Framing Pictures: The Role of Knowledge in Automatized Encoding and Memory for Gist

Article *in* Journal of Experimental Psychology General · October 1979

DOI: 10.1037/0096-3445.108.3.316 · Source: PubMed

1 author:

Alinda Friedman
University of Alberta

Journal of Experimental Psychology: General
1979, Vol. 108, No. 3, 316–355

Framing Pictures: The Role of Knowledge in Automatized Encoding and Memory for Gist

Alinda Friedman
University of Alberta, Edmonton, Canada

SUMMARY

In general, frame theories are theories about the representation and use of knowledge for pattern recognition. In the present article, the general properties of frame theories are discussed with regard to their implications for psychological processes, and an experiment is presented which tests whether this approach yields viable predictions about the manner in which people comprehend and remember pictures of real-world scenes.

Normative ratings were used to construct six target pictures, each of which contained both expected and unexpected objects. Eye movements were then recorded as subjects who anticipated a difficult recognition test viewed the targets for 30 sec each. Then, the subjects were asked to discriminate the target pictures from distractors in which either expected or unexpected objects had been changed.

One consequence of the embeddedness of frame systems is that global frames may function as "semantic pattern detectors," so that the perceptual knowledge in them could be used for relatively automatic pattern recognition and comprehension. Thus, subjects might be able to identify expected objects by using automatized encoding procedures that operate on global physical features. In contrast, identification of unexpected objects (i.e., objects not represented in the currently active frame) should generally require more analysis of local visual details. These hypotheses were confirmed with the fixation duration data: First fixations to the unexpected objects were approximately twice as long as first fixations to the expected objects.

On the recognition test, subjects generally noticed only the changes that had been made to the unexpected objects, despite the fact that the proportions of correct rejections were made conditional on whether the target objects had been fixated. These data are again consistent with the idea that local visual details of objects represented in the frame are not necessary for identification and are thus not generally encoded. Further, since subjects usually did not notice when expected objects were deleted or replaced with different expected objects, it was concluded that if two events instantiate the same frame, they may often be indistinguishable, as long as any differences between them are represented as arguments in the frame. Thus, for the most part, the only information about an event that is episodically "tagged" is information which distinguishes that particular event from others of the same general class. The data reinforce the utility of a frame theory approach to perception and memory.

The Phenomenon of Stimulus Equivalence

When Is a Rose a Rose?

Under what conditions can we say that x is like y, or that x means y, or that x belongs with y?[1] These questions have occupied philosophers for centuries, and are not unknown in other disciplines: Linguists deal with issues of paraphrase, synonymy, and polysemy; physiologists characterize the set of trigger stimuli to which a single cell will respond; art historians discover similarities between Rembrandt and Rothko; physicians find the symptom complex of a disease. Perception of similarity is the rule rather than the exception; people typically react to generic qualities, in most cases emergent, of a myriad environment. They perceive categorically; that they do so is puzzling because it is difficult to understand how such generally effortless categorization is possible, given an environment that under most circumstances is different, at least in detail, from anything previously experienced.

Psychologists and philosophers alike have theorized about the mechanisms underlying the ability to see beyond "superficial" stimulus differences. Consideration of this phenomenon invariably leads to theories about the interaction between the mind and the world, which in turn may focus on the kind of knowledge necessary to support categorization of individual instances (e.g., knowledge about prototypes or common features), and the means by which that knowledge is represented (e.g., images, propositions, or procedures). The major purpose of the present article is to evaluate whether recent ideas about the representation of knowledge by frames or frame systems (e.g., Minsky, 1975) may be used to address the issues implicit in the stimulus equivalence phenomenon.[2] In particular, the general properties of frame theories will be discussed from the point of view of how they might give rise to both "automatic" perception and stereotyping in memory. In addition, it will be argued that frame theories allow the a priori specification of the conditions under which two pictures may be indistinguishable paraphrases of one another. Then, an experiment will be described which demonstrates that this approach makes it possible to specify which objects in a scene will require predominantly "data-driven" analyses of their local visual features in order to identify them, and which objects may be comprehended relatively automatically through the use of global information described in extant memory structures. Further, the experiment demonstrates that it is possible to specify the objects in a picture that either will or will not

This research was supported by National Science Foundation Grants BNS72-02084 and BNS76-81416 and is based on a doctoral dissertation submitted to the Graduate School of the University of Colorado, Boulder. The research was conducted at the Institute for the Study of Intellectual Behavior, University of Colorado, and is Publication No. 84 of the Institute.

I would like to acknowledge my indebtedness to Lyle Bourne, who supervised the research and contributed enormously to its conduct and to the development of my thinking. Thanks are also expressed to Rick Barclay, Walter Kintsch, Dick Olson, and Gary Stahl for their guidance and insights; to Peter Polson for many invaluable discussions of the theoretical issues involved; and to Geoff Loftus, Mike Masson, Doug Nelson, and Linda Sala for their helpful comments on earlier versions of the manuscript. Finally, the manuscript benefited a great deal from a careful, thorough, and insightful critique by Jean Mandler.

Requests for reprints should be sent to Alinda Friedman, Department of Psychology, University of Alberta, Edmonton, Alberta, Canada T6G 2E9.

[1] Implicit in these questions is the assumption that x and y are not identical in the strict sense; though of course they may be, as is the case when we see different perspective views of the same object. The more interesting and problematical cases are those involving two or more physically or temporally distinct events.

[2] Throughout this article, the terms *stimulus equivalence, pattern recognition, stimulus identification, categorization,* and *comprehension* will often be used interchangeably to refer to the phenomenon wherein an event x is taken to be the same as or similar to a physically or temporally distinct event y. Either x or y or both may exist in the mind or in the world. For example, to recognize an object in the world as a chair presupposes that the object bears some resemblance to a previously experienced chair; the recognitive experience therefore qualifies as an instance of stimulus equivalence.

be remembered, regardless of whether it is certain they were encoded, regardless of whether subjects expect a recognition test, and regardless of the fact that subjects are allowed a full 30 sec to explore each scene and are tested immediately thereafter.

Pictures were chosen as the objects of study because they are often viewed as theoretically intractable; we know relatively little about visual memory, and even less about the semantics of pictures. In addition, pictures are complex stimuli that are more nearly "like" the world than text is. They are the proverbial equivalents of 1,000 words, although recent (tongue in cheek) estimates have run as high as 50,000 (Raphael, 1976). Each picture is supposed to be a unique entity, whereas a word can "stand for" many exemplars. For this reason, pictures are often held to be a class of objects that stands in clear contradistinction to the class of objects which are words. Thus, they provide an excellent set of materials for testing whether frame theories about knowledge can deal with issues of stimulus equivalence.

How Do We Know It's a Rose?

There are two classic approaches to the phenomenon of stimulus equivalence, although as Turvey (Note 1) has pointed out, they represent endpoints on a continuum. The relative importance assigned to either the world or the head with respect to perceptual experience determines the place of any particular theory along this continuum. One end of the continuum focuses on the cognitive contributions of the organism to its own perceptual experience; at the other end, the emphasis shifts to the contributions made by the environment to the perceptual experience of the organism. There is enough evidence for both of these extremes to make a compromise position inevitable. For example, there exist physiological data from which it can be plausibly argued that during past millenia, we have evolved neurological structures specialized for detecting certain higher order features of light from the environment, such as lines and angles (Hubel & Weisel, 1962) and depth relations (Bishop, 1970). Indeed, neurons in the inferotemporal cortex have been hypothesized to provide a mechanism for stimulus equivalence across retinal translations (Gross, Bender, & Rocha-Miranda, 1973; Rocha-Miranda, Bender, Gross, & Mishkin, 1975).

On the other hand, the knowledge that a particular rectangular solid is a book or that a particular surface is a tennis court is acquired during an individual's lifetime, which implies that the ability to apprehend the culturally given meaning of things should require an interaction between the output of relatively low-level feature analyzers (e.g., line detectors) and "higher mental processes" that can interpret how those features go together and what they mean. In general, then, it seems likely that past experience is necessary for the interpretation of perceptual input, however physically structured that input may be.

The position taken in this article is that the degree to which both environmental and memorial structural invariants can contribute to the apprehension of meaning will determine the relative automaticity with which that apprehension occurs. The use of the terms *automatic perception* and *automatized encoding* (Kolers, 1975) is therefore meant to distinguish between situations in which the apprehension of meaning occurs on the basis of salient structural features or global properties, and situations in which the observer engages in more attentive or highly interactive "hypothesis testing" procedures during the process of comprehending the events within his or her environment. In other words, a distinction will be made between feature *detection* and feature *analysis*. For the sake of clarity, the terms *automatic* and *interactive* will be used to contrast cases which allow feature detection with those that require relatively more feature analysis. Note, however, that in both cases comprehension involves an interaction between the observer and the information available from the environment, so that this position is different from the "direct perception" view proposed by Gibson and his colleagues (see Gibson, 1966, 1977; Turvey, 1977).

The theoretical approach elaborated here is derived from the artificial intelligence literature about frame theories (e.g., Bobrow & Collins, 1975; Minsky, 1975; Norman & Rumelhart, 1975a; Schank & Abelson, 1975; Schank & Nash-Webber, 1975). In general, frame theories are theories proposing that in the course of cognitive development, people evolve memorial structures which amount to "grandmother cells," so that under normal circumstances, an adult's perception and comprehension of objects and events is relatively automatic and is mediated via higher order memorial structures such as schemes (Norman & Rumelhart, 1975b, 1975c; Palmer, 1975c), or frames (Minsky, 1975), or scripts (Schank & Abelson, 1975). In addition, frame theories generally make it possible to specify the conditions under which more data-driven processing may be necessary for comprehension. In the remainder of this article the following arguments will be presented:

1. Context should be broadly defined to include any information whatsoever which is not inherent in the stimulus event per se, including both structural invariants in the optic array and relatively invariant memorial structures such as frames. That is, the context in which an object or event is comprehended is always both internal and external, and comprehension will therefore always involve both memorial and environmental structural invariants. However, the processes involved in comprehension may be predominantly automatic or predominantly interactive.

2. When higher order environmental or memorial structures are the predominant means by which an organism perceives the meaning of an event, then we may characterize the perception of that event as being relatively automatic. From the frame theory view, this mode of perception is termed *top-down*.

3. When for some reason neither environmental nor memorial global structures are sufficient for categorization or recognition, procedures are invoked that lead to the sort of hypothesis testing behavior which can be characterized as an interaction between bottom-up and top-down procedures. This type of interaction is generally considered to be necessary for frame verification (Minsky, 1975).

