots the role of "object." These roles would be part of a general "transfer" schema. Note that while the final output is an event description, earlier stages in processing the cartoon might be identical to certain aspects of Evans' program; namely, finding transformational rules (ROTATE, DELETE, etc.) which account for the changes between the structural descriptions of adjacent frames.

After deriving a description for the cartoon, the next step is to represent the meaning of each of the phrase alternatives. We will assume that the underlying meaning of these phrases would be represented in the same type of conceptual structure as the meaning of the cartoon (see section J); that is, an event structure consisting of a hierarchy of primitive components which define the relations among the objects mentioned in the phrases. Thus, there would be representations for events whose main actions are infecting, spreading and sawing.

Next the underlying representations of each phrase would be compared to that of the cartoon. In a sense, we can think of this stage as similar to the last stage in the geometric analogies program. It will be remembered that this stage consisted of the comparison of each of the candidate descriptions against an "ideal" description obtained by applying the transformation rules to the new figure. The alternative figure which matched closest -- case-for-case -- was chosen as the best fit. In the present task, the description of the cartoon is the ideal representation against which the meanings of the phrases are compared. The phrase whose event description matches closest the semantic structure of the ideal will be selected as the best fit.

One can pursue the analogy with Evans' program even further. Just as Evans' program might fail to find a match at first because of an overly specific description, subjects in the present experiment might fail at first to see any connection between the drawing and the phrases if they were expecting the phrases to refer to dots or squares. Because the subjects do not grind to a halt, we can infer that, as with Evans' program, they are capable of processing the descriptions to deeper levels of generality; instead of dots and squares, diseases and bread, only generalized schemata and "cases" are compared -- e.g., "the transfer of a substance from an agent to a recipient." Comparing the generalized descriptions, two of three candidate descriptions appear as possible matches. Both "infecting someone with a disease" and "spreading jam on bread" contain the deep notion of "transfer."

How might the comparison process choose between the two candidate descriptions? One possibility is that it returns to the drawing once again in order to examine its visual structure more carefully. The purpose would be to see if there is some feature in the cartoon which might distinguish between the two candidate descriptions -- a feature which is true of one of the descriptions but not the other. Let us assume that upon re-examination of the drawing, the following additional feature is noticed: the dots are not actually "transferred" from one square to the other. Transfer implies an actual change of possession of an object from an agent to a recipient. A closer look at the drawing, however, reveals that the number of dots in the left-hand square do not decrease after contact with the right-hand square; this quantity is conserved despite the appearance of dots in increasing numbers on the right-hand square. A better ideal description of the cartoon would preserve this feature -- something like "the transmission of information" rather than the "transfer of a thing."[1]

Armed with this updated ideal description of the cartoon, the meanings of the two phrases are re-examined to see if one of them is a better fit to the new ideal than the other. In fact, the notion of "infecting someone with a disease" satisfies the new requirements of the description better than "spreading jam on bread." In "spreading" there is an actual transfer of possession of a substance (the jam) from an agent (or more precisely an instrument, the knife) to a recipient (the bread). The more of the jam which is on the bread, the less can be on the knife. In "infecting," on the other hand, an entity is transmitted from an agent to a recipient, but the agent does not lose possession of the transmitted entity. You can't get rid of a disease by giving it someone else. Thus "infecting someone with a disease" is a better "theory" for the drawing than "spreading jam" because it accounts for the data in a more exact way.

---

[1]Appendix B from Expriment 1 shows that a number of subjects made this distinction in their interpretations of the cartoon. Compare "the transfer of flies between two bottles" versus the "transmission of knowledge."

One can see an analogy here to Evans' program which, although recognizing how two candidate figures could fit the ideal description, chooses as the stronger analogy, the figure preserving more of the original elements and transformations. This same strategy could be applied in ranking the remaining two phrases. For example, although "spreading jam on bread" does not match the prototypical conceptual structure of the cartoon on as many elements as "infecting someone with a disease," it still matches the deep component of "transfer." The meaning of "sawing a plank of wood," however, does not match even on this more general level and is therefore ranked as the "poorest fit."

One might also argue that the strategy used by subjects in this task differs from Evans' program. For example, it is possible that a subject could find a fit between a cartoon and only <u>one</u> of the three phrases. If so, there would be no way to compare the number of transformations needed to fit each phrase to the cartoon. The only phrase which fits would be ranked best. The choice between the remaining two phrases could be made at random. That this strategy was not used often is reflected by relatively few cases in which there was lack of preference between first and second or between second and third choices.

Another possibility is that two of the interpretations are fit to the drawing and the one with fewer transformations is ranked best. The second choice could be made without a comparison of transformations between the remaining two phrases if a default strategy was used; i.e., if some match cannot be found for the remaining phrase within a certain time, it would be automatically assigned the lowest rating.

At this point one should note that, given a different set of instructions for the task, even next-best and worst choices could be made to fit the cartoon more closely. Put another way, different instructions could induce a subject to try to "assimilate" a cartoon to the phrase. Suppose the instructions had been to try and imagine how the cartoon could be an example of the kind of event expressed by the phrase "spreading jam on bread." These instructions encourage subjects to use their powers of imagination to try and eliminate inconsistencies between the ideal descriptions of the cartoon and the initial representation of the "spreading jam" phrase. The problem, in the case of Figure 2.5, is how to deal with the equal numbers of dots appearing on both squares. One possible strategy would be to transform the conceptual description of the cartoon in various ways. For instance, one could imagine that the two squares represented slices of bread. At first, only the left-hand slice contains the jam, represented by the dots. But then, an unseen agent folds both slices of bread together like a book, once, then twice. This unseen folding action would account for the gradual appearance of the dots in the right-hand square while allowing the left-hand square to look as if the quantity of dots remained constant. The new method of spreading jam on bread assimilates the cartoon more completely than the old description.

In fact, this "folding bread" solution was actually used by a subject in a pilot study where people were asked to speak aloud as they attempted to "assimilate" the cartoon in Figure 2.5 to "spreading jam on bread." The study showed that one can find a fit between almost any phrase and any drawing given enough time and motivation -- although some correspondences are more difficult to achieve than others. How, for example, could Figure 2.5 be an example of "sawing a plank of wood?"

While it is possible to close the semantic distance between a cartoon and a phrase by transforming the descriptions in imaginative ways, it is not necessary to postulate such active operations to explain the "goodness" rankings in the present study. It seems more likely that the general strategy in the present task consists of comparing the underlying representations of each of the three phrases with that of the cartoon to determine the best match. Sometimes this might require closer examination of the cartoon in order to uncover additional features which distinguish between the candidate phrases. This is not quite the same thing, however, as asking someone to use their imaginative skills to see how a cartoon could be <u>like</u> the kind of event expressed by a phrase. In the next chapter an experiment will be described which explores how the attempt to assimilate a cartoon to a phrase can result, not only in narrowing the semantic distance between them, but can also bias subjects toward using a particular transformational rule when fitting a cartoon to a verbal interpretation.

Our analysis of the comparison process for the items in Figure 2.5 is intended to describe a likely strategy used by subjects in making their judgements. This strategy hinges on the ability of subjects to use their abstract knowledge of event classes in computing the degree of resemblance. Let us extend our analysis to two other problems raised by the data in this experiment: (1) the problem of reciprocal choices and (2) the problem of equally good (or poor) matches.

Table 1 shows that there are several items in which the same two phrases are paired with different cartoons. Consider for example items 4 and 6 in the table. One sees that while the phrase "wiping a dusty blackboard" was chosen as best fit for cartoon 4, it was ranked second-best for cartoon 6. Similarly, the phrase "extracting a tooth," while judged best-fit for cartoon 6, was judged second-best for cartoon 4. Such "reciprocal" judgements demonstrate once again that the comparison process can utilize very abstract features of the drawing upon which to base its decisions. For example, "extracting a tooth" is judged as best fit for cartoon 6 because it satisfies more of the ideal conceptual description of the cartoon than does "wiping a dusty blackboard;" e.g., it matches such lower-order descriptions as "a horizontal row of identically shaped objects (teeth), a 'gap' in the row caused by the contact of an instrument, the upward movement of the instrument." "Wiping a dusty blackboard," on the other hand, does not match cartoon 6 on as many specific features (no single horizontal row, no upward movement of an instrument, etc.). It does, however, match on a deeper, more abstract level; namely, the notion of "part of an array of elements being removed by some instrument." Thus "wiping" is seen as a better fit than "twirling a baton" which does not match even on this deeper level. By the same token, "wiping a dusty blackboard" accounts more exactly for the visual structure of cartoon 4 than "extracting a tooth," which was ranked second-best. This pattern of reciprocal choices could be explained if both pairs of phrases and both pairs of cartoons shared a common deep component - i.e., "an instrument removing part of an array of objects" -- while differing on more specific levels of description.

Finally, there were five instances in which there was no clear preference between two phrases as their relative goodness-of-fit to a cartoon. In two cases, the lack of preference was between best and next-best; in three instances there was no clear preference between second and worst choices. The inability to decide which of two phrases is the better fit to a cartoon can be explained with reference to the following two phrases for cartoon 7 in Table 1: "looking at a charging bull" and "falling off a cliff."

Lack of preference would indicate that both descriptions are equally good "theories" for the cartoon; that is, they account for the elements and changes in the drawing equally well. Both phrases provide a way of explaining the increase in size of the dot in successive frames. The major difference between the two interpretations is that "looking at a charging bull" assumes a frontal view of the event in which a stationary observer is looking at a self-propelled, approaching object. In the case of "falling off a cliff," on the other hand, it is assumed that the observer is looking at the ground as he falls toward it. A less likely, but possible interpretation is that the observer is assumed to be standing on the ground looking up at a falling object. That "looking at a charging bull" was a slight favorite indicates, perhaps, that it is more natural to assume that the expanding dot in the drawing represents a mobile, "looming" object rather than a fixed point toward which an observer approaches.

E. Conclusion

This experiment demonstrates that people show strong agreement in their judgements of goodness-of-fit between perceptual and linguistic events even when there are no superficial correspondences upon which to base their judgements. The ability to rank order phrases according to how closely their meanings correspond to "abstract" cartoons is further evidence that abstract relational knowledge mediates our comprehension of both perceptual and linguistic events.

The experiment can also be interpreted as an argument against the idea that our event concepts can be defined by a finite set of criterial features -- even abstract ones. While subjects do see that some cartoons are closer to the prototypical meaning of phrases than others, they are also able to make "fuzzy" matches; i.e., they are able to match phrases with cartoons on the basis of deep, abstract components of their underlying representations (e.g. transfer). The process by which people compute the closeness of analogy across symbolic modes was shown to share some general principles with the process of solving geometric analogies. Both tasks were seen as examples of problem-solving based on the intelligent comparison of descriptions.

In this experiment people were asked to compute the goodness-of-fit between phrases and cartoon analogies. While requiring cognitive contributions on the part of the comprehender, it was pointed out that evaluating the closeness of analogy is not exactly the same thing as asking someone to imagine how a cartoon could become an analogy for a phrase. The former task involves finding a match at some level of abstraction; the latter task requires something more as well -- the resolving of possible inconsistencies between the initial descriptions of the items being compared. Let us turn now to an experiment which explores how the attempt to resolve inconsistencies can call upon transformational mechanisms which, in turn, can influence the ability to perceive resemblances in novel cases.

CHAPTER III

THE PSYCHOLOGICAL REALITY OF TRANSFORMATIONS IN THE COMPUTATION OF ANALOGY

A. Experiment 3 -- Introduction

A common theme of Experiments 1 and 2 is that transformations play an important role in the discovery of resemblances. In Experiment 2 we argued that evaluating the closeness of analogy between phrases and cartoons involves active operations on the underlying descriptions of both visual and verbal inputs (e.g., generalizing an overspecific description). Similarly, the ability to transform a description was seen to play an important role in Experiment 1 where people were asked to derive their own interpretations for cartoon analogies (e.g., assuming a viewpoint in fitting a cartoon to an event schema).

The present experiment extends the investigation of the role of transformations to a somewhat different task. We pointed out in the previous chapter that there is a difference between evaluating the closeness of fit between a drawing and a phrase and the task of trying to imagine how a cartoon could be like an event described by a phrase. The former task imposes fewer constraints on the degree of resemblance than the latter. For example, it is possible to recognize that Figure 2.5 shares the deep component "transfer" with the phrase "spreading jam on bread" without actually imagining the elements of the cartoon as representing knives, jam and bread.

One can achieve an abstract level of matching and still be aware of specific inconsistencies. But to actually perceive the cartoon as an analogy for "spreading jam on bread" requires more work; it demands using one's imagination to create conditions that make possible a closer equivalence of descriptions. The problem of interest in the present experiment is of this latter kind. A subject is given two inputs -- a phrase and a cartoon -- and is asked to imagine how the cartoon could be an analogy for the phrase. Solving such problems requires using one's imagination, reshaping internal representations in order to find resemblances which may not be apparent at the outset.

It is worthwhile pointing out that such a task, which depends on using the imagination, has the same general form as the mathematical or logical kinds of task upon which well-known theories of problem-solving are based. For example, in Newell and Simon's (1972) studies of human problem-solving, a typical task is where a subject is asked to prove a theorem in logic given another theorem as the starting premise. Solving such problems requires discovering differences between a goal-state (the to-be-proved theorem) and the current-state (the given theorem) and finding means of eliminating those differences. Newell and Simon's General Problem Solver (GPS) is a computer program which can handle such tasks. Its basic strategy is called "means-ends" analysis. This strategy works by first identifying the major differences between the goal and the current-state. Beginning with the most important difference, GPS searches its data base for operators (rules of logic) that can legally be applied to eliminate the major difference; if such an operator is found right away, the problem is solved. If the relevant operator cannot legally apply to eliminate the difference directly, GPS sets up a "subgoal" -- i.e., it asks if there is a way to change the current state to one in which the relevant operator might apply. The program then tries to eliminate the new difference between the current state and the new subgoal. If the operator cannot apply at the level of subgoal, a sub-subgoal can be created until a difference is found which can be eliminated. Thus GPS is "recursive" in that its difference-eliminating procedure can call upon itself in trying to reach its various subgoals.

The task of finding an analogy between a cartoon and a phrase can be thought of as a GPS type problem. The problem is one of getting from the current state (the initial underlying description of the cartoon) to a goal state (the description of the event triggered by the phrase). In trying to reach the goal, the problem-solver may encounter differences or inconsistencies between the underlying descriptions and must call upon the appropriate "operators" or "transformations" to eliminate them. If the appropriate operators cannot apply directly, subgoals may be established and new operators applied in recursive fashion. The goal of the experiment to be described below is to demonstrate the psychological reality of transformational mechanisms in the context of finding analogies between linguistic and perceptual events. Of course, the transformations in this type of task are not the logical operators of GPS, but rather the ones people use in making sense of the visual world, including the mini-world of cartoon analogies. Some of these transformations were discussed in Chapter 1 where it was pointed out that transformations play an important role in visual perception in that they "explain" certain kinds of changes or distortions which ideal forms may have undergone in particular contexts: changes in orientation, position, color, shape, and the like. For example, knowing about viewpoint transformations would allow the identification of an ideal form (a circle) despite a surface distortion -- its manifestation as an ellipse in the visual input.

We propose that transformations such as "seeing it from an assumed viewpoint" have the status of independent rules or operators which can be called upon during the search for resemblance to modify descriptions. If our claim that transformational rules have independant status is true, then it should be possible to manipulate them experimentally -- to show that they are separate components to be reckoned with in theories of resemblance and meaning. The experiment which follows is designed to test this hypothesis.

We have chosen as our test case transformations involving "assumed viewpoint." In addition to examining viewpoint transformations, the experiment will permit us to explore a number of related hypotheses concerning the comprehension of events through linguistic and perceptual media. These hypotheses can best be explained with reference to the following examples.

(a) The comprehension of an event through the medium of language may activate a prototypical vantage point from which the event is assumed to be witnessed. For example, it is more likely that the description "an egg falling off a table" would give rise to a mental representation in which the event was being imagined from a "side-view" rather than, say, a "worm's eye" view in which someone is lying on the floor looking up at a falling egg.

While both points of view generate possible descriptions of the event, they specify different surface features. These differences are characterized by the cartoons in Figure 3.0. To use a linguistic analogy, we might say that the cartoons represent two different surface structures derived from a single, underlying deep structure, the abstract notion of an object falling to the ground. Pursuing this analogy, just as "active" and "passive" surface structures of sentences were thought by Chomsky (1965) to be related to a single deep structure through the application of transformational rules, so viewpoint transformations serve to relate two different surface structures (the cartoons) to a single underlying event schema.

(b) The comprehension of an event through the medium of cartoons also activates prototypical viewpoints. Consider Figure 3.1. Without accompanying verbal context, most likely this cartoon would be interpreted as an event seen from a side-view; e.g., as a good fit to the phrase "a foot stepping on a bug." But, although less probable, it is also possible to make this drawing fit a different kind of event -- "a car colliding with a pole" -- by assuming a different vantage point; namely, seeing the event from above or a top-view. In terms of the linguistic analogy, here is a case where a single surface form -- the a

b

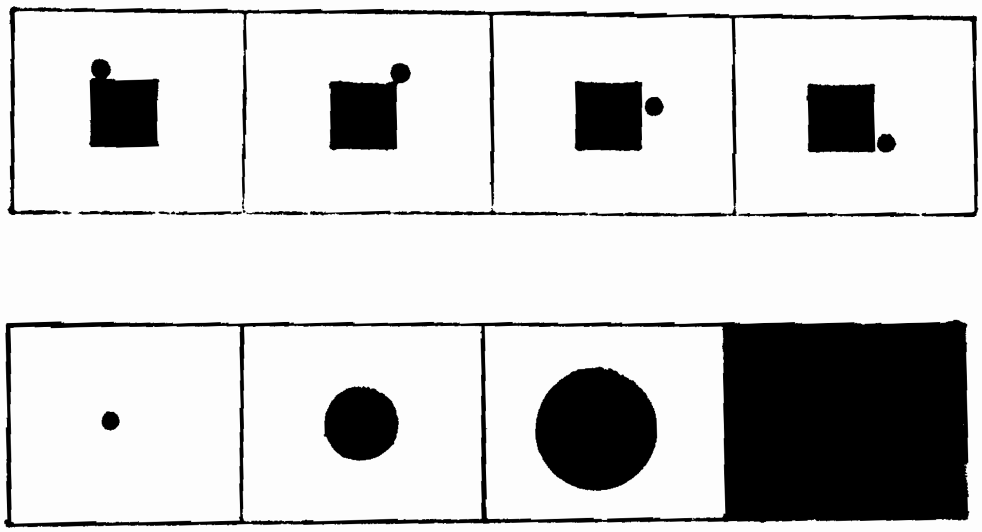


Figure 3.0. - Two cartoons representing the event "an egg falling off a table" from: (a) a side-view and (b) a "worm's eye" view.

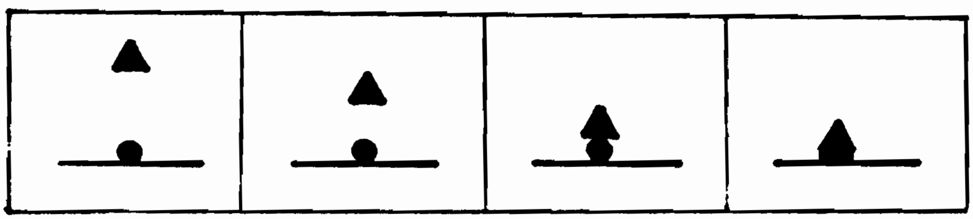


Figure 3.1. - A single cartoon representing two different events:

(a) "A foot stepping on a bug" (side-view);
and
(b) "A car colliding with a pole" (top-view).

cartoon -- can be derived from two deep structures: "colliding" versus "stepping."

(c) Note that if the phrase "a car colliding with a pole" was encountered by itself, without the accompanying drawing, it is unlikely that the "mind's eye" description of the event would be a top-view. This reveals an interesting property of the relation between language and perception which often goes unrecognized in traditional accounts of how perception and language are linked. The process of comparing a linguistic description against a perceptual event is not always well characterized as a process of "verification" -- an all-or-none comparison of the internal representations of visual and verbal inputs in order to determine their truth value. (See section E for a critique of the sentence-picture verification paradigm). Instead, linguistic and perceptual knowledge may interact in the course of searching for a resemblance. For example, the attempt to find a match between a "a car colliding with a pole" and the cartoon in Figure 3.1 may at first be unsuccessful. Their representations do not seem to fit; they seem to describe different sorts of events. Suddenly, imagining the event represented by the cartoon from a top-view allows both representations to mesh. The triangle is perceived as a car seen from above, colliding with a pole represented by the disc. In the course of trying to discover a resemblance, the scope of both event schemata have been widened to include each other. This interaction between event schemata is reminiscent of Piaget's (Furth, 1969) notion that schemata undergo accommodation in the assimilation of other schemata.[1]

(d) Finally, it is proposed that viewpoint transformations are abstract rules which can be manipulated as independent components during the attempt to find analogies across symbolic modes.

This last claim leads to a testable hypothesis about the psychological reality of viewpoint transformations. The basic idea is as follows. A number of cartoons will be designed which, although typically giving rise to side-view interpretations, could also fit an event seen from a top-view. Figure 3.1 is an example of such a cartoon which can be assimilated to either viewpoint: a side-view ("a foot stepping on a bug") or a top-view ("a car colliding with a pole"). Suppose we take a number of such cartoons and pair them with their top-view phrases. Then each pair is presented to a subject whose task is to find an analogy between them. If finding the analogy requires discovering a top-view transformation, then over a number of trials we might be able to induce a perceptual set whereby a subject attempts to assign a cartoon a top-view description even though a side-view would normally be the prototypical way of understanding the cartoon. The effect of perceptual set should be reflected in the amount of time a subject spends trying to find the analogy for a particular cartoon-phrase pair. For example, if after several trials of

---

[1] It is also an instance of what Black (1962) called the "interactionist" view of metaphor. To say "man is a wolf" not only makes the man more wolf-like, but the wolf more human. Both schemata have been widened to accommodate each other.

fitting a cartoon to side-view phrases, a trial is encountered where a cartoon is paired with its top-view phrase, one could predict an increase in the time it takes to see the resemblance. Furthermore, this increase should be greater than the time it would ordinarily take if all the previous cartoons had been processed as top-views.

Similarly, if after several top-view trials, a subject encounters a cartoon paired with its side-view phrase, he should need more time to see the analogy than if all the previous trials had been side-view trials. Thus evidence for an effect of "set" would be an increase in the time spent on the "shift" trial compared to what the time might have been on the same trial had there been no shift.

The attempt to develop in subjects a set or expectation which leads to processing a cartoon from a particular viewpoint is, in some respects, analogous to an earlier experiment by Mehler and Carey (1967) in the processing of sentences. Mehler and Carey demonstrated that it was possible to create in subjects an expectation for a specific type of syntactic structure in the comprehension of sentences. In their paradigm, subjects were presented with a series of eleven sentences embedded in noise. For the control group, all eleven sentences were of the same syntactic form. Either sentences like: (1) "They are forecasting cyclones" or (2) "They are conflicting desires." In (1) the phrase "are forecasting" forms a syntactical unit but "forecasting cyclones" does not. In (2), on the other hand, "conflicting desires" is a unit but not "are conflicting." For the experimental group, the first ten sentences had the same syntactic structure but the eleventh sentence had the contrasting phrase structure. Thus, ten sentences of the form "they are conflicting desires" were followed by the constrasting form "they are forecasting cyclones." Subjects in the experimental group had more difficulty understanding the eleventh sentence than subjects in the control condition. Mehler and Carey attribute this difficulty to a perceptual set induced by the first ten trials. They cite this result as evidence for the psychological reality of syntactic structure in the perception of sentences.

A second experiment by Carey, Mehler and Bever (1970) used the perceptual set paradigm to study the processing of ambiguous sentences. In this study, subjects were set to expect a sentence with either of two syntactic structures: (1) predicate nominative ("they are sleeping lions") or (2) transitive verb ("they are discussing books"). Each sentence type was paired with a picture which would make it either true or false. On the final trial subjects heard an ambiguous test sentence such as "they are visiting sailors" also paired with a picture. The experiment showed that subjects could be preset to perceive only one of the two ambiguous meanings of the test sentence so that their latencies in verifying the picture were identical to subjects receiving unambiguous material.

The present experiment, while using the paradigm of perceptual set to bias perceived meaning, differs from the sentence processing studies in two important ways. First, is that we have no hypotheses concerning the grammatical structure of sentences or the psychological reality of linguistic transformations. Rather, our goal is to show the existence of transformational rules, based on abstract perceptual knowledge, that can influence the semantic representations people assign to visual inputs. Applying different transformations can result in the assignment of different semantic deep structures to the cartoons. A second difference concerns the process by which the perceptual set is established. In Carey, *et al*, for example, it is the syntactic structure of the sentences alone, experienced over trials, which induces the set. The presence of the picture does not contribute to the induction of expectation. In the present experiment, on the other hand, an "interaction" between the linguistic and pictorial message is essential for the abstraction of the transformational rule to occur.

To summarize: The objective of this experiment is to demonstrate the psychological reality of transformations in the comprehension of analogy across symbolic modes. The experiment will attempt to create a perceptual set in subjects by inducing them to process a cartoon from either of two assumed viewpoints -- a side-view or a top-view -- over several trials. We will accept as evidence for the effect of set an increase in time spent finding the analogy on a final, "shift" trial where the viewpoint is switched to the contrasting condition. The average latency for shift trials should be significantly greater than that for non-shift trials within the same viewpoint condition.

B. Method

Materials. Ten cartoon analogies were designed according to the procedures described in Experiment 1. The cartoons were photographed and mounted as 35mm slides. Each cartoon had the property that it could be seen as representing a different event depending on whether or not one imagined seeing it from a *side-view* or a *top-view*. For each cartoon, two phrases were constructed describing two different types of everyday events. One phrase matches the cartoon as a side-view; the other phrase fits the cartoon as a top-view. Figure 3.2 shows the ten cartoon stimuli accompanied by their side and top-view phrases.

Subjects. Forty undergraduate and graduate students from McGill University served as subjects in the study. They participated as unpaid volunteers.