4. There is a trade-off between familiarity (of stimuli or situations) and the degree of automaticity that is possible. In general, the more familiar or expected a particular event is, the more likely it is that comprehension of that event will be relatively effortless and automatic. This is what is normally meant by the statement that "context helps comprehension." Both environmental (external) and memorial (internal) contexts may facilitate (automate) perceptual experience. In contrast, the more unfamiliar or unexpected a particular event is, the more likely it is that its understanding will require attentive, resource-expensive, generate-and-test procedures.

5. There is a similar trade-off between the globality of an event (e.g., an isolated letter vs. a word in a paragraph, a single object vs. an entire scene) and the degree of automaticity that is possible. Both external and internal contexts are "disambiguating"; events occurring within relatively global contexts will tend to be comprehended automatically, whereas isolated events may generally require more interactive processes for their identification.

6. Automatic perception generally requires fewer resources than does interactive perception, and is one of the primary causes of stereotyping in memory.

These ideas will be elaborated in the following sections. Supporting evidence will be given from previous research, and then an experiment whose purpose was to test some straightforward predictors of the theory will be described.

What Is a Frame Theory?

A frame theory is primarily a theory about the representation and use of knowledge for pattern recognition. The implications of such a theory for perception and memory are implicit in the notions that frames represent different types of knowledge (e.g., physical, syntactic, semantic, pragmatic, relational) in an abstract format, that they are organized into systems, and that the knowledge represented by such systems is stereotypical.

Despite the variety of instantiations of the theory (e.g., Bartlett, 1932; Bobrow, 1975; Bobrow & Norman, 1975; Charniak, 1975; Hewitt, 1975; Kintsch, 1974, 1977; Kolers, 1975; Kuipers, 1975; Minsky, 1975; Nash-Webber, 1975; Norman & Rumelhart, 1975b, 1975c; Palmer, 1975b, 1975c; Pylyshyn, 1973, 1975; Riesbeck, 1975; Rumelhart & Norman, 1975a, 1975b; Schank, 1975; Schank & Abelson, 1975; van Dijk, 1977; Winograd, 1972, 1975; Marr & Nishihara, Note 2), it is possible to abstract a few generally agreed-upon principles, although the reader familiar with this body of literature is forewarned of the emergent properties inherent in such generalizations.[3]

The Representation of Knowledge by Frames

Images, propositions, or programs? Most versions of frame theory abide by the notion that knowledge about the world is represented in a format that is more abstract than the sensory or linguistic data used to acquire it. That is, it is generally agreed that perceptual and linguistic knowledge, in addition to a person's skills, is represented by propositions (e.g., Kintsch, 1974; Norman & Rumelhart, 1975a, 1975b) or procedures (Winograd, 1975).[4] The bias that some sort of abstract representation is preferable to, for example, imaginal or verbal representations derives from considerations of power as well as parsimony. The excellent arguments that have been made against a dual or multicode theory of knowledge representation will not be reviewed here (Palmer, 1975b; Pylyshyn, 1973, 1975; see also Kosslyn & Pomerantz, 1977, for some counterarguments), except to note that many of the findings regarded as incontrovertible evidence for imaginal or "analogue" representations (Cooper & Shepard, 1973; Kosslyn, 1975; Moyer, 1973; Moyer & Bayer, 1976; Paivio, 1975) can be readily accommodated by a more abstract representation (Banks & Flora, 1977; Friedman, 1978; Lea, 1975; Palmer, 1975b). It is only necessary that a clear distinction be maintained between the format of the representation (e.g., whether it is propositional or analogical) and its contents (e.g., whether the information contained in the representation is primarily of a perceptual, linguistic, or conceptual nature; see Kieras, 1978).

The structure of knowledge. The second general characteristic of the representation of knowledge by frames is that frames are organized into hierarchies of conceptual "bundles"; what constitutes a bundle depends on the level of the hierarchy.[5] That is, a concept may be viewed as a network of propositions or procedures, and such networks may exist at several levels of abstraction, from single objects and events to scenarios or episodes. For example, the frame for a single object may have propositions or procedures that pertain to what that object generally looks like (e.g., propositions with arguments or "slots" for visual features and their structural relations, or procedures for detecting such features and relations). The frame for an object may also represent what that object does and the kinds of events it can enter into (e.g., where we are likely to find it or how we can interact with it). Thus, a "person frame" might include, among other

---

[3] Some of the authors cited may not regard themselves as frame theorists. The responsibility of assigning them to this category rests with the present author.

[4] There is some debate about whether the data structures that represent knowledge are declarative (i.e., whether knowledge is "knowing what" and is represented propositionally), or whether the data structures are instead procedural (i.e., whether knowledge is "knowing how" and is represented via programs or instructions; see Winograd, 1975). For the present, it is immaterial which of these views is correct. However, a procedural view may be preferable because it seems more adaptable to automatized encoding, in addition to being the approach used by some fairly successful simulations of complex understanding systems (e.g., Charniak, 1975; Winograd, 1972).

[5] The organization may actually be either hierarchical (e.g., Rumelhart & Norman, 1975a, 1975b) or heterarchical (e.g., Winograd, 1972; Minsky & Papert, Note 3). For the present, the distinction is unimportant, and a hierarchical organization will be used to illustrate the major concepts. Note, however, that a hierarchical system of frames is not isomorphic to the more traditional view of hierarchical feature analyzers (Selfridge & Neisser, 1960), since very global frames may be used for feature detection without necessarily having input from lower level frames.

things, arguments for arms, legs, opposable thumbs, bilateral symmetry, ranges of heights, weights, colors, and so forth. Any one of these concepts could be expanded into a frame in its own right. For example, the argument "head" is acutally a subframe that has arguments like "eyes" and "nose," each of which could also be expanded.

Similarly, it is possible for a frame to be more global: The frame for an episode may consist of arguments or slots that are themselves frames and that function as "headings" for the actors, actions, objects, and events most relevant to that particular episode. The person frame described above, for example, could be a subframe of any other frame in which people, as a class of objects, belong, like a "birthday party frame."

In general, then, a frame is a function that specifies the relations that hold among the arguments comprising a particular conceptual bundle at a particular level of abstraction. Importantly, the more abstract (generic) a frame is, the more global are its arguments, both visually and conceptually. This property of frames has implications for the automatic perception of meaning.

Default knowledge and the invariance of frames. The third general property of frames is that their terminal nodes are normally assigned; that is, they have "typical" values or ranges. The assignments depend on an individual's history with his or her environment. History repeats itself; frame systems take advantage of (represent) the repetitive nature of our experience with "mundane reality" (Abelson, 1975a). People know about likelihoods, ranges, and distributions of things and events; this is default knowledge, and is in part what is generally meant by world knowledge. Although an individual's world knowledge will of course be idiosyncratic, default knowledge will tend to overlap among individuals who share the same sociological niche (i.e., age, education, socioeconomic class, race, religion, sex, interests, geography, etc.). For example, North Americans who live in the 1970s know stoves as squarish, usually enameled, electricity- or gas-using things, and not as large and roundish, cast-iron, wood- or coal-burning things.

Default knowledge is responsible for our ability to know or infer things that we do not specifically experience. For example, consider the following two sentences as if they each each began a story: (a) John took Mary to a French restaurant; (b) John took Mary to a run-down diner. Though the sentences differ only with respect to the setting of the action, we are likely to have very different ideas about John, Mary, their possible actions and conversation, the sort of clothing they may have worn, the sort of food they may have eaten, whether there were waiters or waitresses in the establishment, whether there were tablecloths or jukeboxes, candles or fluorescent lights, etc. In a similar vein, consider all the information you have after walking through the kitchen of an unfamiliar house. Even if you were daydreaming or looking at your feet, you would probably respond "yes" to the query "Was there a stove?" simply because you know that kitchens have stoves. Thus, default knowledge may both extend and stereotype our perceptual experiences.

The assignments in a frame may be either tight (i.e., admitting of few or no exceptions, such as "red traffic lights mean stop"), or fuzzy (i.e., admitting a range of possibilities, such as "the size of dogs" or "the number of people at a birthday party"). In addition, though the variables in a frame are often assumed to be bound by a specific experience, it will be argued that frames function as the "given" in most situations (Haviland & Clark, 1974), so that their default values will usually be replaced with specific values temporarily, for example, *during* a particular perceptual experience. So long as the specific values of an experience that instantiates a particular frame or set of frames fall within the range allowed by the default knowledge contained therein, the "new" values will not generally replace the old values permanently. This is part of what is meant by the *invariance* of frames.

Invariance of frames further implies that in general, people perceive the world deterministically. For a particular individual in a particular situation, events are either expected (within the realm of possibility, i.e., the frame) or not. By the time we are adults, we have had a wealth of experience with the world and do not tend to change our belief or knowledge systems impetuously. Both default knowledge and the invariance of frames have implications for what we may expect people to remember about their experiences.

Implications for Psychological Processes

The Role of Frames in Perception and Comprehension: Automatized Encoding

Frames function as semantic pattern analyzers when they assume an interactive role with data-driven procedures; but they may function as semantic pattern detectors once they have been invoked, since they constitute a set of expectations about what belongs where, as well as a set of descriptions about what those things generally look like. Therefore, once an appropriate frame is evoked, its descriptions may be used to guide subsequent processing (Schank & Abelson, 1975). But note that under normal circumstances (e.g., in a familiar environment), the appropriate frame or frames are active as a matter of course, so that perception will tend to be automatic under these circumstances. Even in relatively artificial situations, it has been shown that deriving a context (evoking a frame) for a complex picture can occur in as little as 100 msec (Biederman, Rabinowitz, Glass, & Stacey, 1974). Thus, instantiating a frame has the effect of creating a *perceptual bandwidth* for subsequent object identification. That is, the amount of detail necessary to identify an object decreases as the frame of reference becomes larger (Palmer, 1975b, 1975c) because there will be fewer obligatory slots per item that need to be checked. In the extreme case, we might suppose that object identification could occur holistically. For example, if you know you are walking into a kitchen, you can satisfy the terminal nodes for a "refrigerator argument" by being prepared to see such relatively global features as "a three-dimensional rectangular solid that is shiny and about as tall as a person." In contrast, without a context, or in a strange context (such as the same refrigerator depicted in a living room), the object itself must provide all the cues for its identity, and so more detail should be necessary. Thus, an object in a strange context cannot be comprehended by using global features that are specified by a high-level place frame; instead, local features (e.g., lines, corners, texture) must be used as clues to search for and fill the terminal nodes of a frame that is generic at the level of a single object rather than at the more global level of a place (see Kolers, 1975, for a discussion of automatized encoding of linguistic materials).

It may seem obvious, but we do not generally know *about* novelty without some further processing; however, we can know *that* something is novel with respect to our current frame of reference, because a novel item will be missing from that frame. Thus, one of the functions of frames as pattern detectors is to free us and our resources for processing the novel and unexpected aspects of the environment.