Procedure. Each subject was assigned to one of two "set" conditions: (1) a "side-view" condition in which the first nine trials required matching a cartoon with its side-view phrase, or (2) a "top-view" condition in which the first nine trials required matching a cartoon to its top-view phrase. The tenth and final trial for both conditions was a "shift" trial which paired a cartoon with the phrase from the contrasting viewpoint. Figure 3.3 gives an example of one experimental run for the side-view condition. Each subject saw a different order of cartoon-phrase pairs such that each pair appeared an equal number of times in each of the ten serial positions. In this way, the comparison of shift versus non-shift trials would be based

FIGURE 3.2

CARTOON ANALOGIES WITH THEIR SIDE AND TOP VIEW PHRASES

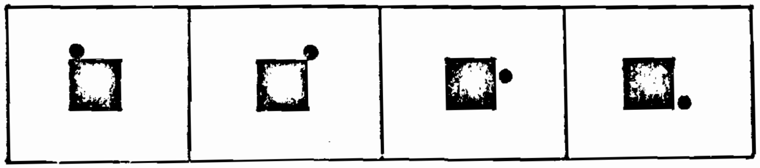
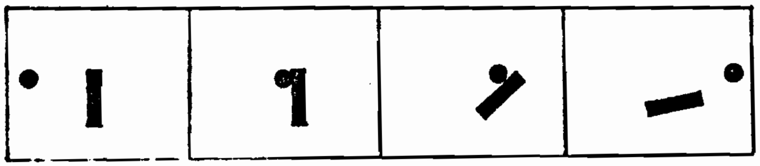
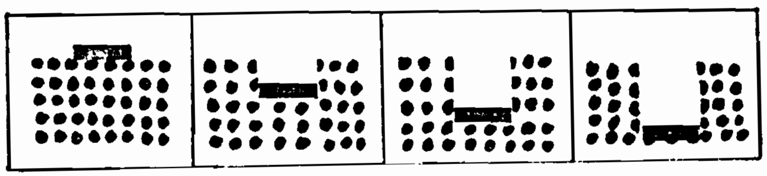
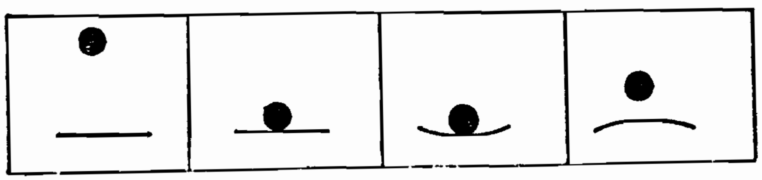
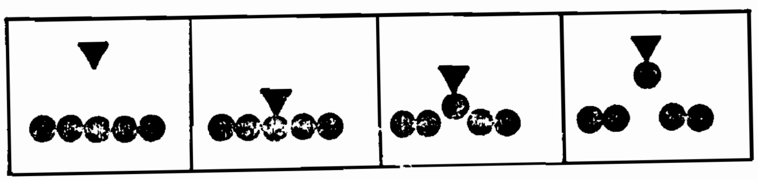
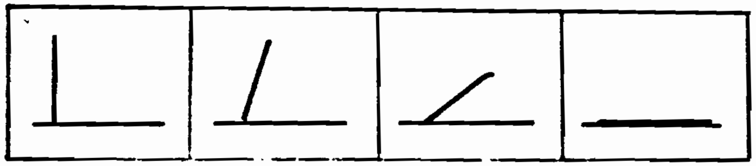
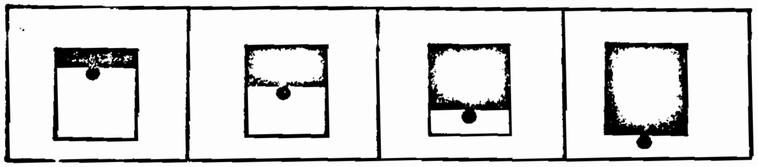
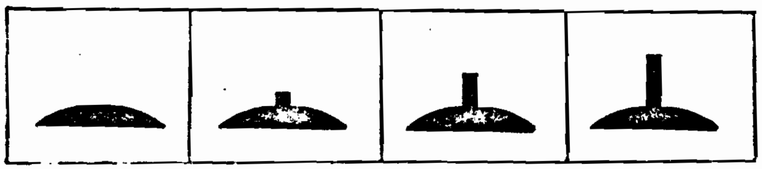
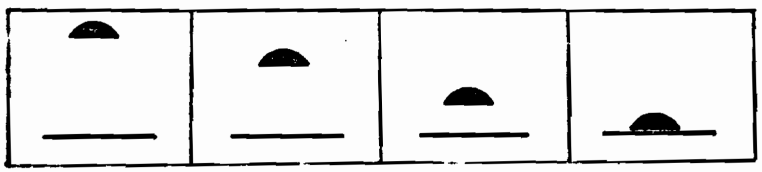
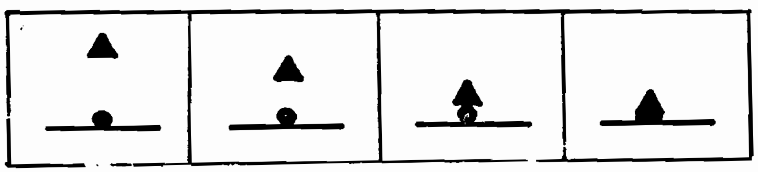
Cartoon Analogy	Side-View Phrase	Top-View Phrase
	AN EGG ROLLING OFF A TABLE	A TRUCK ROUNDING A SHARP TURN
	GETTING HIT ON THE HEAD WITH A GOLF BALL	A CUSTOMER PASSING THROUGH A TURNSTYLE
	A STONE SINKING INTO QUICKSAND	A BULLDOZER CLEARING SNOW
	A BOY BOUNCING ON A TRAMPOLINE	A COP TRYING TO BREAK DOWN A DOOR
	EXTRACTING A TOOTH WITH FORCEPS	CHOOSING A PARTNER AT A DANCE

FIGURE 3.2 (continued)

Cartoon Analogy	Side-View Phrase	Top-View Phrase
	A TREE TOPPLING TO THE GROUND	WIND BLOWING AGAINST AN OPEN GATE
	PULLING DOWN A WINDOW SHADE	COVERING A MATTRESS WITH A SHEET
	A RABBIT POPPING ITS HEAD OUT OF A HOLE	A SNAKE STICKING OUT ITS TONGUE
	A ROCKET LANDING ON THE MOON	A SWIMMER COMING TO THE EDGE OF A POOL
	A FOOT STEPPING ON A BUG	A CAR COLLIDING WITH A POLE

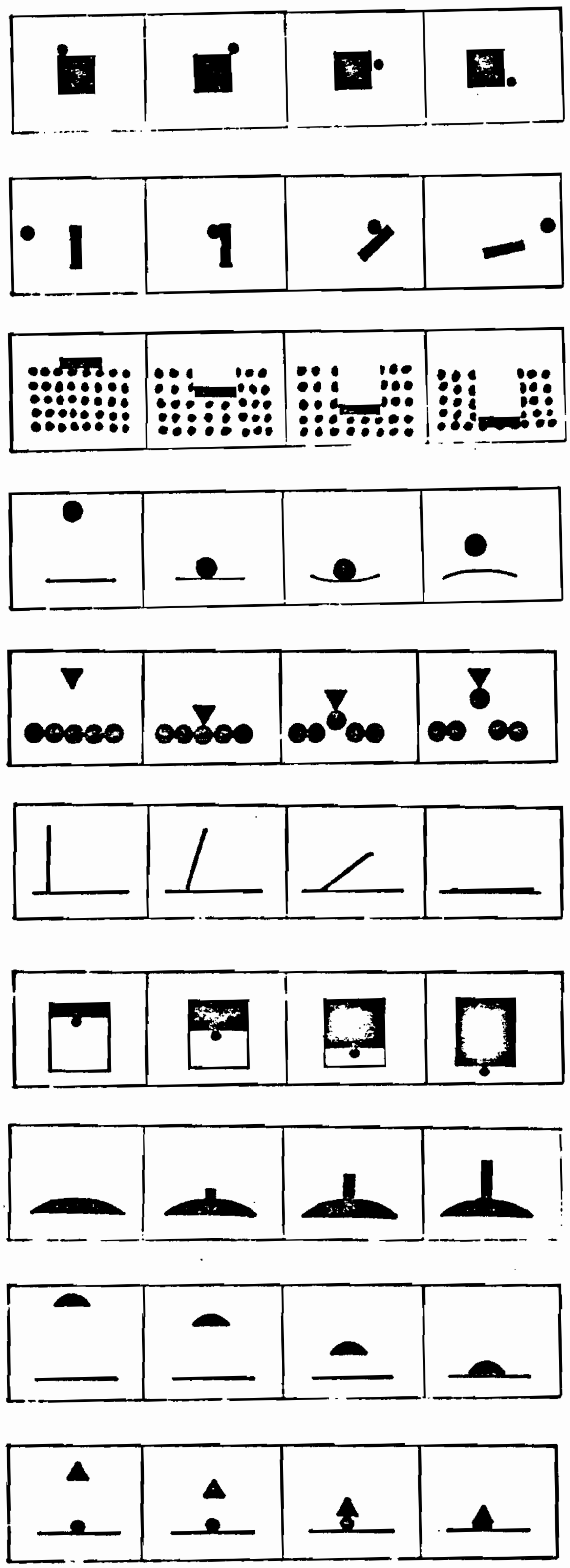


Figure 3.3. - Phrases and cartoons for one subject in the side-view condition. The last trial shifts to a top-view phrase.

on scores for each of the ten cartoon-phrase pairs.

Subjects were told that they would see a series of phrases describing everyday events, each phrase being followed by a cartoon-strip. (Full instructions are given in Appendix A). They were told that their task was to see how the cartoon could be an analogy for the event described by the phrase and that the time it took them to find the analogy would be recorded. Subjects then went through a practice trial in which they were familiarized with the nature of the cartoons and instructed in using the reaction time apparatus. On each trial, the sequence of events was as follows: (1) the subject read the phrase projected on the screen; (2) the subject responded "OK" when he understood the phrase; (3) the phrase disappeared; (4) the cartoon-strip was projected on the screen; (5) the switch advancing the cartoon slide simultaneously started a timer; (6) the subject pressed a button as soon as he found an analogy; (7) the button-press stopped the timer; (8) the elapsed time was recorded; (9) the next trial began.

Although no limit was placed on the time a subject could take responding, they were asked to respond immediately once they were sure to have found an analogy for the cartoon.

Slides were projected by a Kodak Carousel projector onto a screen about 3 feet in front of the seated subject. The elapsed time between the onset of the cartoon slide and the subject's response was measured by a Lafayette clock, model 54417-A.

C. Results

The results of the experiment are portrayed graphically in Figure 3.4. The figure plots the mean reaction time for each of the ten trials in both side and top view conditions. The main question of interest concerns the predicted difference in latency between the shift and non-shift trials: for example, is a subject worse off having been shifted suddenly to a top-view on trial 10 than he would have been if he had also been seeing the top-view on trials 1 to 9? This question can be most clearly grasped by looking at the regions encircled by the hashed lines in Figure 3.4. Each region encompasses the scores for trials 5 through 10 for the same viewpoint condition.[1] Notice that the mean latency for the shift (10th) trial is greater than those for each of the five previous trials within the same viewpoint condition (although not from the same subjects). The hypothesis is that the mean of the shift trials is significantly higher than those of the previous five trials.

A two-way analysis of variance was performed on the latency scores for trials 5 through 10.[2] The analysis tested for a main effect

---

[1] Only scores for trials 5 through 10 are included because the generally decreasing latencies for the first four trials indicate the presence of a practice effect which begins to level off at trial 5.

[2] Because of the particular crossover design used in this experiment where a subject is switched to the contrasting viewpoint condition on a final trial, the usual way of partitioning out subject variance in a treatment-by-subjects ANOVA cannot be used because subjects are not actually nested in viewpoint conditions. However, a correction procedure applied prior to ANOVA had the desired result of partitioning out subject variance. The correction procedure is outlined in Appendix C.

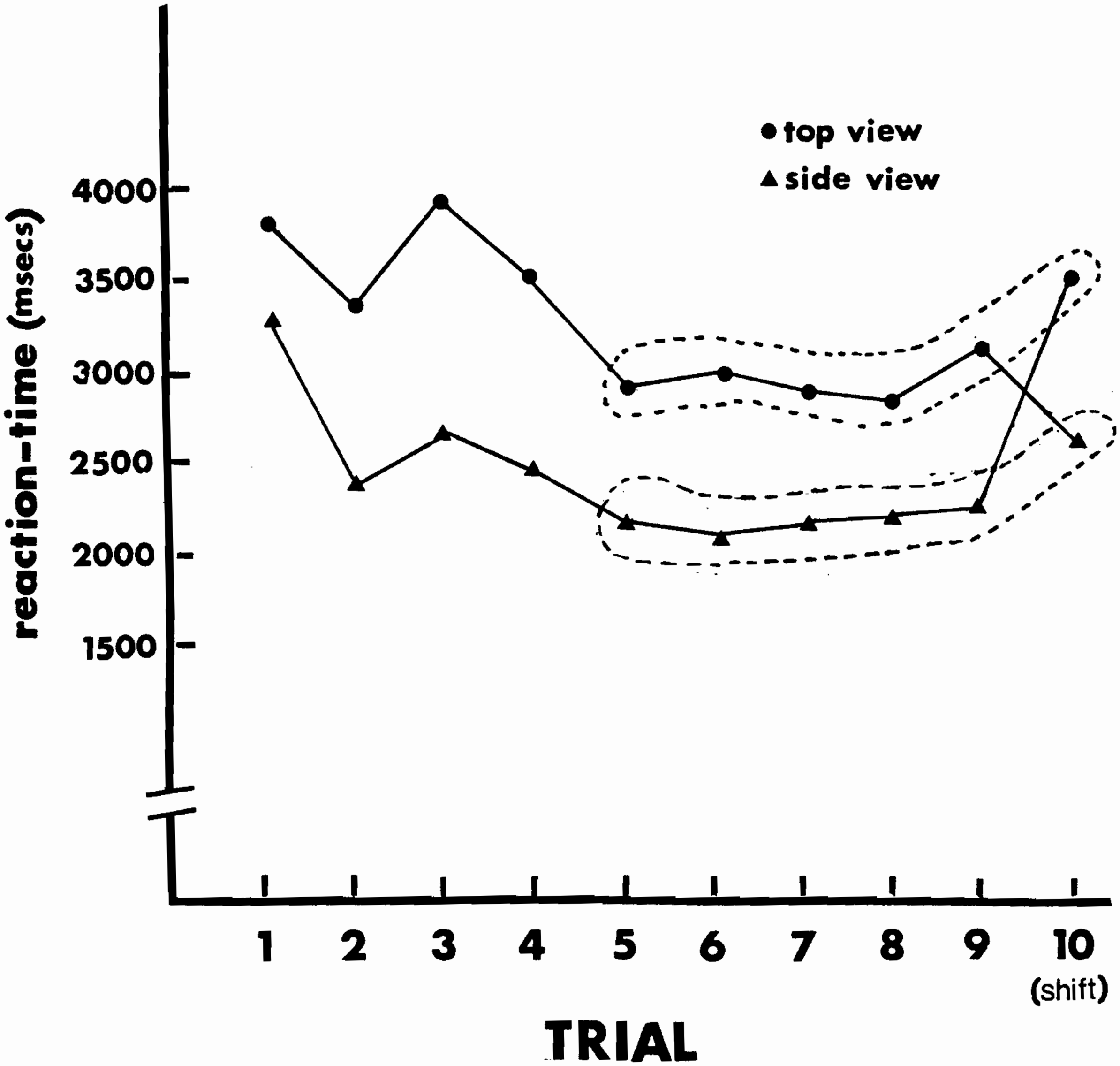


Figure 3.4. - Mean reaction-time to find an analogy between a phrase and a cartoon for two viewpoint conditions. The tenth trial in each condition is a "shift" trial pairing a cartoon with the phrase from the contrasting viewpoint. The solid lines join scores for the same subjects, while the hashed lines encircle scores in the same viewpoint condition.

of conditions (side-view, top-view); a main effect for trials (5 through 10); a trials x conditions interaction; and most importantly, a test of the hypothesis -- a planned comparison of the shift versus non-shift trials.

The results of the analysis of variance are summarized in Table 2. There is a large main effect of conditions, $F(1,227) = 56.89$, $p < .001$; no main effect for trials, $F(5,227) = 1.81$, $p < .10$ and no trials x conditions interaction, $F(5,227) = .506$. The planned comparison of the means for the shift versus non-shift trials proved significant, $F(5,195) = 7.78$, $p < .001$.

D. Discussion

The results support the claim that transformations are psychologically real and can affect our ability to find analogies between perceptual and linguistic events. Subjects can be biased to attempt to assign a meaning to a cartoon consistent with a particular point-of-view. A sudden shift to a phrase specifying a shift in assumed viewpoint requires recomputing the description of the cartoon so that a match can be achieved. This additional recomputation is reflected by significantly longer reaction times for the shift trials compared with the non-shift trials within the same viewpoint condition. Let us discuss in more detail the nature of this recomputation.

Prototypical Viewpoints. The large main effect for conditions supports the hypothesis that it is easier to match cartoons to

TABLE 2

SUMMARY OF ANALYSIS OF VARIANCE FOR EXPERIMENT 3

Source	d.f.	Mean Square	F
Conditions	1	49,417,005	56.89
Trials	5	1,573,593	1.81
Conditions x Trials	5	435,313	.501
Error	227[a]		

Planned comparison of Shift versus Non-Shift trials, $F(1,195) = 7.78$[b]

[a]Error term has adjusted degrees of freedom due to the correction procedure described in Appendix C.

[b]Critical value of F tested under reduced degrees of freedom due to correction procedure described in Appendix C.

their sideview than to their top-view phrases. Figure 3.4 shows that reaction times for top-view trials take consistently longer than side-view trials. In addition, a switch from a side-view to a top-view trial produces a large increase in reaction time while a shift from a top to a side-view shows a slight drop. Why should it take longer to match cartoons to their top-view descriptions?

This result can be explained by making two assumptions about the nature of our event concepts: (1) our knowledge of event classes includes information about how events look from various viewing angles; and (2) some viewing angles are more typical than others, in particular side or frontal views. The assumption of prototypical viewpoints triggered by verbal descriptions stems from the following reasoning. Because we perceive the visual world mostly from an upright, vertical orientation, our experience of events is more often from a side or straight ahead perspective than from the air or lying on the ground. If this is so, it is possible that our event concepts come to reflect this preferred spatial orientation. That is, one of the "slots" in an event schema may be that of "typical viewpoint" and, unless there is contextual information specifying otherwise, the prototypical "default" value for viewpoint will be a side-view.

To cite our earlier example: the surface representation of the event activated by the phrase "an egg falling off a table" would most likely be from a side rather than from a "worm's eye" view, unless context specified another viewpoint. In other words, all things being equal, the comprehension of an event through language will assign a side-view description to an event.

The same assumption holds for the encoding of the cartoon-strips. We pointed out in Chapter 1 that the flat picture plane of a drawing can be interpreted in terms of Arnheim's "horizontal space" or "vertical space." Vertical space is consistent with a side view; horizontal space with a top-view. As in the comprehension of phrases, it appears more natural to interpret cartoon analogies in terms of vertical space -- as representing side-views -- than as depicting top-views. Thus, although it is possible to imagine the cartoon in Figure 3.1 as representing either viewpoint, the side-view would likely be the preferred encoding.

Postulating that the default assignment for both phrases and cartoon-strips is a side-view helps explain the relative ease of matching cartoons to their side view phrases. If both cartoons and phrases have initially been encoded as side-views, the comparison process proceeds smoothly because no inconsistencies in viewpoint are revealed. The case of top-view trials is quite different, however. Let us characterize the processing of a typical early top-view trial as follows: (1) the phrase elicits a prototypical side-view description; (2) the cartoon is also encoded initially as a side-view; (3) a comparison of the underlying descriptions results in a mismatch; (4) a search process begins to look for ways of resolving the mismatch; (5) one strategy is to imagine the event described by the phrase from a different viewpoint, say a top-view; (6) if the new viewpoint accounts for the visual features of the drawing, a fit is achieved and the prototypical side-view is abandoned; (7) the comparison process notes that a top-view transformation has been effective in finding the analogy and stores this information for use in later trials.

This account would explain why top view trials take longer. But note that to account for the data on the final shift trials, one has to assume, not only that side-views are inherently easier than top-views, but that there has been an effect of set. For example, Figure 3.4 shows that the shift from a top to a side-view trial produces a slight drop in latency. This drop can be explained by the assumption that side views are inherently easier to match for the reasons stated above. However, the fact that the final side-view trial is still significantly slower than the mean of the five previous side-view trials (bottom curve) is attributable to an effect of set; i.e., at least some of the subjects first try to process the cartoon as a top-view but quickly discover that the prototypical side-view description is a good fit.

Figure 3.4 shows that shifting from a side to a top-view trial produces a large increase in reaction-time. Again, the inherent difficulty of matching top-views can account for part of this increase. But note that the mean for this final shift trial surpasses the mean of the five previous top-view trials (top curve). This additional increase in latency argues strongly for an effect of set.

Finally, post-experiment interviews revealed that view-point transformations may play a role in the search for resemblance without subjects necessarily being aware of it. A few subjects did report being aware of having to switch viewpoint on the final shift trial. Most subjects, however, while aware that there was something different about the last trial, were not conscious of a viewpoint change compared to the previous trials until it was pointed out to them by the experimenter. The fact that people are not conscious of using viewpoint transformations does not necessarily mean that they are not psychologically real. After all, we are not usually aware of using perceptual transformations in making sense of scenes or of syntactic transformations in processing sentences.

E. Conclusion

In concluding this chapter we will discuss some implications of the present experiment for several studies whose broad concern has been the mental processes involved in interpreting linguistic and pictorial messages.

Theories of Sentence-Picture Comparison. Clark and his colleagues have undertaken a program of research in recent years to examine the nature of the interface between language and perception (Cf. Clark, Carpenter and Just, 1973 and Clark and Chase, 1972). These authors have proposed detailed models of the process by which people compare various types of perceptual and linguistic inputs. Among the important assumptions of these models is that both language and perception are avenues into a single, abstract representational system and that any comparison of visual and verbal material must necessarily occur in this common format. They envisage this common format in terms of elementary propositions which encode the names of objects, properties and relations. For example, both the sentence "the star is above the plus" and the picture ($\overset{*}{+}$) would be represented by the proposition, "star (above) plus."

In addition to the notion of a common representational format, a second assumption of Clark, *et al* is that our perceptual experience is mediated by an organized spatial framework derived from the fact that we stand upright, have eyes in front of our heads, walk in a forward direction, use ground level as reference point, etc... This spatial framework influences our perceptual encoding of physical dimensions. For example, the picture ($\overset{*}{+}$) showing two objects arrayed along the vertical axis would be encoded as "star (above) plus" even though a second encoding, "plus (below) star" is also a true description of the scene.

In support of these hypotheses, Clark *et al* have conducted experiments in which a subject is typically presented with a sentence (e.g. "the star is above the plus") and asked to verify it against a picture ($\overset{*}{+}$); that is, say whether or not the sentence is true or false of the picture. The time taken to respond is partitioned into the various operations assumed to make up the comparison process.

The present study shares certain views put forward by Clark, *et al*; namely, the notion that both visual and verbal inputs undergo a process of interpretation whose output is an abstract representation. However, the present experiment also departs from Clark *et al's* approach in several important respects.

The first difference concerns the nature of the stimuli used as a basis for theory building. Clark *et al*'s goal is to develop a "general" theory of how people compare sentences with pictures. Yet the verbal and visual materials used in their experiments seriously limit the potential generality of such a theory. It is true that the configuration $(\overset{*}{+})$ technically qualifies as a picture or diagram in that it portrays a spatial relation between two objects. It is also true that "star above plus" is a meaningful sentence describing a spatial relation. Yet one should be wary of basing a general theory on such restricted kinds of stimuli. They lack what Neisser (1976) calls "ecological validity."

A general model of how people comprehend events through linguistic and pictorial media must go beyond explaining how people encode static spatial relations. The previous chapters have stressed the dynamic, changing nature of our perceptual experience. Moreover, we have emphasized that comprehension, whether through words or visual images, is best thought of as the grasping of dynamic abstract (often causal) relations among objects. Dynamic relations are important because the meaning of an object can often only be understood by observing its effects on other objects -- the role it plays in an event schema. Bransford and McCarrell (1974) have been forceful advocates of this position.

Clearly, the kinds of simple, static displays used by Clark *et al*, do not easily lend themselves to the investigation of how language or pictorial media afford the comprehension of dynamic causal events. For this reason, they are a poor foundation upon which to build a general theory of how language and perception are related.[1]

In light of these arguments, one of the contributions of the present experiment has been to extend the scope of theories of sentence-picture comparison to include dynamic events. The consequence of using a richer set of stimuli such as cartoon-strips is to encourage the search for a richer set of theoretical constructs to handle them. Thus, elementary propositions of the form "A (above) B" are inadequate to represent dynamic events such as "colliding" or "bouncing." New relational concepts such as "cause," "change" and "possesion," must be the building blocks of _event_ representations. Current research on the representation of meaning mentioned in Chapter 1 should prove useful in this regard.

A second problem with accepting Clark _et al_'s model as a general theory of sentence-picture comparison concerns the use of the

---

[1]One might argue that since painting, drawing, photography and sculpture are "immobile" media, a useful starting point for a theory of picture comprehension might be to study the encoding of the types of static spatial relations depicted by Clark's stimuli. Even this may not be a good idea, however. Arnheim (1964) has pointed out that successful portrayal in photography or painting often demands knowledge of "visual dynamics." He illustrates this point with the example of a sequence of still photographs representing various phases of a blacksmith at work. Only those pictures showing the hammer lifted high convey the full impact of the blow; the intermediate phases are not perceived as transitional stages but rather as a quiet lifting of the hammer. Presumably, knowledge of dynamic relations underlying events affects, not only the artist's ability to select the best "frozen moment" to portray, but also affects the observer's ability to comprehend the portrayed event.

verification paradigm. As with the present experiment, verification requires a comparison of the internal representations of verbal and visual inputs. There is a critical difference between the two tasks, however. The task of verification channels the comparison process into an "all-or-none" mode of decision-making. If there is an exact match between the components of the representations, a "true" response is generated. If a mismatch occurs at early stages, extra operations are required which involve changing "truth indices" until the sentence is determined to be either true or false of the picture. The problem with verification is that by limiting the response to either true or false, it precludes any opportunity to observe the role that imaginative processes might play in how people compare linguistic and perceptual events.

In contrast to the verification paradigm, subjects' responses in the present experiment are not limited to true or false. Instead, the focus is on how to transform an initial mismatch into a match. "True" and "false" are not meaningful responses. Mismatches at the early stages of the comparison process do not result in the changing of "truth indices," but rather activate the search for transformations (e.g., change of viewpoint) which narrow the semantic distance between the event representations.

Shifting the emphasis from verification to the study of how people discover analogies between linguistic and perceptual deep structures has the advantage of highlighting the creative, problem-solving nature of the comparison process; i.e., where people must use their abstract knowledge of events and transformations in finding resemblances.

Encoding of Ambiguous and "Improverished" Pictures. Psychologists have long been fascinated by the problem of ambiguous pictures; that is, where the same visual pattern gives rise to multiple percepts. Some well-known examples include Wittgenstein's "rabbit-duck" drawing in which the rabbit's ears become a duck's bill when one switches attention from one animal to the other. An even more striking example is the Necker cube where the same line of a schematic drawing of a cube can be either a front edge or a back edge depending on which of two possible three-dimensional interpretations the viewer assigns to the two-dimensional display. Ambiguous figures are often cited by perception theorists as evidence of the interpretive, inferential nature of perception. They demonstrate that a single visual input may be accounted for by any number of candidate descriptions.

For example, Bugelski and Alampay (1961) showed that the description assigned to an ambiguous drawing (rat/man) could be biased toward either meaning by using the technique of set. A subject could be induced to see the percept "rat" by showing him a preceding series of pictures of animals. Similarly, the present experiment demonstrated that a complex visual input, a cartoon event, could be accounted for by two different event descriptions depending on the induction of a set for one of two assumed viewing angles. The present study differs from the problem of ambiguous figures, however, in two important respects. First, the cartoon-strips are not ambiguous in the sense of spontaneously oscillating back-and-forth from one viewpoint to another, as the Necker cube can for example. Rather, the meanings assigned to the drawings are usually consistent with a single, side-view description. Second, the dual meaning only becomes apparent as the result of a complex interaction between underlying descriptions; i.e., a kind of problem-solving where the subject attempts to match the visual input to the verbal description.

By using language as a means of guiding the processing of the cartoons, the present experiment shares some similarities with experiments demonstrating the effect of language on the comprehension of pictures. One of the earliest was that of Carmichael, Hogan and Walter (1932) who demonstrated that language can bias peoples' reproductions of tachistoscopically presented drawings of objects. By simply announcing the name of one of a pair of objects (glasses versus dumbell) before the presentation of the picture, the experimenter could bias subjects' drawings toward one of the two items.

More recently, Bower, Karlin and Dueck (1975) tested peoples' memory for "droodles." Droodles are impovershed pictures which present unlikely fragments of hidden objects which are difficult to interpret without some clue. Examples of droodles plus the verbal clues used by Bower, et al are shown in Figure 3.5. Subjects receiving the verbal clues during interpretation of the pictures performed better on recall and recognition tasks than did control subjects who did not see the accompanying verbal descriptions.