Evidence for automatized encoding. That both external context and knowledge structures facilitate comprehension is not a particularly new idea; however, effects of priming, set, redundancy, expectancy, and so on, are usually viewed as evidence for the highly constructive or interactive nature of comprehension. In the present discussion, it is proposed that at least some of these effects can be taken as evidence for the relative automaticity with which most things are comprehended, and that in fact, it is the unexpected and unfamiliar stimulus whose comprehension will demand more of an interaction between top-down and bottom-up procedures.

Perceptual facilitation by structure that is external to the stimulus (including knowledge structures) has been demonstrated with meaningful, nonmeaningful, linguistic, and nonlinguistic materials in a wide range of tasks, including simple and complex pattern, word, and object recognition, visual search, reading, lexical decisions, and simultaneous matching. For example, using nonmeaningful materials, Weisstein and Harris (1974) demonstrated that the identification of even simple features was facilitated by external structure, and Palmer (1977) showed that global features are more easily recognized than figural details. Similarly, with meaningful materials, identification occurs more rapidly when objects are presented in coherent scenes (Biederman, 1972; Biederman, Glass, & Stacy, 1973), or words are presented in paragraphs (Lefton & Fisher, 1976), than when either type of stimulus is presented alone or in a jumbled (nonstructured) surround. In addition, that perception can be guided by world knowledge has been demonstrated by manipulating the rules of physical plausibility (Hock, Gordon, & Whitehurst, 1974), linguistic rules (Just & Carpenter, 1976), the prior associations among items in a lexical decision or object naming task (Meyer, Schvaneveldt, & Ruddy, 1975; Oldfield & Wingfield, 1965; Wingfield, 1968), and the rated informativeness of areas to be searched among or recognized in a picture (Antes, 1977; Pollack & Spence, 1968). Further, schematic drawings, which emphasize global properties, are identified faster or better than detailed photographs (Ryan & Schwartz, 1956); and relatively abstract primes, such as category and object names (Potter, 1975; Rosch, 1975) or thematic descriptions (Potter, 1976), result in more rapid identification or matching of meaningful pictorial stimuli. Finally, misidentifications occur as a result of expectations (Mial, Smith, Doherty, & Smith, 1974; Palmer, 1975a), most likely because such expectations evoke a frame of reference with a set of feature detectors for a particular *class* of objects; any object whose shape is at all similar to the expected shape is thus "recognized" as such. In general, then, the ability to detect rather than analyze a stimulus may be induced by a variety of experimental operations which serve as bridges between the stimulus and the knowledge structures that can facilitate its comprehension.

The experiments cited above are by no means exhaustive, and are difficult to reconcile with the idea that recognition occurs via the stage-wise analysis and subsequent assembling of very elementary features. The data provide converging evidence that both environmental and memorial contexts are instrumental in automating the detection of structure and meaning, since these contexts allow the use of global and sometimes quite abstract information. In the experiment below, this issue will be addressed by using the duration of an eye fixation as a relatively direct and unobtrusive measure of the time required for predominantly automatic versus predominantly interactive object identification.

The Role of Frames in Memory: Keeping the Baby and the Bath Water

Given that a picture or real-life setting has been comprehended as an instance of a relatively global place frame (e.g., a kitchen scene), there are at least two broad classes of information that can be remembered or forgotten about any particular object in it: (a) whether or not the object occurred, which is *episodic* information, and (b) what the object looked like, which is *descriptive* information. Logically, both episodic and descriptive information can be remembered or forgotten about either objects that exist as arguments in the frame used to comprehend a picture, or objects not represented in that particular global frame, whose comprehension should theoretically require finding a lower level, single-object frame through the use of interactive verification procedures. However, one of the implications of the fact that much of what we encounter as "new input" is an allowable variation of an "old theme" is that the expected and mundane portion of any new experience could be stored as an instance of the particular global frame that it instantiates, without regard to specific episodic or descriptive details. More precisely, two events which instantiate the same global frame may qualify as paraphrases of each other, and may often be memorially indistinguishable, so long as any episodic or descriptive differences between them are represented as arguments in that frame.

As an example of what is meant by pictorial paraphrase, assume that your kitchen frame contains the knowledge (in its argument list) that kitchens have counters, and that counters often have things like toasters, blenders, coffee percolators, and the like, on top of them. If this is so, then toasters, blenders, percolators, etc., are "synonyms" with respect to things that belong in kitchens. Now suppose you are first shown a picture of a perfectly normal and mundane kitchen with a percolator on one of its counters, and a little later a second picture which is identical to the first, except that the percolator has been replaced by a blender. The fact that both objects exist as arguments in your kitchen frame permits you to comprehend both pictures as instantiations of that frame. Thus, the two pictures are paraphrases because a percolator and a blender could be substituted for each other without changing the overall meaning of either picture, that is, *without the necessity for any other frames to be evoked during comprehension.* In contrast, a picture of a kitchen with a fireplace in it should require at least two frames for its comprehension—one global frame for the kitchen in general and one single-object frame for the fireplace. Such a picture could not paraphrase any other picture that could be comprehended with a kitchen frame alone. Thus, in situations in which a perceptual experience instantiates a frame, part or all of the knowledge embodied in the frame may be predicated to that experience, so that under certain circumstances it will not be possible to distinguish between what was understood and the knowledge structures that facilitated the understanding. In practically any situation except a memory experiment, the fact that frames automatically extend our direct observations is beneficial, as this allows us to make inferences about the real world (e.g., "Since this is a kitchen and I don't see a garbage pail anywhere, there's a good chance it's in the cabinet under the sink").

Storage and forgetting heuristics for expected objects: The bath water. The discussion above suggests that the default knowledge in frames may underlie a kind of *storage heuristic,* which will be referred to here as "remembering by prototyping." Remembering by prototyping can occur with respect to either episodic or descriptive information, and reflects the fact that it is not normally useful to take particular note of either type of information for objects that have a reasonably high a priori likelihood of being found in a particular place and that look the way they are expected to look. These types of objects will be referred to as *obligatory.* The percolator/blender example given above is an example of how this storage heuristic can be used to predict the sort of episodic information we might expect people to confuse; that is, if episodic information about obligatory objects is indeed not normally tagged, then these objects constitute things that could be substituted between pictures to form paraphrases that would then be memorially indistinguishable.

Obligatory objects are also prime candidates for having their visual details remembered by prototyping, which is to say that the descriptive information of obligatory objects may not usually be stored at all (e.g., "There was probably a stove in the kitchen I just walked through, but I can't say for sure whether it was a gas or an electric one"). The most likely reason for this is that under normal circumstances, the visual details of obligatory objects are unnecessary for their identification when they are being comprehended within a global context. Indeed, objects can only take on an obligatory or nonobligatory or unexpected status with respect to a context. Thus, one of the behaviorally observable implications of remembering by prototyping is that memory for either episodic or descriptive information about obligatory objects may generally be poor, even when tested shortly after those objects have been perceived.[6]

---

[6] Whether or not a particular expected object is episodically tagged or remembered in descriptive detail may of course depend on many sorts of factors: For example, if you are in the market for a drip coffee maker, you may very well begin noting and remembering not only who among your acquaintances owns one, but what make and model seems to be the most popular, etc. You might unbind this information only after having made your purchase. Similarly, it seems reasonable that if the descriptive information of an expected object falls outside the range specified by the global frame, then that object becomes essentially unexpected, and should share the same memorial fate as other unexpected objects. For example, a stove of some sort is expected to be in a modern kitchen, but not a cast-iron stove, which is descriptively "wrong"

The default knowledge in frames may also underlie at least one kind of *forgetting heuristic* (Schank & Abelson, 1975), namely, "forgetting by prototyping." Forgetting by prototyping can also occur with respect to either episodic or descriptive information. Further, both remembering and forgetting by prototyping are instances of *normalization;* the distinction between them derives from when that normalization takes place. In contrast to remembering by prototyping, forgetting by prototyping is an example of normalization that occurs over time. That is, information about the occurrence or the descriptive details of certain objects and events may persist in memory for some time before it finally becomes unbound from the representation. Thus, the stereotyping or normalization of a picture in memory reflects a prototypical rather than veridical memory for either the existence of the objects in it or their visual details.

Objects that are *nonobligatory* with respect to a frame (i.e., objects that have been associated with a particular place often enough that their occurrence is not untoward, such as plants in a kitchen) are prime candidates for the eventual forgetting of descriptive and perhaps episodic information, while *unexpected* objects (i.e., objects not on the argument list of the global frame) should only be subject to forgetting of their visual details, for reasons that will be elaborated below. Note that in the long run, however, both remembering and forgetting by prototyping yield the same memory representation; obligatory and nonobligatory objects (i.e., the expected objects represented within a frame) will generally be remembered stereotypically.

Before turning to a discussion of memory for unexpected objects, there are several distinctions between obligatory and nonobligatory expected objects that warrant highlighting. First, obligatory objects are more closely related to the theme of an event, while nonobligatory objects generally correspond to extraneous or less important details. Second, obligatory objects are diagnostic with respect to instantiating or verifying a frame of reference (cf. Biederman, 1972; Biederman et al., 1974; Loftus, 1976), but by definition must be nondiagnostic with respect to discriminating among instances of the same frame (e.g., a stove is diagnostic for kitchens in general, but not for a particular kitchen). Third, obligatory objects should be sufficient, but not necessary, for frame instantiation, since any particular instance of a frame will not generally contain all the obligatory objects that are possible (e.g., not every picture of a kitchen will depict the stove which is assumed to be there). In contrast, nonobligatory objects are neither necessary nor sufficient for frame instantiation; they are thus nondiagnostic with respect to instantiation, but may be somewhat helpful with respect to discriminating among instances of the same frame (e.g., "Aren't the geraniums in Mary's kitchen lovely?").

The fourth difference between obligatory and nonobligatory objects is that obligatory objects should be perceived relatively automatically, using predominantly top-down procedures. However, since nonobligatory objects are nondiagnostic with respect to frame instantiation, it is possible that their frame descriptions are less specific than the descriptions for obligatory objects. For example, a kitchen frame would probably have an obligatory argument for "stove" rather than only one general argument for "appliance," while the same frame might have a nonobligatory argument for "plants" rather than one that specified "geraniums." Consequently, identifying nonobligatory objects may require more analysis of visual detail than identifying obligatory objects would. The details of nonobligatory objects which are recruited for the exigencies of identification may therefore persist for at least a little while before they become unbound. This is

---

even though it could perform the necessary functions. Conversely, a brick fireplace is not an expected kitchen object (i.e., it is episodically wrong), even though it could look like a perfectly normal brick fireplace. Thus, with respect to the instantiation of a particular frame, there can be expected or unexpected objects whose details do or do not conform to the ranges specified by default knowledge; expected objects with unexpected details are likely to be comprehended and remembered similarly to unexpected objects.

why episodic and descriptive information about obligatory objects is likely to be subject to remembering by prototyping (e.g., "I don't really know for sure, but I assume that Mary has a toaster"), while information about nonobligatory objects might be forgotten by prototyping (e.g., "Mary has some lovely plants in her kitchen"). In general, then, episodic and descriptive information about objects represented as either obligatory or nonobligatory arguments in a relatively global frame may be perceived and even consciously acknowledged, but may not generally be permanently stored or even episodically tagged.