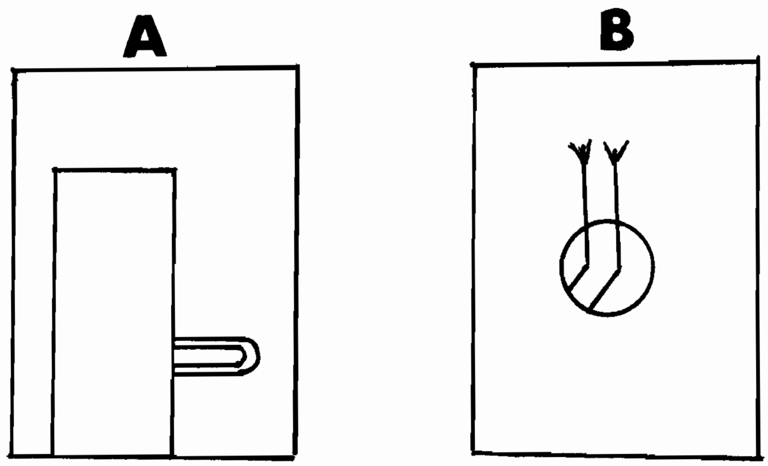


Figure 3.5. - Droodles: (a) A midget playing a trombone in a telephone booth; (b) An early bird who caught a very strong worm.

Source: Bower, Karlin and Dueck (1975)

The authors point out that verbal clues provide the subject with a familiar "schema" or "frame" which guides the assimilation of the visual features to meaningful categories. Although the authors offer no theory as to how this assimilation process works, the present experiment investigates one rule that might have been part of the assimilation process: e.g., assumed viewing angle. Both verbal descriptions are effective schema for the droodles because they suggest to subjects a way that the observer's own viewing angle could account for the partial fragments present in the drawings. For example, changing one's viewing angle from the side to directly in front of the phone booth would expose the currently hidden parts of the dwarf and his trombone. Thus, transformations may have played a role in the interpretation of droodles.

The present study goes beyond the droodles experiment by showing, not only that language can provide a familiar schema for the interpretation of visual stimuli, but also that the process by which the interpretation is achieved can itself become a schema. That is, the experiment demonstrated that language does not just furnish meaningful categories into which features are fit, but also can trigger a method of fitting -- a transformational schema -- which can apply to new cases.

SUMMARY AND CONCLUSION

One aspect of cartoon analogies that deserves mention in this closing section is that people often find them amusing.[1] It tickles us to see that a bunch of interacting squares and triangles can be classified as crushing or wiping. Why should this be so? The act of comprehending a phrase such as "wiping a dusty blackboard" is not usually considered to be amusing. Why should it be fun to find such a meaning for a cartoon?

Part of the reason may be that these cartoons do not yield their meanings without a bit of a struggle. A resemblance may not be perceived immediately and as we seek to make sense out of the interacting elements, a slight tension builds. When suddenly the pieces of the puzzle fall into place, we feel a kind of release from tension as we manage to fit a cartoon to a familiar schema. Such feelings of release from tension are pleasureable and often accompany the final stages of solving a problem or even getting the point of a bad joke.

Another, perhaps deeper, reason why we find these cartoons amusing is that they force us to see how flexible our cognitive frameworks can be. We are amused when we realize that a set of objects belonging to one domain -- geometric shapes -- can be

---

[1] I am grateful to Adam Gopnik for calling to my attention the problem of why someone should find an abstract cartoon humorous or amusing and for suggesting a possible explanation.

used as a pathway to a different set of target realities -- the class of everyday events. The cartoons force us to recognize a kind of tension between code and reality; we know that a triangle is not an animate agent, yet despite this awareness we cannot resist endowing the triangle with agency. Thus, the cartoons not only induce us to perceive resemblances, they can also make us aware that a clever manipulation of a symbolic medium has occurred; a manipulation which brings us face-to-face with the non-rigid nature of our classification system. This same awareness of the flexibility of our cognitive frameworks is partly what underlies our enjoyment of a Norman McLaren film when a dot or a line begins to "behave" in human-like ways.

Gombrich (1969) reminds us that a particularly clever manipulator of the duality between code and reality is the cartoonist Saul Steinberg. Consider the drawing in Figure 4.0. Through the use of context, Steinberg is able to employ a single pictorial element, the straight line, to represent many kinds of realities: a water-line, a wash-line, a horizon-line, a train track, and so on. Steinberg's cartoon is also a good reminder that the boundary between the "metaphorical" and so called "normal" modes of pictorial comprehension is blurred. The cartoon shows that the tension between code and reality is always potentially present in picture perception, although our awareness of it can vary. Even understanding a simple line-drawing of a cube requires the ability to interpret entities

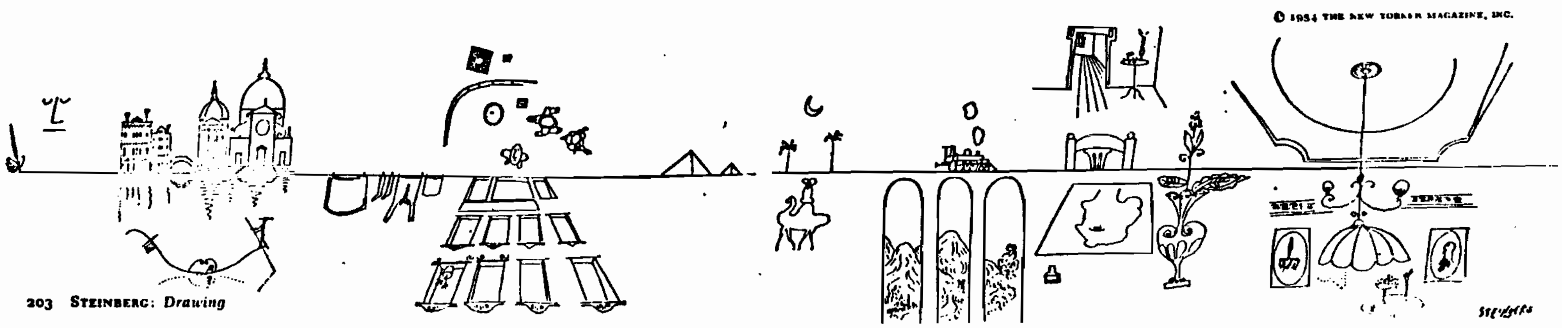


Figure 4.0. - A Steinberg cartoon

Source: The New Yorker Magazine (1954)

from one domain -- the class of lines -- in terms of the domain of solid, three-dimensional objects. And, as recent work in scene analysis shows, the ability to map such picture fragments onto the domain of three-dimensional objects requires a complex conceptual structure which relates the representational domain (line fragments) to the semantic domain (solid, 3-D objects).

Of course the picture fragments in cartoon-strip analogies are not just straight lines (except for the contours of the frames), but also two-dimensional geometric shapes. The challenge of Chapter 1 was to examine the process by which the conceptual system is able to map the geometric picture fragments, composed within the medium of cartoon-strips, onto the domain of familiar actions and events. Using peoples' interpretations of a number on abstract cartoon-strips, we asked what kind of information-processing system could have performed the input-output transformations? We quickly discovered that the commonly held view that people recognize resemblances between events because they belong to the same "equivalence class" (i.e., share a set of physical features) could not account for the ability to understand cartoon-strip analogies. It was shown that the physical features of the individual visible elements were not as critical to comprehension as abstract relations among the elements. But more than this, subjects' interpretations showed that many concepts -- such as assumed viewing angle -- involved in shaping the meanings have no counterpart either in the elements of the drawings or the actual words of the interpretation.

It was suggested that, instead of classification, a better way to characterize cartoon comprehension was as a kind of problem-solving in which the conceptual system forms a description which explains or accounts for the visible elements in the cartoon display by a finite number of concepts or "ideals." The concepts that form descriptions were called ideals because, although they do not usually arrive at the sense organs as simple, separate pieces of the visual input, they are the conceptual system's assumption that separate, simple factors may actually be present in a scene in complex ways. That is, it was shown now that the conceptual system must know how ideals can interact -- undergo transformations -- in order to form accurate descriptions of a visual input.

The notion that comprehending cartoon-strip analogies is a process of description encouraged us to describe some of the different kinds of ideals that seemed to be shaping the interpretations. In so doing, we proposed that the description process occurs in hierarchical stages, each stage generating a description of the input in terms of certain kinds of ideals and passing along its results to the next highest stage of analysis until the input is recognized as a familiar event. The early stages of description include such basic concepts as: "object," "movement," "overlap," "next-to," etc., which serve to describe the spatial relations among the elements in the different frames as well as to account for changes in shape, position or orientation of the separate elements.

At higher levels of description, we postulated the existence of more complex ideals such as "causality" which describe the changes in one element as a consequence of changes in another element. At even higher level, we postulated the presence of event schemata (e.g., "wiping") which encode our prototypical knowledge of events in terms of a number of abstract categories: agent, recipient, instrument and so on.

Although portraying the description process as occurring in hierarchical stages helped us to see the various kinds of ideals involved in cartoon comprehension, we pointed out that not only did the process work "bottom-up," but also "top-down." That is, it is not the visual features alone which drive the description process, but the conceptual system itself may generate a hypothesis, based on partial evidence, as to the kind of event being portrayed by the cartoon. In this case, it will try to test its hypothesis against the visual features of the drawing. This testing procedure may reveal that certain features are inconsistent with the hypothesis. When this happens, we proposed that the conceptual system has certain strategies for dealing with inconsistencies; i.e., it may modify the description of the cartoon by performing certain kinds of transformations on it: for example, imagining the event from a different viewpoint or seeing the event depicted by the cartoon as the consequence of an earlier unseen event.

Acknowledging that the description process can work top-down and that fitting the visible elements in a cartoon to an event schema can call upon transformational strategies helped us better understand two main problems raised by the interpretations: (1) how the same cartoon can elicit different interpretations and (2) how some subjects are able to treat the cartoons as metaphors for "abstract" ideas.

Perhaps the most difficult question raised in Chapter 1 was how the linguistic and perceptual systems are related in the comprehension of the cartoons. One theory found wanting, is that picture-like mental images could serve as mediators linking words to the cartoons. The problem with this theory was not the idea that words can become associated with mental images; but the claim that mental images are faded, percept-like entities bearing isomorphic resemblance to their external referents. The ability to label cartoon-strip analogies as complex events, not just geometric shapes, presents a serious challenge to the mediated associates model. It is not clear how a picture-like mental image could embody knowledge of abstract relations underlying events or procedures necessary for identifying them in abstract cartoon strips.

An alternative model was that both language and perception are linked to a common conceptual system in which the meanings of visual and linguistic inputs are mapped onto a common representational format. We ended the chapter by citing two examples of how semantic networks have been used recently to represent the meaning of linguistic utterances and visual events. For language, an utterance is decomposed into its basic underlying propositions; these propositions are themselves formed out of primitive semantic components (change, cause, possession, etc.). For vision, a computer program which simulates the mental processes involved in the perception of causal event sequences was cited as an example of how structural networks could be used to represent the meaning of dynamic visual events. A weakness of both approaches is that they do not provide procedures which could use transformations, based on the perceiver's tacit knowledge of the perceptual effects of his own activity, in matching the meanings of linguistic and perceptual inputs.

In Chapter 2, the ideas of description and transformation were extended to the problem of how people compute the degree of resemblance between an abstract cartoon strip and several alternative verbal interpretations. An experiment was conducted to demonstrate (1) that people are able to compute the degree of resemblance between events when they are encoded in two different media -- phrases and abstract cartoons and (2) that they can do this despite the fact that there are no surface similarities between the physical features of objects in the cartoon display and the physical features of the objects mentioned in the phrases. The results showed that people agree strongly as to which of a set of phrases is best, next-best and worst fit to an abstract cartoon. The ability to perform such "fuzzy" matches was taken as further evidence that our knowledge of event classes is flexible and not restricted to a set of criterial features.

In exploring the cognitive competence underlying peoples' intuitions about goodness-of-fit between cartoons and phrases, we looked for help to a computer simulation of the processes involved in another kind of analogy task -- the solving of geometric analogies -- where people are asked to find structural similarities between sets of geometric patterns. These problems are of the form: A is to B as C is to which of a set of alternative patterns?

We argued that the program which solves these problems embodies some general principles of analogical reasoning which can apply as well to the problem of computing the goodness-of-fit across modes. The success of the program depends on its forming descriptions of the patterns being compared, describing the differences, and accounting for the differences by transformations. Once the appropriate transformations are discovered which change A to B, they are applied to C to give the "ideal" solution. Similarly, evaluating the goodness-of-fit between a cartoon and a set of alternative phrases requires generating underlying descriptions of the events, comparing the descriptions, assessing the differences and choosing the phrase coming closest to the ideal description of the cartoon; i.e., requiring the fewest transformations to find a fit. As with geometric analogies an important strategy is the ability to transform an initially overspecific description to deeper levels of abstraction; for example, ignoring a mismatch in terms of specific classes of object (squares versus bread) while basing the resemblance on an abstract component alone (e.g. transfer).

One of the main themes of the thesis has been that the theoretical constructs of description, ideal and transformation, which have already proved useful in the study of perception, thinking and problem-solving, can be fruitfully applied to the study of interactions between language and perception.

To this end, Chapter 3 presented an experiment where people were asked to use their imagination to find analogies between phrases and abstract cartoon-strips. The experiment was designed in such a way that the ability to find the analogy would be affected by the use of a transformation involving assumed viewing angle. Our claim was that if such transformations are psychologically real, independent components of the problem-solving process, then one should be able to manipulate them experimentally.

To do this, we used the technique of "set." In each of nine trials subjects were asked to find an analogy between a cartoon and a phrase while the amount of time it took to find the analogy was recorded. In one condition, the cartoon more easily fit the meaning of the phrase if imagined from a side-view. In the other condition, the analogy to the phrase could be most easily apprehended by imagining the cartoon event from a top-view. By having subjects receive the same type of viewpoint over nine trials we hoped to induce in them a perceptual set for expected viewpoint. Such an effect was observed on a final "shift" trial where subjects encountered a phrase from the other viewpoint condition.

In addition to providing evidence for the psychological reality of viewpoint transformations in finding resemblances across modes, the experiment also suggested that interpreting events through linguistic and pictorial media involve assumptions about prototypical viewpoints; namely, an event described in language or portrayed by a cartoon drawing will be imagined as representing a side or front view unless contextual information specifies otherwise. It was hypothesized that the tendency to assign side-views reflects the fact that we perceive events most often from an upright spatial orientation rather than from bird's-eye or worm's-eye views.