Memory for unexpected objects: The baby. A further means by which frames can influence memory is implicit in the distinction between feature detection and feature analysis, as well as in the fact that frames are generally invariant. Any new input that is truly new (e.g., objects which are not represented in the argument list of a currently active frame) will tend to be remembered more veridically than either obligatory or nonobligatory expected objects for two reasons. First, since automatized recognition procedures should require few or no resources, while interactive procedures are resource expensive, it is likely that when interactive procedures are necessary for object identification, the result is a relatively more elaborate representation with respect to descriptive information. Remember, however, that automaticity is a matter of degree; so that although we assume that nonobligatory objects require some interactive processing for comprehension, they should require less of this than unexpected objects, since they are represented in the global frame. For example, since plants are nonobligatory but certainly acceptable things to find in a kitchen, the perception of green near a kitchen window might readily become assigned to the terminal node for plants.

The second and more important reason why the occurrence of unexpected objects should be well remembered is that since frames are relatively invariant, a stimulus that is unusual or unexpected with respect to a particular frame should not generally change the contents of the frame; rather, the representation of such a stimulus may be "stuck on" to the frame. For example, Schank and Abelson (1975) believe that one thing which is stored after reading a text is a "weird list," whose function is to mark departures from normality (i.e., from the frame, in the present terminology). Unusual events are invaluable for distinguishing among instantiations of the same general frame or theme. Indeed, they are generally what makes a text or situation interesting and different from previous experiences of the same class of text or situation. Thus, in the case of a picture, a weird list is essentially a list of object frames representing objects that were not in the argument list of the global frame used to comprehend most of the picture. The weird list thus represents the episodic differences between that particular picture and the prototype it instantiates. This episodic difference structure is therefore the most important means by which a person may be able to uniquely identify and remember any particular instantiation of a frame, since it is what makes one picture mean something different from another picture that instantiates the same global frame. Thus, a weird list which is appended to a frame allows important episodic information to coexist with, but be distinguishable from, default knowledge, while more commonplace episodic information may be indistinguishable from default knowledge. Note, however, that descriptive information about unexpected objects is relatively inconsequential with respect to discriminating among particular frame instantiations; it may therefore be subject to eventual forgetting by prototyping.[7]

[7] In the original version of this article, it was stated that nonobligatory objects were nondiagnostic with respect both to frame instantiation and to discriminating between instances of the same frame. I am grateful to Jean Mandler for pointing out that under normal circumstances, environments do not contain objects that are truly unexpected (e.g., a swimming pool in a living room). Thus in real-life situations, it is likely that nonobligatory objects are used to discriminate between instantiations of the same general frame. Generally speaking, people probably note the differences between any particu-

To summarize, it is likely that additions, deletions, substitutions, and other changes that may distinguish two events may be undetected if those events instantiate the same frame during their perception, and if the new or changed objects exist as arguments in the instantiated frame. In contrast, unexpected objects may be represented as a list of object frames in a separate difference structure, whose terminal nodes probably have much more descriptive detail than those for the nonobligatory objects in the frame. The separate representation of unexpected objects makes it likely that their occurrence will be better remembered than will the occurrence of expected objects.

Evidence for the "gistification" of events in memory. The implications of frame theory for event memory have not really been directly tested, but there are data that support them to a certain extent. First, world knowledge affects what is retained from a text, so that the types of things remembered can be directly attributed to the frames which were likely to have been instantiated during comprehension (Abelson, 1975b; Anderson & Pichert, 1978; Frederiksen, 1975; Anderson, Reynolds, Schallert, & Goetz, Note 4). Second, people seem to remember primarily the theme or gist of textual materials, while forgetting subordinate or surface structure details (Bartlett, 1932; Kintsch, 1977; Kintsch & Keenan, 1973; Mandler & Johnson, 1977; Rumelhart, 1977; van Dijk, Note 5). Third, theme-related paraphrases or additions to a text are generally not noticed as being new in a recognition test, or else are found as intrusions in recall protocols (Barclay, 1973; Bransford, Barclay, & Franks, 1972; Frederikson, 1975; Johnson, Brandsford, & Solomon, 1973; Sulin & Dooling, 1974).

It is difficult, however, to use the data from experiments of text comprehension and memory as overwhelming support for (a) the applicability of frame theory to perceptual experience in general and (b) the relationship between automatic and interactive identification procedures and memory. In the first place, our daily interactions with the world involve much that is beyond the domain of textual materials. Further, although Kolers (1975) found that verbatim sentence memory decreased as the ability to read inverted text became automatized, understanding a text as a *story* may involve procedures different from those used to understand a perceptual event or a static representation of a perceptual event, such as a picture. Thus, while it is generally agreed that frame-like structures are involved in understanding the events within a text, the interaction between frames and text structure is a complex issue (cf. van Dijk, 1977), and will not be discussed here. Instead, reading a text will be regarded as tantamount to acquiring secondhand knowledge about a sequence of events, which probably involves different procedures than does acquiring knowledge through direct perceptual experience.

When we turn, however, to experiments that have used pictorial materials, it appears at first as if the same subjects who can only manage to remember the gist of a text are virtual eidetic geniuses with respect to pictures (e.g., Bahrick, Bahrick, & Wittlinger, 1975; Nickerson, 1965, 1968; Shepard, 1967; Standing, Conezio, & Haber, 1970). Nevertheless, these findings are amenable to a frame theory interpretation because sub-

lar instance of an event and their frame representation (i.e., prototype) of that event, and then store a difference structure. In the present experiment, it was reasonably certain that in contrast to the nonobligatory objects, the unexpected objects that were used would not be represented in the frames for each of the places depicted. Thus these objects should have maximally discriminated the pictures from any prototypical representations which subjects might have brought with them to the experiment. Under more naturalistic circumstances, nonobligatory objects (e.g., a fireplace in a living room) would most likely play more of a role in discriminating among instances of the same frame of reference. Nevertheless, the main point of this section still holds: Within a particular context, objects that become stored in a difference structure, and thus best discriminate among particular frame instantiations, can be expected to be better remembered (episodically) than those objects which are not. In the present experiment, the difference structures for the pictures used were predicted to be comprised of the unexpected objects; in more naturalistic settings, they are probably comprised of nonobligatory objects.

jects may appear to be very accurate in recognition when what they actually remember is an abstracted theme. That is, a set of distractors which is thematically heterogeneous with respect to the target pictures may allow subjects to appear to have excellent verbatim memory, when in fact performance is actually based on memory for gist. For example, there are data that support the view that people do indeed generally remember just the themes of pictures, so that false recognitions occur to distractors that are conceptually consistent with these themes (Bower & Glass, 1976; Jenkins, Wald, & Pittenger, 1977; Mandler & Johnson, 1976; Mitterer & Rowland, 1975). However, while these data are encouraging, the evidence they provide for specific frame theory predictions is circumstantial, at best. Moreover, these studies do not directly address the relationship between automatized encoding and memory, or the circumstances under which details may be either forgotten or remembered verbatim. Further, there have been no systematic efforts to determine the precise circumstances under which two pictures would be *semantic* paraphrases of one another (i.e., when they would instantiate the same frame). Jean Mandler and her colleagues (Mandler & Johnson, 1976; Mandler & Parker, 1976; Mandler & Ritchey, 1977; Mandler & Stein, 1974) are developing a taxonomy of the effects of different distractor transformations on recognition memory for pictures; the materials they use, however, are very simple (e.g., eight-object), highly schematic drawings. Further, while they correctly assume that a jumbled–nonjumbled manipulation is sufficient to make a distinction between a person's ability or inability to see a picture in terms of a frame, they fail to manipulate the likelihood of the objects within their pictures (however, see Parker, Note 6, who did make use of some unexpected items). Transformations may have quite different effects on recognition, depending on whether they have been applied to obligatory, nonobligatory, or unexpected objects. This issue will be addressed in the experiment below by manipulating the a priori likelihood of the objects in the target pictures, and then subjecting them to a variety of transformations in the distractors.

A Test for Frame Theory

Consider the following situation: People are asked to view six complex pictures of coherent, real-world scenes for 30 sec each, and their eye movements are recorded while they do so. Each picture represents a different place, and the subjects are primed with the names of the places before they see the pictures. Further, they are forewarned of a difficult test of their memory for the objects in the pictures. The pictures consist primarily of expected objects, although they each have a few unexpected objects; categorizing the objects is a function of independently obtained ratings.

The experimental context just described was used to test three hypotheses. The first prediction tested was that comprehending expected objects is relatively automatic, while comprehending unexpected objects entails more interactive recognition procedures, which should be manifested by a difference in the amount of time necessary to identify the two different types of objects. This hypothesis follows from the fact that when a particular frame is activated by a prime, an object that is represented by a frame argument should take less time to identify than an object that is not; it should take less time because an expected object should be detectable on the basis of relatively global features, given an activated memorial structure to preinterpret what those features are. In contrast, identifying an unexpected object should theoretically require more of an interaction between bottom-up and top-down procedures; this should take longer because the "pieces" have to be put together from scratch, so to speak, and a new frame of reference, at the level of the object rather than at the more global level of the scene, must be found.

It has been known for some time, of course, that people tend to look longer or more often at "informative" areas of pictures (Antes, 1974; Antes & Stone, 1975; Buswell, 1935; Durham, Nunnally, & Lemond, 1971; Loftus, 1972; Mackworth & Morandi, 1967), but definitions of informativeness have varied from amount of visual contour to ratings of what subjects say (without other instruction) is informative. Experiments that vary semantic informativeness are notably rare (however, see Loftus & Mackworth, 1978, and Scinto, Note 7). In the present experiment, the rated likelihood of the objects in the target pictures was systematically varied. Since several investigators have suggested that the locus of an eye fixation may be used to indicate that a specific item has been encoded (Just & Carpenter, 1976; Loftus, 1972), the additional assumption will be made here that the duration of the *first* fixation to an object reflects the amount of time it takes for that object to be identified, so that expected objects should receive much shorter first fixations than unexpected objects.