Perhaps the main contribution of experiment 3 has been to contribute to the widening of the scope of current research on the relations between language and perception. Research on this problem has been hampered by an overly restrictive task, that of sentence-picture verification. The present experiment has presented an alternative to the verification paradigm in two ways: first by using a richer type of verbal and visual stimuli which incorporate important dynamic, causal relations and secondly, by showing that the human conceptual system can deal with mismatches between verbal descriptions and visual events in more interesting ways than changing "truth indices" if imaginative processes are allowed to operate.

Some Questions for Further Research. This thesis has been concerned with how people find resemblances across pictorial and linguistic modes. We have tried to show that this task can be studied as a kind of problem-solving in which cognitive structures (ideals, schemata, frames) interact, sometimes in complex ways, during the attempt to understand an abstract cartoon-strip or to justify a resemblance between a cartoon and a verbal description. The experiments reported in the thesis were designed to challenge the human conceptual system so that we might see more clearly the effects of schemata, strategies and assumptions in the interpretation process. The people participating in these experiments were all adults. One of the premises of the experiments then was that the subjects performing the tasks had fully-developed or mature conceptual systems. That is, while the experiments could challenge the conceptual system in particular ways, we assumed that the basic cognitive building blocks employed to solve the various problems posed by the experiments had already been acquired; i.e., those concepts necessary to understand the conventions of comic strips, to read and understand phrases, to think at a certain level of abstraction, etc.

Now many years of research in cognitive development suggest that the human conceptual apparatus does not arrive full-blown in the young child. Instead, Piaget and others have proposed that the child's conceptual system progresses through a sequence of stages such that his schemata or concepts become increasingly powerful: that is, the child acquires more sophisticated formulas for generating and manipulating his internal representations of events, objects and ideas. Roughly, the formulas at later stages of cognitive development, although derived from the earlier, more basic formulas, permit the child to progress from thinking only in terms of concrete objects, properties and situations to thinking about the world in more general, abstract ways. For instance, at later stages of cognitive development the child can think about his own thought processes deriving new rules or strategies for guiding his thinking about new problems as well as old ones.

Given this view, it strikes me that an interesting line of future research would be to examine the problem of analogical thinking across symbolic modes in a developmental context. A developmental approach might be worthwhile, not only because it could increase our understanding of the growth of imaginative processes but also because it could teach us something about the cognitive competence necessary to understand and use symbolic forms other than language: i.e. the comprehension of events in the pictorial medium of cartoon-strips. Let me conclude therefore by outlining a number of questions for further research on analogical thinking across symbolic modes from a developmental perspective.

In an earlier section of this thesis we suggested an alternative interpretation of Piaget's experiment on children's judgements of animacy -- i.e., where they were asked to classify things as "living" versus "nonliving." We argued that the apparent willingness of the child to classify a cloud as "alive" might not reflect so much a cognitive deficit, but rather a willingness to engage in analogical thinking: to seek partial and to some extent abstract resemblances between living and non-living movement patterns.

In fact, recent research supports the view that the ability to engage in a kind of metaphorical perception ("symbolic play") begins very early, around the second year of life (Gardner, Winner, Bechhofer and Wolf, 1977); for example, a child using a pencil to imitate the flight of a rocket. Now the ability to understand abstract cartoon-strips as familiar events also demands what Perkins (1978) has called the ability to engage in a kind of "contrary seeing;" i.e., finding a likeness between two events despite perceptual evidence to the contrary. An example would be seeing a face in the clouds or, as in the case of abstract cartoon-strips, seeing a visual pattern as representing a specific event class (e.g. "wiping") despite evidence that the pattern consists of two-dimensional geometric shapes. It would be interesting to know if this propensity for metaphorical perception in children transfers easily to the problem of interpreting abstract cartoon-strips such as those used in the present thesis. If young children can identify abstract cartoon-strips as representing specific event classes, this would suggest that our event concepts are abstract and flexible even at very early stages of cognitive development.

Of course cartoon-strips can be designed so that some require more sophisticated schemata than others in order to be understood. It would be interesting to make use of this fact by studying the ability of children at different levels of cognitive development to interpret cartoons which pose certain kinds of problems. For example, cartoon-strips can be designed such that certain kinds of information are ambiguous, missing or even anomalous: e.g., Cartoon-1 in Appendix B where a change-of-state is represented with an ambiguous cause. How might children at different levels of cognitive development handle such ambiguities? Would the younger child feel any obligation to identify a causal agent? Have they already acquired the notion of a "well-formed" event? How might they deal with anomalies, say a piece of an event in a cartoon that does not fit in with the rest of the sequence?

In addition to examining developmental differences in how children tackle different kinds of problems posed by cartoon-strips, one might begin to explore the child's ability to achieve correspondences across symbolic modes. In Chapter 2, for example, we saw that adults had little difficulty in computing the goodness-of-fit among cartoons and verbal descriptions. In doing this task, we argued that the conceptual system restructures its initial representation of the meaning of a cartoon or a phrase. One type of restructuring involved the generalizing of an overly specific description in order to uncover a deeper level at which a fit was possible. Performing such a task demonstrates a kind of "meta-knowledge" of one's own conceptual system. At what developmental level do children begin to show such meta-knowledge in seeking deeper resemblances between events encoded in two different modes? That is, could they make goodness-of-fit rankings between verbal descriptions and picture sequences which required evaluating the closeness of analogy? Could they verbalize the reasons why one phrase was chosen as a better fit than another? How would children at different levels of development deal with metaphorical interpretations -- seeing a cartoon as a visual metaphor for a concept such as "domination?" This too would require the ability to process a description to deeper levels of abstraction. Another way to probe the child's evolving conceptual system would be to ask him to explain how the same cartoon could fit alternate interpretations.

Finally, we saw in Chapter 3 that adults are able to draw on their knowledge of how events look at different viewing angles in dealing with mismatches between cartoons and phrases. Piaget has argued that at certain levels of cognitive development young children have great difficulty in imagining how a scene will look from a point of view other than their own. Piaget calls this "ego-centrism." Would ego-centrism reflect itself in children's ability to imagine how a cartoon-strip could fit either of two event classes depending on a change in viewing angle? Are some vantage points easier to imagine than others?

These kinds of questions seem to me to be a promising area for future research. Such an approach, based on the use of abstract cartoon-strips, would begin to unite studies of problem-solving, analogical thinking and imaginative processes with research on the child's emerging conceptual apparatus and how if affects his ability to communicate verbally about what he sees.

BIBLIOGRAPHY

Arnheim, R. _Art and Visual Perception_. Berkeley and Los Angeles: University of California Press, 1964.

__________. _Visual Thinking_. Berkeley: University of California Press, 1969.

Bartlett, F.C. _Remembering_. Cambridge: Cambridge University Press, 1932.

Bickerton, D. "Prolegomena to a Linguistic Theory of Metaphor." _Foundations of Language_, 1969, _5_, 34-52.

Black, M. _Models and Metaphors_. Ithaca: Cornell University Press, 1962.

Bobrow, D.G., & Norman, D.A. "Some Principles of Memory Schemata." In D.G. Bobrow & A.M. Collins (Eds.), _Representation and Understanding: Studies in Cognitive Science_. New York: Academic Press, 1975.

Boden, M.A. _Artificial Intelligence and Natural Man_. New York: Basic Books, 1977.

Bower, G.H. Karlin, M.B., & Dueck, A. "Comprehension and Memory for Pictures." _Memory & Cognition_, 1975, Vol. 3(2), 216-220.

Bransford, J.D., & Johnson, M.R. "Considerations of some Problems of Comprehension." In W.G. Chase (Ed.), _Visual Information Processing_. New York: Academic Press, 1973.

Bransford, J.D., & McCarrell, N.S. "A Sketch of a Cognitive Approach to Comprehension." In W. Weimer & D. Palermo (Eds.), _Cognition and the Symbolic Processes_. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1974.

Bregman, A.S. "Perception and Behavior as Compositions of Ideals." _Cognitive Psychology_, 1977, _9_, 250-292.

Brown, S.J.M. _The World of Imagery: Metaphor and Kindred Imagery (2nd ed.)_. New York: Russell Y Russell, 1966.

Bugelski, B.R., & Alampay, D.A. "The Role of Frequency in Developing Perceptual Sets." _Canadian Journal of Psychology_, 1961, _15_, 205-211.

Carey, P.W., Mehler, J., & Bever, T.G. "Judging the Veracity of Ambiguous Sentences." Journal of Verbal Learning and Verbal Behavior, 1970, 9, 243-254.

Carmichael, L., Hogan, H.P., & Walter, A.A. "An Experimental Study of the Effect of Language on the Reproduction of Visual Perceived Form." Journal of Experimental Psychology, 1932, 15, 73-86.

Chomsky, N. Aspects of the Theory of Syntax. Cambridge: MIT Press, 1965.

Clark, H.H., & Chase, W.G. "On the Process of Comparing Sentences against Pictures." Cognitive Psychology, 1972, 3, 572-517.

Clark, H.H., Carpenter, P.A., & Just, M.A. "On the Meeting of Semantics and Perception." In W.G. Chase (Ed.), Visual Information Processing. New York: Academic Press, 1973.

Collins, A.M., & Quillian, M.R. "Experiments on Semantic Memory and Language Comprehension." In L.W. Gregg (Eds.), Cognition in Learning and Memory. New York: Wiley, 1972.

Edwards, A.L. Statistical Methods for the Behavioral Sciences. New York: Holt, Rinehart and Winston, 1954.

Evans, T.G. "A Program for the Solution of Geometric-Analogy Intelligence Test Questions." In M.L. Minsky (Ed.), Semantic Information Processing. Cambridge, Mass.: MIT Press, 1968.

Fillmore, C.J. "The Case for Case." In E.U. Bach & R.T. Harms (Eds.), Universals in Linguistic Theory. New York: Holt, Rinehart & Winston, 1968.

Furth, H. Piaget and Knowledge. Englewood Cliffs, New Jersey: Prentice-Hall, 1969.

Gardner, H. The Arts and Human Development: A Psychological Study of the Artistic Process. New York: John Wiley & Sons, 1973.

Gardner, H., Winner, E., Bechhofer, R., Wolf, D. "The Development of Figurative Language." In K. Nelson (Ed.), Children's Language, New York: Gardner Press, 1977.

Gibson, J.J. The Senses Considered as Perceptual Systems. Boston: Houton-Mifflin, 1966.

_________. "The Theory of Affordances." In R.E. Shaw & J.D. Bransford (Eds), Acting, Perceiving and Knowing: Toward an Ecological Psychology. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1977.

Gombrich, E.H. Art and Illusion. New York: Pantheon, 1960.

Gopnik, M. "Meaning may be a Catastrophe: A Solution for Jackendoff." Unpublished manuscript, Department of Linguistics, McGill University, 1976.

Guzman, A. "Computer Recognition of Three-Dimensional Objects in a Visual Scene." Artificial Intelligence Technical Report-228. Cambridge, Mass.: Massachussetts Institute of Technology Artifical Intelligence Laboratory, 1968.

Heider, F., & Simmel, M.L. "An Experimental Study of Apparent Behavior." American Journal of Psychology, 1944, 57, 243-259.

Hester, M.B. The Meaning of Poetic Metaphor: An Analysis in Light of Wittgenstein'slaim that Meaning is Use. The Hague: Mouton, 1967.

Hume, D. A Treatise on Human Nature. Vol. 1. Originally published in 1739. London: Dent, 1960.

Katz, J.J., & Fodor, J.A. "The Structure of a Semantic Theory." Language, 1963, 39, 170-210.

Lakoff, G. "Hedges. A Study in Meaning Criteria and the Logic of Fuzzy Concepts." Papers from the Eighth Regional Meeting, Chicago Linguistics Society, Chicago: Department of Linguistics, University of Chicago, 1972.

Langer, S. Philosophy in a New Key. Cambridge, Mass.: Harvard University Press, 1957.

Lindsay, P.H., & Norman, D.A. Human Information Processing (2nd ed.), New York: Academic Press, 1976.

Locke, J. An Essay Concerning Human Understanding. Vols. 1 and 2. Originally published in 1690, London: Dent, 1961.

Mehler, J., & Carey, P.W. "Role of Surface and Base Structure in the Perception of Sentences." Journal of Verbal Learning and Verbal Behavior, 1967, 6, 335-338.

Miller, G.A., & Johnson-Laird, P.N. Language and Perception. Harvard University Press, 1976.

Michotte, A. The Perception of Causality. London: Methuen, 1963.

Minsky, M. "A Framework for Representing Knowledge." In P.H. Winston (Ed.), The Psychology of Computer Vision. McGraw-Hill, 1975.

Neisser, U. Cognitive Psychology. New York: Appelton-Century-Crofts, 1967.

__________. Cognition and Reality. San Francisco: W.H. Freeman, 1976.

Newell, A. & Simon, H.A. Human Problem Solving. Englewood-Cliffs, New Jersey: Prentice-Hall, 1972.

Norman, D.A., & Rumelhart, D.E. Explorations in Cognition. San Francisco: W.H. Freeman & Sons, 1975.

Norman, D.A., & Bobrow, D.G. "On the Role of Active Memory Processes in Perception and Cognition." In C.N. Cofer (Ed.), The Structure of Human Memory. San Francisco, W.H. Freeman, 1976.

__________. "Descriptions: An Intermediate Stage in Memory Retrieval." Cognitive Psychology, 1979, 11, 107-123.

Osborn, M.M., & Ehninger, D. "The Metaphor in Public Address." Speech Monographs, 1962, 29, 223-234.

Osgood, C.E., Suci, G., & Tannenbaum, P. The Measurement of Meaning. Urbana: University of Illinois Press, 1957.

Paivio, A. Imagery and Verbal Process. New York: Holt, Rinehart and Winston, 1971.

Perkins, D.N. "Metaphorical Perception." In E. Eisner (Ed.), Reading, the Arts and the Creation of Meaning. Reston, Virginia: National Art Education Association, 1978.

Piaget, J. The Child's Conception of the World. New York: Harcourt Brace, 1929.

_________. La psychologie de l'intelligence. (1947). Translated as The Psychology of Intelligence. London: Routledge and Kegan Paul, 1950.