The second implication of frame theories that was tested was that predominantly top-down versus predominantly interactive encoding procedures would yield representational differences between expected and unexpected objects, which would be revealed by differences in retention of descriptive details as a function of the rated likelihood of the objects. In particular, unexpected objects should have more details in their representations than either obligatory or nonobligatory expected objects do, because their local visual features should be essential for identification. Subjects should therefore easily be able to acknowledge any changes made to such objects. Changes made to the descriptive information of expected objects, on the other hand, may go entirely unnoticed.

The third implication tested was whether knowing which frames were used to encode the original pictures would allow prediction of which distractors would be indistinguishable from the original pictures. For example, a distractor that contains new or missing objects may be falsely recognized if those objects can be reasonably assumed to be arguments in the originally instantiated frame. In contrast, the hypothesized weird list (i.e., the difference structure) should be something like a very short and detailed object inventory, and deleted unexpected objects should be easily noticed. In addition, new unexpected objects should be noticed regardless of what type of object they replace, because they will not exist either as arguments in the instantiated global frame or as objects on the weird list.

A particularly nice feature of the experiment is that the fixation records for each subject may be matched with his or her recognition performance, thus ensuring that false alarms actually do reflect memory losses, or paraphrase confusions, rather than encoding failures.

Method

Likelihood Norms

Normative data were collected to be used as a guideline in constructing pictures for the recognition experiment. Fifteen subjects were given the names of 15 different places, each on a separate sheet of paper, with three category headings typed under each name. The *must* category was reserved for objects that are found in a particular place 95% of the time (e.g., a stove in a kitchen). The *usually* category was for objects that, while often found in a particular place, are not definitively necessary (e.g., dishes in a kitchen), and the *sometimes* category was for objects that may be seen in a particular place but are not necessarily associated with it (e.g., plants in a kitchen). For each place named, subjects had 5 min to list as many things as they could think of for each of the categories.

After the data were tabulated, the seven places that yielded the greatest number of responses were chosen to be included in a rating study. The name of each place was typed on a separate sheet of paper, and an assortment of objects, with the numbers 1 through 7 typed beside them, were listed below that. The objects listed for each place included all of the items listed by the previous group of subjects; in addition, the lists were increased to include low probability objects and objects of a size and shape similar to those previously listed. A new group of 51 subjects rated the items, having been read instructions in which they were asked to rate the likelihood of seeing each of the objects in each particular place as a function of their experience (world knowledge) of such places. It was assumed that the ratings given to the names of these objects would be appropriate for their pictures, since the objects were drawn as normatively as possible and were not displayed in physically impossible relationships (e.g., a floating stove).

Stimulus Pictures

Targets. Complex and detailed line drawings of six different scenes (city, farm, kindergarten, kitchen, living room, and office) were used as the

Table 1
Mean Ratings of Objects Changed to Make Distractor Items for the Recognition Test

Transformation	Likelihood of object: High	Medium	Low
Token	1.453	3.745	6.011
Rearrange	1.722	3.835	5.429
Delete	1.576	3.626	5.665
Type[a]			
High	1.790	3.515	6.244
Medium	1.815	3.871	5.945
Low	2.038	3.742	6.484

[a] Each type transformation involved two objects, the original and the new object. In the row labeled "High," it is the original objects whose ratings were high ($M = 1.837$); the new objects had either a high, medium, or low rating, and the mean ratings of the new objects are given in the appropriate columns. For the type-medium transformation, the mean rating for original objects was 3.675; for the type-low transformation, the mean rating for original objects was 6.205.

target stimuli. Each scene was drawn in correct perspective, and contained between 25 and 34 different "areas." Across pictures, there were 170 different target areas which were scored; most of these (151) were simply the perimeters of single objects, but 19 areas included several objects with similarly rated probabilities (see Appendix A).[8]

Most of the objects within each scene had either a high or medium rated likelihood; in addition, each target picture had a few items whose likelihood of being seen in such a place was rated to be relatively low (e.g., a fireplace in a kitchen). It was assumed that obligatory, nonobligatory, and unexpected objects corresponded to objects that had received high, medium, and low ratings, respectively. Schematic representations of the target pictures are shown in Appendix A, which also contains a list of the objects in each picture and their rated likelihood.

Distractors. The distractor items for the recognition test were constructed by applying token, rearrangement, deletion, and type transformations to objects that had received either high, medium, or low ratings. Both the token and the rearrangement transformations were presumed to be tests of memory for different types of descriptive information about the objects (i.e., their visual details in the case of token changes and their location in the case of rearrangements); deletions and type changes were presumed to test memory for episodic information about the objects.

For the token transformations, an object from the same conceptual class as one of the original objects was substituted for it; for example, a different kind of stove replaced the original stove in the kitchen scene. For rearrangements, two objects from the original picture with similar sizes, shapes, and ratings switched places with each other.[9] It should be noted that since the scenes were drawn in correct perspective, it was sometimes necessary to change the spatial relationships among the features of an object to make it fit into the space occupied by the object it was to change places with, and vice versa; however, all of the details of both objects were exactly preserved. In the deletion transformation, a high, medium, or low probability object was deleted from the picture altogether. Finally, for the type transformations, an object from a conceptual class different from that of the original, with either the same or different rated likelihood, was substituted for the original; for example, a radio took the place of the original toaster on the kitchen counter. There were three kinds of type transformations (type-high, type-medium, and type-low), depending on whether the original object was of high, medium, or low rated probability. Thus, the nomenclature being used for the type transformations refers to the rated probability of the *original* target objects. The objects that were used for each distractor in each picture are listed in Appendix B.

Each distractor differed from the original scene by a transformation to only one object; the new objects were always the same general size and shape as those they replaced. Occasionally, a particular target object was changed for more than one distractor (e.g., in the office, a chest of drawers became a file cabinet for the type-low to high distractor, and a refrigerator for the type-low to low distractor); however, no subject ever saw any object changed more than once. Table 1 shows the mean likelihood ratings across pictures for the particular objects that were transformed for each of the 18 distractor types, and Figure 1 shows examples of the token and type transformations. It can be seen that in the absence of context, it is

[8] Four of the multiobject areas included objects that were to be transformed in the recognition test: the sofa in the living room (sofa and pillows), the stove in the kitchen (stove and pots), the piano in the kindergarten (piano and piano stool), and the teacher's desk in the kindergarten (desk, blotter, apple, and pencil cup.) When these target objects were changed for the recognition test, the objects contiguous to them remained unchanged. These decisions were made prior to scoring the data.

[9] There were two exceptions to this rule, both motivated by a desire to prevent the pictures from becoming too cluttered with unusual objects. In the city picture, a mailbox was switched with an old-fashioned standing radio, and in the farm picture, a factory chimney stack was switched with a silo. Both of these cases, though they involved items of differing likelihoods, were counted as rearrangements of low probability objects.

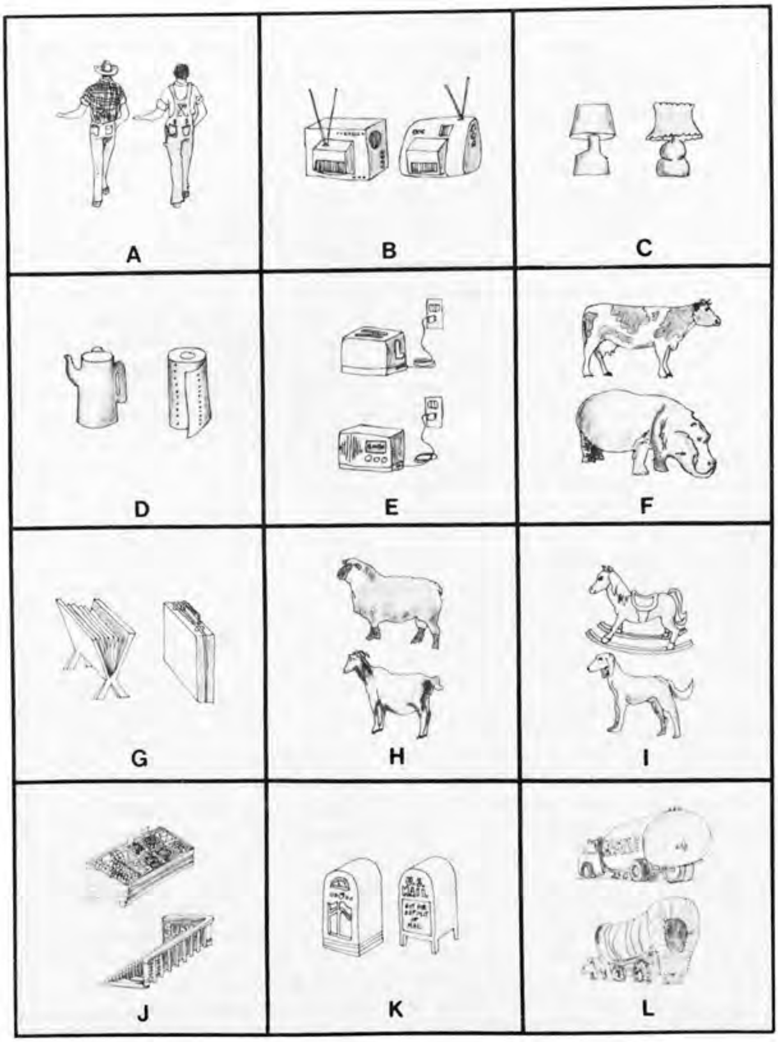


Figure 1. Examples of objects used in the token and type transformations. (Figure 1 was made by rephotographing the original drawings of the objects, printing them so that they would all conform to the same general dimensions, and then redrawing them from these new prints. In order to facilitate black and white photographic reproduction, the shading of the redrawn objects is not as subtle as in the original slides, which were photographed in color. A—token-high, Farm; B—token-medium, Living room; C—token-low, Kitchen; D—type-high to high, Kitchen; E—type-high to medium, Kitchen; F—type-high to low, Farm; G—type-medium to high, Office; H—type-medium to medium, Farm; I—type-medium to low, Kindergarten; J—type-low to high, City; K—type-low to medium, City; L—type-low to low, Farm.)

basically impossible to use visual details to distinguish among objects of differing likelihoods.

To ensure that there were no systematic size differences among the objects used for the 18 distractors of each picture, the area of each of the target objects that were to be used in the recognition test was obtained and expressed as a proportion of the total area of its picture. These data were subjected to a Probability × Transformation × Picture analysis of variance. Neither the main effects nor the interaction even approached significance; the largest F ratio obtained was for the interaction, $F(10, 50) = 1.32$.

Design

Each subject viewed all six target pictures. The picture presentation order was constant throughout the recognition test for a particular subject (i.e., every *n*th picture was of the same scene), and was the same as the order in which he or she had seen the targets; however, the probability that a particular instance of a scene was either a target or distractor version was .5. In addition, the picture presentation order was counterbalanced across 36 subjects so that every picture preceded and followed every other picture exactly once, yielding six subjects in each of six picture presentation order conditions.

The recognition test consisted of 36 targets and 36 distractors. A particular subject saw six distractors per picture, one for each of the transformations (token, rearrangement, deletion, type-high, type-medium, and type-low). For each subject, two of the transformed objects from each picture were of high rated probability, two were medium, and two were low. Thus, across the six pictures, each subject contributed two data points to the 18 different kinds of distractors. In addition, on the distractor trials for a single subject, the six types of transformations each occurred once every six trials, counterbalanced so that across pictures and blocks, each transformation preceded and followed each other transformation exactly once for every subject.

The 18 types of distractors for a particular picture were represented across subjects and presentation orders as follows: The six distractors seen by any subject for a particular picture depended entirely upon where that picture was situated in the presentation order, although there was always one distractor of each transformation type. For example, the distractors for whichever was the first picture in the sequence were always delete-high, token-low, type-low to low, type-high to high, type-medium to medium, and rearrange-medium. There were 36 distractor trials for each subject, divisible into six blocks of six trials each, one trial per block on each picture. The order of the blocks was cycled so that the six different subjects within each picture presentation order received a different cycle of blocks, and, consequently, of transformations. This manipulation made it possible to look at the data by halves, thirds, etc.

In summary, there were six possible transformations made to objects in three probability categories, making 18 possible distractor types for each picture. Across pictures, each of 36 subjects could make from zero to two correct rejections per distractor type; and across subjects, each of six pictures could have from 0 to 12 different subjects correctly rejecting each distractor type. Across both subjects and pictures, then, each distractor type could yield a maximum of 72 correct rejections, and each transformation and probability category could yield a maximum of 216 and 432 rejections, respectively.

Procedure

Twenty undergraduates participated in the experiment to fulfill a course requirement, and 16 additional subjects received $2.00 for their participation. Upon arriving, subjects were told that they were participating in an experiment that would test their memory for pictures, that they would be seeing six pictures for 30 sec each, that their eye movements would be recorded, and that they would have to later be able to distinguish between the original pictures and new pictures in which, for example, only a small detail on one object would be different.

Each subject was seated in front of a rear-projection screen and a Whittaker eye movement monitor, which is a device that determines the point of fixation by measuring the center of the pupil with respect to the center of the corneal reflection. A TV camera, pointed at the rear-projection screen, was interfaced with the eye monitor, so that the set of cross hairs representing the subject's current point of fixation was superimposed on the scene that was being viewed by the subject, and was simultaneously projected to a TV monitor in front of the experimenter and to a Sony videotape recorder. The output of a second TV camera, pointed at the subject's left eye, enabled the experimenter to view the eye as she adjusted the pupil and corneal reflection delimiters.

The subject was asked to place his or her head on a chin rest, find a comfortable position, and remain still until the calibration procedure was finished and all of the pictures had been shown. The calibration procedure consisted of projecting onto the screen a slide that had a cross in its center and a cross at the centers of each of its sides. The subject looked alternately at each cross, while the experimenter ensured that the cross hairs were located in the center of the cross being fixated, and adjusted the pupil and corneal reflection delimiters until they remained stable while the subject looked around the screen. After the eye monitor was calibrated, the experimenter repeated the instruction to look carefully at each picture and try to remember as much as possible, and then proceeded to show the pictures for 30 sec each. The topic of each picture was announced in advance of the slide. The pictures subtended a horizontal visual angle of approximately 30°, and a vertical angle of approximately 20°.

After the subject saw all of the pictures, he or she was instructed about the recognition test in as much detail as possible, with the exception of the likelihood manipulation. The subjects were told that half of the pictures would differ from the originals by a change to only one object, and that once an object had been changed, it would go back to its original version and never be changed again. Then the different transformations were described in detail, using examples from a place that was different from the six target places.

Before beginning the recognition test, the subject was instructed about the operation of the response keys, which were two blocks of wood each having two metal strips across them. The subject placed the heels of his or her hands on one of the strips, kept both index fingers raised, and responded by touching either finger to the second metal strip.

Two signs, labeled "old" and "new," were taped in front of the left and right wooden blocks, respectively. The subject was given six practice trials in the use of the response keys; the experimenter displayed slides of single letters and the subject was told to respond "old" to capital A and "new" to any other letter. The subject was further encouraged to look at the signs before responding, if necessary. After this practice, the recognition instructions were reiterated briefly, and the recognition trials proceeded as follows: The experimenter pushed a button; 2 sec later the slide for that trial was exposed and a clock was simultaneously started. The subject stopped the clock by touching either index finger to the second metal plate on the appropriate response key, and the experimenter recorded the response and the time for that trial. No feedback was given. The entire procedure took between 45 and 60 min.

Results and Discussion

Eye Movements During Original Viewing

Scoring the data. Of the 170 target areas scored, 151 were areas delineated by the perimeters of single objects, and 19 areas included several objects within their perimeters that had similarly rated probabilities (see Appendix A). While scoring the data, it was possible to discern separate fixations to the different objects in these multiobject areas. Therefore, before any analyses were carried out, each of the eye movement measures for the multiobject areas was adjusted for each subject by dividing the obtained time or number of looks by the number of objects in the area, in order to render them comparable to the 151 single-object areas. The rated probabilities for these areas were similarly adjusted by calculating the average of the ratings of all the objects in each area.

Every subject produced approximately 3 min of eye movement recordings, which consisted of a set of cross hairs representing his or her point of fixation in real time, superimposed on each of the target pictures. The video recorder made a mark on the videotape every 1/60 of a second, which provided a means of quantifying the eye movement data. There were approximately 10,800 frames of videotape per subject; the four dependent time measures to be discussed represent the amount of time in milliseconds per fixation per object, obtained by counting the number of frames that occurred while the cross hairs were superimposed on or contiguous to (approximately 1/8 in. [3.2 mm] away on the video screen) a given object, dividing the result by 60, multiplying by 1,000, and truncating any remainder. Thus, each frame counted added approximately 17 msec to the total for each of the fixation duration measures.

Relationship between rated probability and fixation durations. The mean durations of the first and second fixations and of the average of the third through *n*th fixations to each object were obtained across the subjects who had actually looked at each particular object for the ordinal fixation number under consideration. These means were then correlated with the rated probability of the objects. Excluding the city and kindergarten scenes, which will be discussed below, between 27.7% and 52.4% of the variation in the amount of time it took to encode an object for the first time was accounted for by its rated probability. In contrast, rated probability accounted for between 14.4% and 38.6% of the time spent on second fixations, and between 5.6% and 24.1% of the time spent on subsequent fixations. Clearly, the a priori likelihood of an object had its most potent effect on the duration of the first fixation to it.

The city and kindergarten pictures followed this trend, but did not generally yield reliable correlations. An inspection of these pictures led to a plausible ad hoc hypothesis: The city and kindergarten both included many people and objects with writing on them (e.g., a street sign or a blackboard). In fact, across all of the pictures, there were 25 areas that depicted people or objects with writing, and 18 of these were in the city and kindergarten. Although people are ubiquitous, and thus deserve high probalility ratings, they may also be reasonable "objects" for a human observer to take note of while viewing a scene. Similarly, reading may involve time parameters or procedures different from those required by simple object identification.

To determine whether this admittedly post hoc explanation of the differences between the city and kindergarten and the other four pictures had any validity, the first fixation durations were again correlated with the

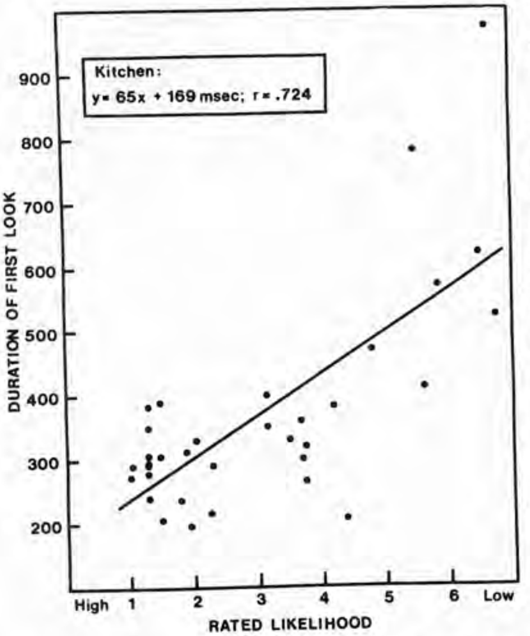


Figure 2. Duration of the first fixation to the objects in the kitchen scene as a function of their rated likelihoods.

rated probability of the objects, but all of the people and objects with writing were excluded. Table 2 shows both the original and new correlations, and Figure 2 shows the data for the kitchen scene, which was entirely free of people and writing. It can be seen that excluding these areas generally yielded better correlations for all of the pictures, and that those for the city and kindergarten improved dramatically and in fact became reliable, despite fewer degrees of freedom. As a result of this, all subsequent analyses were performed both with and without these areas.

Table 2 also shows the regression equations for each picture, as well as the equation obtained across pictures, using only the 145 areas without people or writing. The variation attributable to the regression was reliable, $F(1,\ 143) = 108.35$, $MS_e = 16{,}649$; and the 95% confidence interval for the slope indicated that for each unit that objects decreased in likelihood, the durations of the first fixations were increased by between 56 and 81 msec. In addition, although there was a relationship between rated likelihood and the total amount of time spent on the target objects, $r = .473$, $r^2 = .224$, $df = 143$, that relationship was relatively insubstantial when the variance attributable to the duration of the first fixations was partialed out, $r = .168$, $r^2 = .028$, $df = 142$.

Since conclusions drawn from the correlational analyses might be objected to on the grounds that there were a disproportionate number of high probability objects in each of the pictures, the duration measures for each subject and picture were collapsed into three probability categories; the cutoff points for the categories never overlapped within a picture, and across pictures ranged from 1.020 to 2.863 for high probability objects, 2.647 to 4.776 for medium probability objects, and 4.843 to 6.745 for low probability objects. The denominators used to calculate each mean were the numbers of high, medium, and low probability objects within each pic-

Table 2
Relationship Between an Object's Rated Probability and the Duration of the First Fixation to It

Picture	All areas r	All areas n	Excluding writing and people r	Excluding writing and people n	Regression equation
City	.321	26	.598	13	$y' = 54x + 265$ msec
Farm	.657	26	.586	23	$y' = 51x + 232$ msec
Kindergarten	.228	25	.748	20	$y' = 73x + 103$ msec
Kitchen	.724	34	.724	34	$y' = 65x + 169$ msec
Living room	.664	25	.701	24	$y' = 70x + 109$ msec
Office	.526	34	.716	31	$y' = 110x + 47$ msec
Across all pictures			.657	145	$y' = 69x + 156$ msec[a]

[a] SE slope = 6.58 msec.

ture that had been looked at by each subject. Thus, each subject provided three scores per picture per fixation number, and the means discussed below and displayed in Figure 3 represent the fixation durations for a "typical" object in each probability category. A second set of means was calculated for each subject in a manner identical to the first, excluding the people and objects with writing.