Pollio, H.R., Barlow, J.M., Fine, H.J., & Pollio, M.R. Psychology and the Poetics of Growth. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1977.

Pylyshyn, Z.W. "What the Mind's Eye Tells the Mind's Brain: A Critique of Mental Imagery." Psychological Bulletin, 1973, 80, 1-24.

Richards, I.A. The Philosophy of Rhetoric. New York: Oxford University Press, 1936.

Rock, I. An Introduction to Perception. New York: MacMillan, 1975.

Rosch, E. "Cognitive Reference Points." Cognitive Psychology, 1975, 7, 532-547.

Schank, R.C. "Identification of Conceptualizations Underlying Natural Language." In R.C. Schank & K.W. Colby (Eds.). Computer Models of Thought and Language. San Francisco: W.H. Freeman, 1973.

Selfridge, O.G. "Pandemonium: A Paradigm for Learning." In Symposium of the Mechanisation of Thought Processes. London: H.M. Stationery Office, 1959.

Thomas, O. Metaphor and Related Subjects. New York: Random House, 1969.

Turbayne, C.M. The Myth of Metaphor. Columbia, S.C.: University of South Carolina Press, 1971.

Verbrugge, R.R. The Comprehension of Analogy. Unpublished doctoral dissertation, University of Minnesota, 1974.

_________. "Resemblances in Language and Perception." In R.E. Shaw & J.D. Bransford (Eds.). Perceiving, Acting and Knowing: Toward an Ecological Psychology. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1977.

Verbrugge, R.R., & McCarrell, N.S. "Metaphoric Comprehension: Studies in Reminding and Resembling." Cognitive Psychology, 1977, 9, 494-533.

Waltz, D.L. "Understanding Line Drawings of Scenes with Shadows." In P.H. Winston (Ed.). _The Psychology of Computer Vision_. McGraw-Hill, 1975.

Weir, S. "The Perception of Motion: Michotte revisited." _Perception_, 1978, _7_, 247-260.

Wheelwright, P. _Metaphor and Reality_. Bloomington, Ind.: University of Indiana Press, 1962.

Whorf, B.L. _Language, Thought and Reality: Selected Writings of Benjamin Lee Whorf_. J.B. Carroll (Ed.). New York: Wiley, 1956.

Winograd, T. "Frame Representations and the Declarative/Procedural Contreversy." In D.G. Bobrow & A.M. Collins (Eds.). _Representation and Understanding: Studies in Cognitive Science_. New York: Academic Press, 1975.

Winston, P.H. "Learning Structural Descriptions from Examples." In P.H. Winston (Ed.). _The Psychology of Computer Vision._ McGraw-Hill, 1975.

__________. _Artificial Intelligence._ Reading, Mass.: Addison-Wesley, 1977.

APPENDIX A

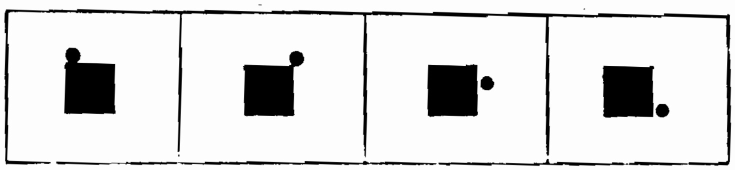
Instructions

INSTRUCTIONS

Experiment 1

This is an experiment on how people understand analogies. In an analogy two things are compared to see how they could be alike in some way. Usually analogies compare pairs of words, sentences or pictures to see what they have in common. In this experiment, however, you are going to be given only one part of the analogy in the form of a cartoon-strip. Your job will be to see how this cartoon could be an analogy for a familiar everyday event.

For example, you are going to see a series of cartoons like:

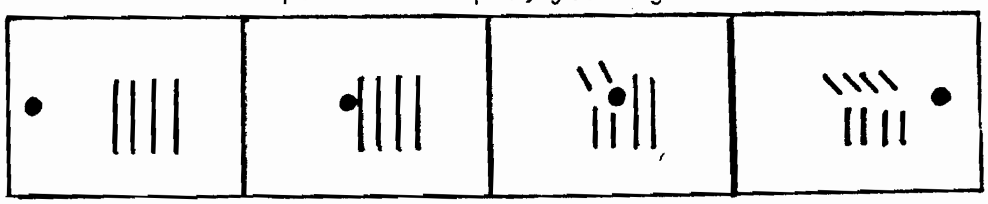


What I would like you to do is to look at each cartoon and try to understand how the kinds of changes happening to the figures in the drawing remind you of a familiar event. The cartoons are to be read like an ordinary comic strip from left-to-right. The cartoon depicts four frames of a single, continuous event -- not four different events. You might say, for example, that this cartoon reminds you of an egg falling off a table.

Once you have thought of an analogy for the cartoon, you will write it down on the sheet of paper provided starting with number 1. As you can see there is enough space beside each number to write down more than one analogy if you wish for any particular slide. Each slide will be projected for about three minutes before the next slide comes on.

Experiment 2

This is an experiment on how people understand analogies. In an analogy two things are compared to see how they could be alike in some way. Usually analogies compare pairs of words, sentences or pictures to see what they have in common. In this experiment, however, you are going to be given one part of the analogy in the form of a cartoon-strip. For example, you might see a cartoon like:



projected on the screen. For each cartoon you will find a list of three phrases on the rating sheets you have been given. In this practice example the three phrases are:

☐ sharpening a pencil

☐ mowing a lawn

☐ swinging a club

What I would like you to do is to look at the cartoon and try to understand how the kinds of changes happening to the shapes remind you of a familiar, everyday event. The cartoons are to be read like an ordinary comic-strip from left-to-right and represent four frames of a single, continuous event -- not four different events.

Once you feel you understand the cartoon, read the phrases and think about the kind of event they describe. After reading each of the phrases, choose which phrase you think is the best, next-best and worst analogy for the cartoon. Please indicate your order of preference by writing a 1 (best), 2 (next-best) or 3 (worst) in the box at the left of each phrase. You must assign a rank to each phrase and no two phrases should have the same rank.

Let's work through a practice example. How could you rank the phrases for the cartoon on the screen? (work through example).

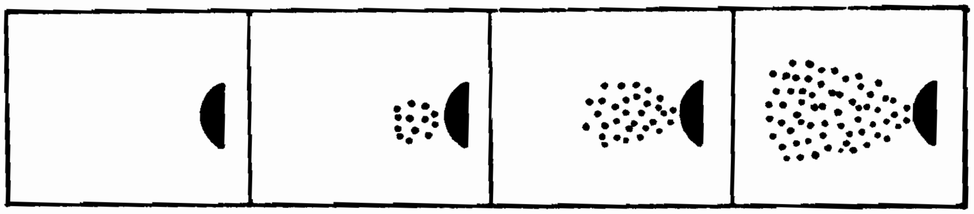
Experiment 3

This is an experiment on how people understand analogies. In an analogy two things are compared to see how they could be alike in some way. Usually, analogies compare pairs of sentences or pairs of pictures to see what they have in common. In this experiment, however, you are going to try to find analogies where the things being compared are in two different media: phrases and cartoons.

You are going to see a series of phrases describing various kinds of events. Following each phrase will be a cartoon strip. For example, you might see the phrase (show slide):

PUSHING THE BUTTON ON AN AEROSOL SPRAY

followed by the cartoon (show slide):



Your job will be first to read the phrase and think about the kind of event it describes. When you feel you understand the phrase, say "OK!" The phrase will disappear from the screen and a cartoon-strip will appear in its place. Look at the cartoon and try to imagine how the kinds of changes happening to the geometric shapes in the drawing could be an analogy for the kind of event described by the phrase you saw.

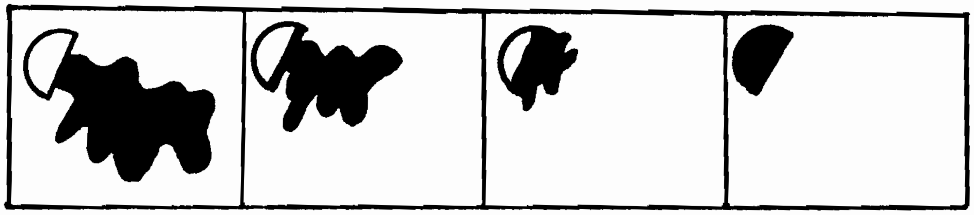
These cartoons are read from left to right and depict four frames of a single continuous event -- not four different events. In this slide, for example, you might see the cartoon as a kind of "spraying" in that the dots coming from the half-circle remind one of droplets emanating from a spray can.

What I'm interested is in how long it takes you to find an analogy between the phrase and the cartoon. That is, once the cartoon comes on I would like you to press this button (give S control switch in strongest hand) just as soon as you see how it could be an analogy for the phrase. Although there is no time limit, please react as quickly as possible once you see the analogy. Don't press the button, however, unless you do see a correspondence.

APPENDIX B

Interpretations for 10 Cartoon-Strip Analogies

CARTOON 1



Action[a]	Interpretation[b]
Absorbing; Vacuuming; Inhaling	1. a cloth absorbing ink
	2. an ink spot being sucked into a well
	3. vacuuming
	4. inhaling smoke
	5. a vacuum absorbing an ink spot
	6. a siringe absorbing liquid
	7. a vacuum sucking up something
	8. a rag wiping a stain
	9. a sponge absorbing liquid
	10. absorbing an oil stain
	11. a magnetic ray attracting a liquid
	12. an oil spot absorbed by a cloth
	13. absorbing information
	14. liquid filling a transparent container
	15. a mouth inhaling something
	16. someone inhaling smoke
	17. a sponge absorbing a spot
	18. a vacuum picking up dirt
Spilling; Emptying	19. emptying a glass of something
	20. spilling ink from an inkwell
	21. spilling coffee
	22. spilling the contents of a bowl
	23. spilling coffee
	24. spilling ink from an inkwell

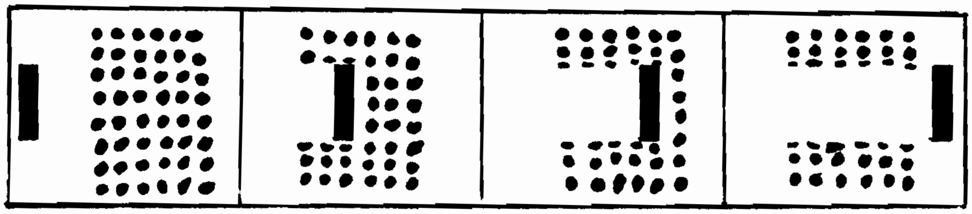
CARTOON 1 (continued)

Action[a]	Interpretation[b]
Filling	25. filling a container with liquid 26. filling a transparent container with liquid 27. filling up an inkwell
Other	28. a cloud covering the moon 29. smoke staining a lamp 30. a cloud approaching the sun 31. an animal hiding in a hole 32. putting out a light 33. a country disappearing from a map being swallowed up by... 34. brilliant sunshine 35. sticking out a tongue

[a]Interpretations describing similar kinds of action have been grouped together.

[b]The interpretations have been translated from the french.

CARTOON 2

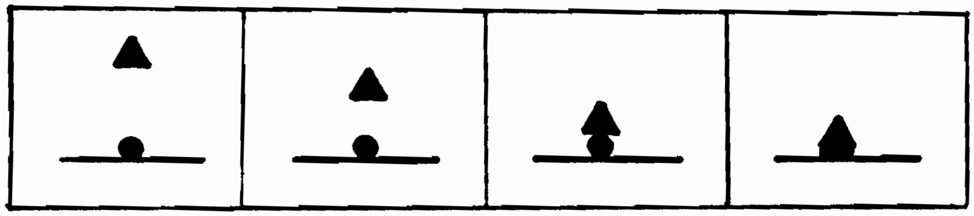


Action	Interpretation
Mowing; Plowing; Erasing; etc.	1. a lawnmower cutting grass
	2. a lawnmower cutting grass
	3. a lawnmower cutting grass
	4. a lawnmower cutting grass
	5. a lawnmower cutting grass
	6. mowing the lawn
	7. mowing the lawn
	8. mowing the lawn
	9. mowing the lawn
	10. mowing the lawn
	11. mowing the lawn
	12. mowing the lawn
	13. a bulldozer plowing snow
	14. a bulldozer clearing snow
	15. plowing a gap in a roadway
	16. a tractor removing snow
	17. a tractor plowing a field of cabbage
	18. a vehicle plowing through a crowd
	19. clearing snow with a shovel
	20. plowing some stones
	21. a bulldozer removing snow
	22. a shovel scooping up pebbles
	23. a bulldozer ravaging trees in a forest
	24. a plow clearing snow
	25. a tractor crushing rocks
	26. erasing a blackboard

CARTOON 2 (continued)

Action	Interpretation
	27. erasing a dirty blackboard 28. erasing a dusty blackboard 29. wiping up dust 30. a beam sweeping across a TV screen 31. a tank piercing an enemy's defenses 32. elite troops eliminating part of a population
Other	33. a plank caving in a wall 34. a game of dominos 35. two sidewalks and a car

CARTOON 3

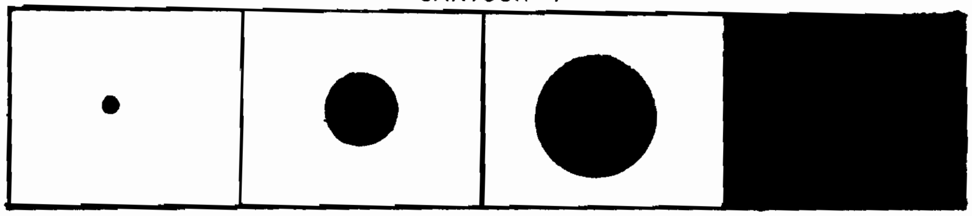


Action	Interpretation
Crushing; Mashing	1. a spoon crushing a potato
	2. one object falling on another and crushing it
	3. crushing a mushroom with a rock
	4. crushing a nut with hammer
	5. a hammer crushing a foot
	6. a mass flattening a sphere
	7. dropping a lampshade on a foot
	8. crushing a cigarette
	9. parachuting onto a rock and crushing it
	10. a foot crushing a ball
	11. a fork mashing potatoes
	12. crushing a ball
	13. a floating lampshade crushes a lamp
	14. a hand crushing an insect
Covering; Putting	15. covering a pea with a can
	16. a cover protecting food
	17. a hat covering a head
	18. putting a cover on a cake
	19. putting a roof on a house
	20. placing a roof on a house during construction
	21. covering a candle with a can
	22. putting out the flame of a cigarette with an unlit cigarette