Both sets of fixation durations were analyzed in a four-factor repeated measures analysis of variance; the factors were probability (high, medium, or low); fixation number (first, second, or third–*n*th), picture (1–6), and subjects (1–36). Both pictures and subjects were simultaneously treated as random variables, so that the effects of fixation number and probability, as well as their interaction, were tested with quasi F ratios (F') (Myers, 1972, pp. 308–309). Since the two analyses yielded identical results, the means and error terms reported are from the analysis that excluded the people and writing.

The main effect of fixation number was reliable, $F'(2, 16) = 14.91$, $MS_e = 86{,}611$, with later fixations being shorter on the average than the first two (the means for the first, second, and later fixations were 432, 447, and 363 msec). However, this main effect was largely due to the dramatic decline in the amount of time subjects spent looking at low probability objects, as indicated by the Fixation Number × Probability interaction, $F'(4, 22) = 15.74$, $MS_e = 73{,}229$. Figure 3 shows the data for this interaction.

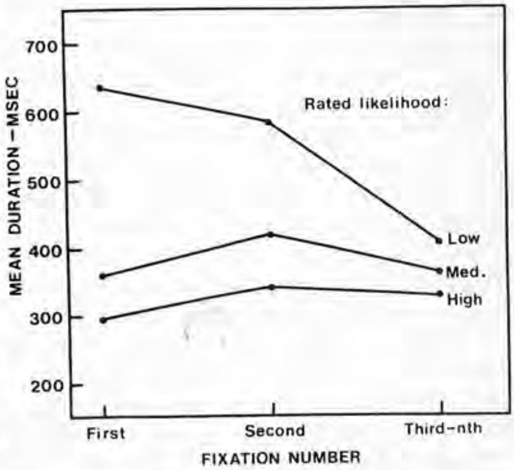


Figure 3. Relationship between rated likelihood and the duration of the first and second fixations and the average duration of the third through *n*th fixations to an object (people and writing excluded).

Newman-Keuls tests showed that subjects spent an equal amount of time on all three fixations to the high and medium probability objects, while there was a significant decline in looking time after the second fixation to the low probability objects. In addition, low probability objects were looked at reliably longer than high probability objects at all three fixations, but the time difference decreased from 342 msec during the first fixation to only 78 msec during the later fixations. Third, the high probability objects were looked at for reliably less time than the medium probability objects during both the first and second fixations, although the differences were relatively small (65 and 79 msec, respectively). Finally, it is of interest to note that the mean durations of the first, second, and third–*n*th fixations to the people in the pictures were 324, 328, and 225 msec, respectively, and the means for the written objects were 807, 584, and 392 msec, respectively.

In addition to the effects described above, there were reliable main effects of probability, $F'(2, 12) = 21.08$, $MS_e = 403{,}939$, and pictures, $F(5, 175) = 9.44$, $MS_e = 47{,}418$, and reliable interactions between pictures and probability, $F(10, 350) = 7.33$, $MS_e = 50{,}360$, subjects and probability, $F(70, 350) = 1.69$, $MS_e = 50{,}360$, and subjects and fixation number, $F(70, 350) = 1.47$, $MS_e = 55{,}963$. The latter two interactions, tested as they were against a great many degrees of freedom, are not too worrisome; they simply mean that not all subjects had the same magnitude of time differences between early and late fixations, or between high and low probability objects. Potentially more troublesome is the Picture × Probability interaction; however, Newman-Keuls tests showed that low probability objects were looked at reliably longer than high probability objects for all six pictures, though the differences between these two categories ranged from 88 msec for the city to 328 msec for the office.

There are several major conclusions to be drawn from the data thus far. First, both the correlations and the analyses of variance support the view that when a person knows in advance the general context of what he or she is going to see, this information allows global memory structures such as frames to be used to detect objects expected to be in that context, resulting in relatively short first fixations to them. Second, identification of low probability objects appears to require a more complete and presumably more data-driven analysis of their visual details; first fixations to low probability objects were twice as long, on the average, as first fixations to high probability objects. Third, the differences in fixation durations between high and low probability objects decreased during the later fixations, probably because subjects were trying to take in as many details from as many different objects as they could, in anticipation of the recognition test. Fourth, the medium probability objects required slightly but reliably longer first fixations than the high probability objects did, which suggests that their identification may have involved more of an interaction between top-down and bottom-up procedures. Finally, rated probability did not appreciably influence the total time spent viewing the objects when the variance attributable to the first fixation durations was partialed out.

Relationship between rated probability and the number of different fixations. During the 30 sec of original viewing, subjects were able to fixate an average of 81.7% of the objects in each of the pictures. The number of different fixations to any particular object, however, was not reliably predicted by that object's likelihood, contrary to the expectations expressed by Loftus (1976). For the six pictures, the largest positive correlation between probability and the number of fixations was .284; at best, then, probability accounted for only 8.1% of the variation in the number of different fixations made to the objects. In addition, when the average number of fixations to high, medium, and low probability objects was subjected to a Probability × Pictures × Subjects analysis of variance, in which both subjects and pictures were treated as random variables, the only reliable effects were for pictures, $F(5, 175) = 25.10$, $MS_e = .376$, and the Picture × Probability interaction, $F(10, 350) = 9.93$, $MS_e = .341$. The interaction was unsystematic; the mean number of different fixations to any particular object was 2.25, and the difference between the number of fixations made to low versus high probability objects ranged from −.61 for the farm to .65 for the kindergarten. Thus there was essentially no relationship between the number of different fixations to an object and that object's rated probability.

Although Loftus and Mackworth (1978) found that their subjects looked at unexpected objects more often than they looked at expected objects, there are several reasons for believing that their data do not compromise the general conclusion reached above. First, their subjects saw 78 relatively simple line drawings, half of which contained only one unusual object. It seems possible that the subjects could therefore have become aware of the implicit demand characteristics of this experimental manipulation, so that they did in fact begin to return to the unexpected objects more frequently. More important, subjects in the present experiment viewed the pictures for 30 sec each and their instructions emphasized the fact that the recognition test would consist of changes to specific objects; therefore they may have purposely distributed their fixations more evenly across all the objects in the pictures. Loftus and Mackworth's (1978) subjects, on the other hand, had only 4 sec to look at each picture, and were told to look at the pictures *as if* they were going to have a recognition test on each picture as a whole. In this case, the single unusual object becomes the only item in the difference structure for each picture, and might be returned to more frequently as a result. Thus, whether or not the number of different fixations to an object is a function of that object's contextual likeliness may be a function of the number of objects in the scene, the amount of time subjects have to look at the scene, and their expectations regarding what they will have to do subsequently.

Recognition Performance

Overall speed and accuracy. The total number of hits and correct rejections to each picture was tallied for each subject, and these data were used in a three-factor analysis of variance in which presentation order (1–6) was between subjects, and picture (1–6) and version (target or distractor) were within subjects. Both subjects and pictures were treated as random variables.

Subjects made more correct responses to the targets ($M = .73$ for hits) than to the distractors ($M = .59$ for correct rejections), $F'(1, 5) = 7.36$, $MS_e = 10.64$. In addition, there was a reliable interaction between pictures and version, $F(5, 25) = 5.49$, $MS_e = 1.791$. More targets than distractors were correctly recognized for every picture except the kindergarten; that the other five pictures had a reliable difference between versions was confirmed by finding a reliable difference for the farm scene, $F(1, 25) = 5.21$, $MS_e = 1.791$, which had the smallest difference.

The mean reaction times for each subject to make hits, correct rejections, misses, and false alarms were calculated across pictures and subjected to a Subjects × Response Type within presentation order analysis of variance. Only the main effect of response type was reliable, $F(3, 90) = 28.91$, $MS_e = 11.23$. Newman-Keuls tests showed that subjects were fastest on correct rejections (9.76 sec), next fastest on misses (13.39 sec), and slowest on hits (16.02 sec) and false alarms (16.18 sec), which did not differ from each other. It should be noted that the time taken to make hits and false alarms suggests that subjects looked at most of the objects at least once before deciding to call a picture old. That is, since subjects made an average of 2.25 fixations to each of approximately 82% of the target objects within the 30-sec time period of original viewing, the fact that they spent approximately half that time to accept a picture as old suggests that they first checked most of the objects in it. The reaction time data were not subjected to any further analyses, primarily because accuracy was stressed far more than speed was.

Distribution of correct rejections. In order to strengthen the argument that the proportions of correct rejections reflect representational differences between expected and unexpected objects, and are not the result of encoding failures, it was desirable to ascertain that the appropriate target objects had in fact been encoded. Accordingly, each of the Probability × Transformation analyses to be described below was performed twice—once with pictures random and once with subjects random; and whenever pictures were treated as the random variable, the dependent measure was the proportion of times a particular distractor was correctly rejected, *provided that the subjects who were tested on that distractor had eye movement records showing that they had looked at the appropriate target object at least once during original viewing.* This measure will be designated "P(CR|F)." There were missing cells when the data were conditionalized on whether a subject had originally looked at either of the particular two objects that were tested for each of the 18 distractor types; so when subjects were treated as random in the analyses below, the dependent measure was the unconditional probability of correct rejections for each distractor type. Both sets of analyses, as well as comparisons among the means, yielded identical results. Moreover, all of the effects to be discussed were tested with either min F' ratios (Clark, 1973) or combined error terms from both analyses, in order to be able to simultaneously generalize to subjects and materials.

To summarize, the data to be discussed reflect the number of times a particular distractor type was correctly recognized as being new, counting only those cases in which the target object was known to have been looked at and depicted neither a person nor an object with writing.[10]

[10] Since the 16 distractors that involved target objects that had writing or depicted people were to be excluded, the following strategy was adopted for testing the effects of probability, transformation type, and their interaction: The P(CR|F) for

Effect of transformation type. When the probability of a transformed object is ignored, the distribution of correct rejections among the different transformations may be viewed as an indication of the memorability of different kinds of pictorial information. The main effect of transformation was reliable, $F(5, 150) = 12.16$, $MS_e = .112$, for subjects; $F(5, 20) = 6.08$, $MS_e = .036$, for pictures; min $F'(5, 44) = 4.05$. The mean $P(CR|F)$s for the token, rearrangement, deletion, type-high, type-medium, and type-low transformations were .401, .503, .536, .620, .715, and .702, respectively, and these proportions were based on *n*s of 167, 171, 153, 129, 144, and 171.