CARTOON 3 (continued)

Action	Interpretation
Landing	23. an object landing
	24. a space capsule landing on the moon
	25. a capsule landing on a moon
	26. a spaceship landing on a planet
	27. landing a flying saucer
Other	28. eating soup with a spoon
	29. hammering a nail into wood
	30. a spider descending and gobbling up an insect
	31. a symbol of evolving domination
	32. two people joined together in marriage
	33. gathering up objects
	34. two lovers attracting each other
	35. a symbol of a youth hostel

CARTOON 4

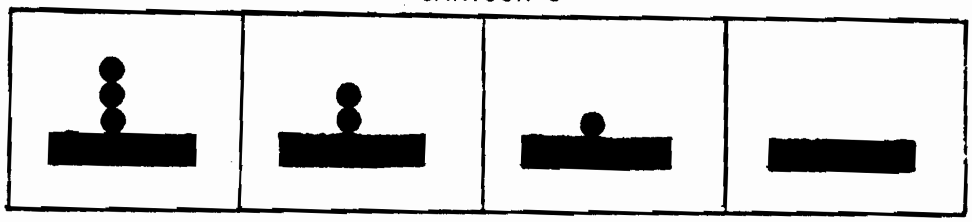


Action	Interpretation
Looking at an oncoming object	1. looking at an approaching ball
	2. getting a ball on the nose
	3. getting a rock thrown at you
	4. an object approaching me
	5. getting crushed by an oncoming car
	6. a ball coming toward a camera until it covers it
	7. seeing a ball drop from the sky
	8. a ball coming toward a camera lens
	9. watching a flying saucer come towards earth
	10. watching an object descend from the sky
	11. seeing a flying saucer approach from space
	12. a black bird coming at me
Approaching a stationary object	13. looking at a nail-head from different heights
	14. approaching the end of a dark tunnel
	15. watching a planet get bigger as I descend in a space ship
	16. walking toward the end of a tunnel
	17. seeing a planet get bigger as I come closer in space
	18. walking toward the opening in a tunnel
	19. watching a planet as I approach it

CARTOON 4 (continued)

Action	Interpretation
	20. zooming in on a dot with a microscope 21. seeing a Japanese flag at different distances
An object undergoing a size transformation (expanding)	22. an ink spot growing bigger 23. a spot filling a sheet of paper 24. an ink spot expanding on paper 25. painting a circle on a square 26. a car draining its oil 27. an expanding circle caused by a pebble dropped in water 28. a black spot fills a surface 29. a pebble forming a growing circle in water 30. an expanding spot filling a surface 31. a spot growing
Other	32. an eclipse of the sun 33. an eclipse 34. a traffic light 35. a nonresolvable problem you have to live with

CARTOON 5

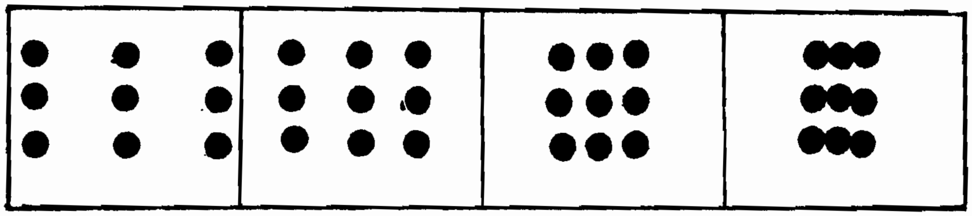


Action	Interpretation
Falling; Dropping; Pouring; Diving	1. a man diving into the sea 2. immersing a stick in water 3. drops of rain falling into a stream 4. three people falling off a boat into the sea 5. balls falling into a container 6. peas dropping onto a plate 7. water being poured into a container 8. dropping pears into a box 9. dropping coins in a piggy-bank 10. a submarine lowering its periscope 11. pearls falling down a hole
Entering; Passing through	12. people entering a hall 13. three persons entering an elevator 14. a caterpillar entering a mailbox 15. three people passing through a doorway 16. billiard balls entering a side pocket
Moving off toward the horizon	17. an individual moving off toward the horizon 18. someone entering the water and disappearing toward the horizon 19. someone moving across a plain toward the horizon
Melting	20. snowballs melting 21. a snowman melting

CARTOON 5 (continued)

Action	Interpretation
Hammering; Sticking	22. hammering a nail into a block
	23. sticking a pin into a pincushion
Removing	24. removing pieces of something until nothing remains
	25. removing balls one-by-one
Other	26. piling meatballs on a plate
	27. a rolling stone gather no moss
	28. assimilation of the minority by the masses
	29. the leveling of differences in society
	30. a toy for teaching counting
	31. olives disappearing as they get eaten
	32. a starting signal in a race
	33. a child retracting his tongue
	34. a man descending down into the earth

CARTOON 6

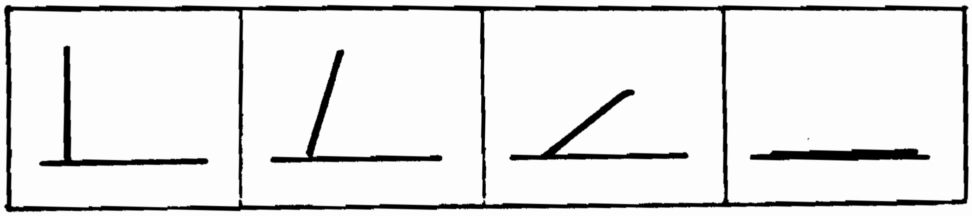


Action	Interpretation
Closing ranks; Converging	1. soldiers forming ranks 2. people coming together 3. nine people coming together to make friends 4. a battalion huddles together in fear 5. a group sandwiched between two others 6. majorettes coming together in a demonstration 7. soldiers executing manoeuvres 8. atoms converging to form a whole 9. people marching, coming together 10. children coming together in solidarity 11. nine people tightening ranks 12. military troops closing ranks 13. soldiers closing ranks after orders 14. soldiers tightening ranks 15. a square dance where people come together 16. a dispersed crowd coming together 17. people coming closer together during therapy 18. people sharing interests come together 19. people huddling together in the cold 20. flies converging on sugar 21. soldiers closing ranks 22. people coming together

CARTOON 6 (continued)

Action	Interpretation
	23. majorettes marching at half-time
	24. dancers coming together
	25. dancers coming together during square dance
	26. a group closing ranks in a parade
	27. sheep squeezing together to get through a narrow passage
Other	28. arranging cans of jam on a shelf
	29. walls closing in on prisoners in a cell
	30. seeing a leaf from head-on, then from the side
	31. a game of dominos
	32. an elastic taking on its initial shape
	33. union makes strength
	34. I like symmetry
	35. clarification of fuzzy ideas

CARTOON 7

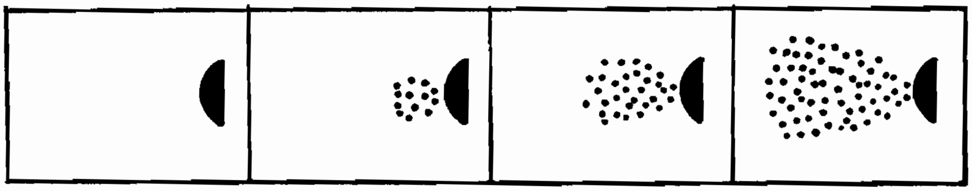


Action	Interpretation
Falling over	1. a pole falling over
	2. a thin pole falling
	3. a tree falling
	4. a ruler falling
	5. a pole falling to earth
	6. a wall falling over
	7. a stick losing its balance and falls
	8. a drunk falling down
	9. a stick falls to the ground
	10. a woman falling
	11. a telephone pole falling
	12. a pole falling to the ground
	13. chopping down a tree
	14. a tree falling
	15. a tree struck by lightening falls over
	16. a falling object
	17. a man falling
	18. a tree falling
	19. a plank of wood falling
	20. a wall falling
	21. a wall falling over
	22. wind pushing over a candle
	23. a person shaken, falls to the ground
	24. the leaning tower of Pizza
	25. it's hard to keep one's balance
	26. what goes up must come down

CARTOON 7 (continued)

Action	Interpretation
Closing; Folding	27. closing a book
	28. folding a camper
	29. a train signal closing
	30. a trap door closing
	31. closing a book
	32. seeing a door closing
	33. a notebook being closed
	34. opening a book
Other	35. time passing

CARTOON 8

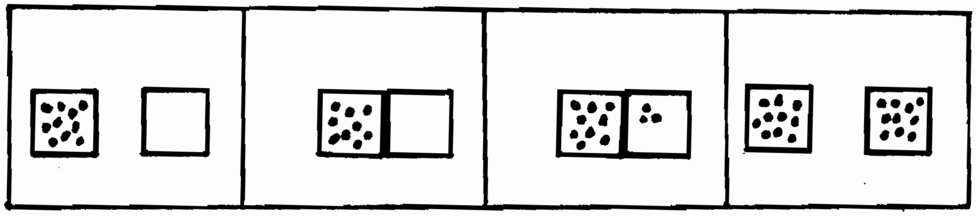


Action	Interpretation
Spraying; Blowing	1. water spraying from a shower
Emmitting	2. opening a shower faucet
	3. something blowing particles
	4. a flower emmitting pollen
	5. blowing, sneezing
	6. a shower begins to spray
	7. a hose spraying water
	8. eruption of a volcano
	9. tears flowing from someone crying
	10. opening a shower tap
	11. a boat leaving spray in its wake
	12. the diffusion of electrons
	13. the sun emmitting heat
	14. dandilion petals blowing away
	15. a volcano erupting
	16. soap bubbles detaching from a spoon
	17. an atomizer spraying something
	18. words emmitting their meanings
	19. someone with a cold, sneezing
	20. expiring air from one's lungs
	21. something emmitting words
	22. someone transmitting their voice
Leaving	23. bees leaving a hive
	24. "
	25. "
	26. "

CARTOON 8 (continued)

Action	Interpretation
	27. bees leaving a hive
	28. "
	29. "
	30. wasps leaving their nest
	31. "
	32. wasps flying from their nest
	33. wasps leaving their hive
	34. ants leaving their hill
	35. a crowd leaving a stadium

CARTOON 9

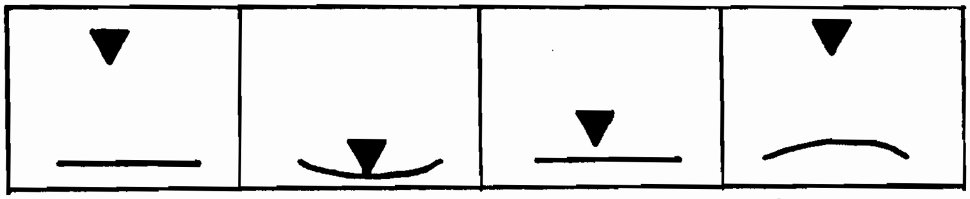


Action	Interpretation
Transferring	1. bacteria transferring from one tube to another
	2. two spacecraft exchanging personnel
	3. transfer of kinetic energy between two blocks of wood
	4. the idea of transfer
	5. the transfer of insects from one bottle to another
	6. the transfer of flies between two vials
	7. animals moving from one cage to another
	8. substances being transferred from one tube to another
	9. microbes transferring between test tubes
	10. transfer of sand in an hourglass
Infecting; Transmitting	11. sperm fertilizing an egg
	12. osmosis
	13. infecting someone with a virus
	14. osmosis between cells
	15. cellular division
	16. infecting someone with a virus
	17. the transmission of knowledge
	18. the transmission of microbes

CARTOON 9 (continued)

Action	Interpretation
	19. infecting someone with a virus
	20. reproduction by mitosis
	21. the act of mating and fertilization
	22. infecting someone with a disease
	23. the process of fertilization seen through a microscope
	24. infecting someone with a disease
	25. communicating ideas
	26. the idea of transmission
	27. the idea of transmission between two cells
	28. infection of molecules
Other	29. a game of dice
	30. homeostasis -- water rising to same level in two tubes
	31. two windows getting hit with raindrops, first one, then the other
	32. a game of dominos
	33. a game of dominos
	34. an exchange between extremes
	35. looking at a bunch of heads through a window

CARTOON 10



Action	Interpretation
Bouncing; Jumping	1. bouncing on a trampoline
	2. "
	3. "
	4. "
	5. "
	6. "
	7. "
	8. "
	9. "
	10. "
	11. "
	12. "
	13. "
	14. "
	15. "
	16. "
	17. "
	18. "
	19. "
	20. "
	21. "
	22. an object bouncing
	23. a diver bouncing on a springboard
	24. rebounding
	25. someone on a diving board
	26. a child bouncing on a bed
	27. a bouncing ball

CARTOON 10 (continued)

Action	Interpretation
Hitting; Ricocheting	28. hitting a metal bar with a hammer
	29. hitting a tennis ball with a racquet
	30. hitting a tennis ball
	31. a tennis ball being hit with a racquet
	32. a bullet ricocheting off a wall
Other	33. laughing then crying
	34. kicking a metal plate
	35. concavity then convexity

APPENDIX C

ANOVA Correction Procedure for the Latency Data from Experiment 3

ANOVA Correction Procedure for the Latency Data from Experiment 3

Because of the particular crossover design used in this study where a subject is switched to the contrasting viewpoint condition on the final trial, the normal way of partitioning out subject variance in a treatment-by-subjects ANOVA cannot be used since subjects are not actually nested in viewpoint conditions. However a correction procedure, applied prior to ANOVA had the desired result of partitioning out subject variance.

The procedure involved obtaining a correction factor for each subject based on the deviation of his mean score for trials 5 through 10 from the grand mean for these trials for subjects in his group. This factor was then added to each subject's score for all six trials. This removes subject variance from consideration. The remaining "subject variance" is really subject-by-trial interaction, the proper error term for the planned comparison.