The general pattern of transformation difficulty was not the same for every picture, although token changes were generally the most difficult and type-low changes were generally the easiest. Newman-Keuls tests showed that the three type transformations were the easiest to recognize and did not differ from each other. Token changes and rearrangements were the most difficult to recognize and also did not differ from each other. The deletions and all three of the type transformations were easier to notice than token transformations; in addition, the type-low and type-medium transformations were reliably easier than both deletions and rearrangements.

Without taking probability information into account, the data suggest that the descriptive information measured by token changes was either not stored or else was quickly forgotten; that there was something akin to an object inventory in the representation, at least to the extent that type changes were well noticed; and that spatial composition and spatial location information (tested with deletions and rearrangements) was also relatively poorly remembered.

Effect of rated probability. When type of transformation is ignored, the distribution of correct rejections reflects the overall memorability of objects that had either a high, medium, or low rated likelihood of being seen in the pictures to begin with. The effect of probability was reliable and was the most robust effect in the data, $F(2, 60) = 43.32$, $MS_e = .097$, for subjects; $F(2, 8) = 24.38$, $MS_e = .032$, for pictures; min $F'(2, 19) = 15.83$. Indeed, for 32 of the 36 subjects and for all six pictures, more distractors were correctly rejected when the change involved a low probability item than when it involved a high probability item. The low-high difference ranged from 22.8% for the kindergarten to 42.5% for the kitchen. Across pictures, there were 32.6% more correct rejections to low than to high probability distractors, min $F'(1, 18) = 33.10$, and 23.1% more correct rejections to low than to medium probability distractors, min $F'(1, 18) = 16.62$. The 9.5% difference in accuracy between medium and high probability distractors was reliable only for subjects, $F(1, 60) = 8.37$, $MS_e = .097$.

The robustness of the probability effect is further indicated by the fact that it did not diminish over blocks, even though the test trials could potentially have been used as further study trials. The conditional probabilities of rejecting high, medium, and low probability distractor changes on the first block of trials, given that the changed objects had been looked at, were .468, .558, and .705, respectively; by the last block of distractor trials, these probabilities were .396, .550, and .754. Moreover, it will be seen in a subse-

---

each of the 18 distractor types, as well as each probability category and transformation, were calculated separately, across pictures and excluding cases that involved people or writing. These proportions were taken as the best estimates of the population proportions for the particular distractor type under consideration, and were the only proportions used in any of the comparisons. However, *all* of the observations were included in the two analyses of variance in which the factors were probability and transformation type, and either pictures or subjects were random. The error mean squares from these analyses were taken as the most conservative estimates of the population variances for each effect; and in testing the effects from the analyses, as well as in further comparisons among the means, the degrees of freedom were adjusted to reflect the loss of the 16 distractors involving writing or people (i.e., approximately one-sixth of the data) by assuming that there were five instead of six pictures, or 30 instead of 36 subjects. Min F' values were calculated using the F ratios from both analyses, with the adjusted degrees of freedom.

quent section that the effect of rated probability on recognition accuracy was not a function of the total time spent viewing the target objects.

Finally, it should be noted that the conditional probability of rejecting a changed written object was .673, while the proportion for a "people distractor" was .667. The recognition accuracy for these items was thus similar to the accuracy of the low probability items, when in fact only 4 of 16 such cases were actually from the low probability category. This discrepancy once again supports the exclusion of these distractors from the estimates of the population proportions.

Transformation × Probability interaction. On the whole, using conditional proportions as the dependent measure of recognition, it appears that rated probability and transformation type had independent effects on recognition accuracy, since the Transformation × Probability interaction was not reliable for subjects or for pictures. Indeed, the low–high difference was reliable for the token, deletion, and type-medium transformations, min $F'(1, 66) = 4.56$, 8.01, and 6.57, respectively, and approached reliability for the rearrangement and type-high transformations, min $F'(1, 66) = 2.94$ and 3.13, $p < .10$. The difference was not reliable for the type-low transformation, however, min $F' = 1$. Table 3 and Figure 4 show the data for the interaction.

It should be emphasized that the effect of rated probability was almost insensitive to the kind of information being tested; for example, token changes made to low probability objects were noticed 14.5% more often than categorical replacements of high probability targets with new high probability distractors. Only one proportion was discrepant from what was expected from a frame theory approach: the low probability target objects that were replaced by medium probability objects should have been rejected with a higher frequency than that which was observed. Apart from this case, however, the data are extremely clear and are supportive of the theory.

Lest the reader believe that (a) excluding the 16 distractors involving people and writing has biased the data, or (b) using conditional rather than unconditional proportions has biased the data, or (c) that d' analyses would have been a more appropriate treatment of the data, it should be pointed out that since the analyses of variance included all of the distractors, using the error mean squares from these analyses increased the variance against which the effects and comparisons among means were tested. In addition, Table 4 shows the values of d' that were obtained for each of the distractors by using the probability of a hit and both the conditional (upon a fixation) and unconditional probabilities of a false alarm for each of the

Table 3
Proportion of Correct Rejections, Given an Original Target Object Fixation, as a Function of Rated Likelihood and Transformation Type (People and Writing Excluded)

Transformation	Rated likelihood		
	High	Medium	Low
Token	.250 (60)	.386 (57)	.600 (50)
Rearrange	.391 (46)	.433 (67)	.672 (58)
Delete	.265 (34)	.500 (60)	.729 (59)
Type-high	.455 (44)	.647 (34)	.745 (51)
Type-medium	.469 (32)	.653 (49)	.889 (63)
Type-low	.696 (56)	.585 (65)	.860 (50)
Across transformations	.426 (272)	.521 (332)	.752 (331)

Note. Across pictures, each subject could contribute a maximum of 2 correct rejections to each of the above cells. Consequently, across the 36 subjects, each distractor could have a maximum of 72 correct rejections. Numbers in parentheses were used in the denominators of the proportions, and are the total number of times, out of the 72 possibilities, that the appropriate target object was fixated at least once.

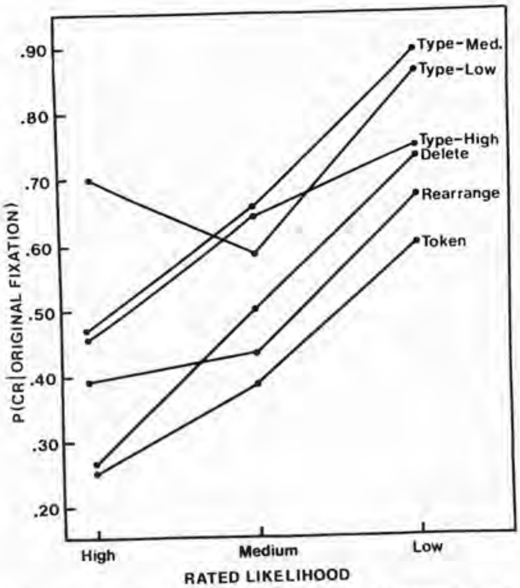


Figure 4. Conditional probabilities of correctly rejecting each transformation type as a function of rated likelihood (people and writing excluded). (Names of the type transformations refer to the rated likelihoods of the original target objects; for example, reading from left to right, the line labeled "Type-High" depicts the proportion of times subjects noticed when an original high probability target object was replaced by another object with either a high, medium, or low rated likelihood.)

individual pictures, and then taking an average across the pictures. Analyses of variance on these data supported all of the major conclusions reached above. Thus, by limiting the discussion to objects which are certain to have been encoded, and by omitting distractors that involved people or writing, stronger statements regarding differences between the memory representations of expected and unexpected objects, and more accurate statements regarding the magnitude of those differences, can be made.

Relationship Between Rated Likelihood, Total Viewing Time, and Distribution of Correct Rejections

The conclusions reached above would be less interesting if the differences in proportion of correct rejections were entirely due to differences in the total amount of time that subjects had spent viewing the target objects. Remember, however, that there was virtually no relationship between total viewing time and rated likelihood when the variance attributable to first fixation durations was partialed out. We therefore expected that the relationship between likelihood and proportion of correct rejections would be preserved, even when viewing time was controlled. In order to be certain of this, partial correlations between rated likelihood and the conditional proportion of correct rejections were calculated for each transformation, controlling for any variability that could be accounted for by either total viewing time, the duration of the first fixations, or the combination of these two variables. Then the relationships between either total time and the P(CR|F)s or the duration of the first fixations and the P(CR|F)s were obtained, controlling for the variability accounted for by rated likelihood. It can be seen in Table 5 that with the exception of rearrangements, rated likelihood accounted for more of the variance in correct rejections than did either total viewing time, the first fixation durations, or their combination. These data indicate that in general, a "differential study time" hypothesis cannot be used to explain the recognition data.

General Discussion

The major findings of the present experiment may be summarized very simply. Subjects who anticipated a difficult recognition test viewed a few rather complex pictures for a relatively long period of time, having been told the topic of each. The duration of their first fixations to the objects in the pictures was a function of the a priori likelihood that these objects would be found in the particular settings in which they were depicted. These data were used to validate the hypothesis that expected and unexpected objects are distinguishable on the basis of whether predominantly automatic or predominantly interactive procedures are used for their identification. Further, the recognition data showed that even when original viewing time was controlled, and the data were conditionalized on fixation of the target objects, subjects confused what they had actually seen with new objects they could plausibly have seen. They almost never noticed miss-

Table 4
Conditional and Unconditional Values of d′ *Across Pictures, as a Function of Rated Likelihood and Transformation Type (People and Writing Excluded)*

	Rated likelihood					
	High		Medium		Low	
Transformation	UC	C	UC	C	UC	C
Token	−.15	−.14	.29	.11	.84	.96
Rearrange	.36	.41	.44	.54	.84	1.12
Delete	.09	−.03	.67	.63	1.37	1.43
Type-high	.48	.47	1.03	1.02	1.49	1.42
Type-medium	.62	.46	.95	1.00	1.82	1.88
Type-low	1.32	1.33	.85	.85	1.91	1.86
M	.45	.42	.71	.69	1.38	1.45

Note. Analyses of variance on both the unconditional (UC) and conditional (C) d' data yielded reliable main effects of transformation, $F(5, 20) = 7.31$, $MS_e = .341$, and $F(5, 20) = 5.38$, $MS_e = .415$, respectively; and of probability, $F(2, 8) = 21.09$, $MS_e = .291$, and $F(2, 8) = 25.59$, $MS_e = .319$, respectively.

ing, new, or descriptively changed expected objects, while they almost always noticed deletions or replacements of unexpected objects. This result was obtained despite the fact that there were no significant size differences among the objects used for the 18 distractor types. In addition, this result was obtained despite the fact that the instructions about the recognition test could have led the subjects to treat the scenes as collections of individual, unrelated objects. Thus, recognition accuracy was largely a function of having varied the conceptual similarity between the targets and distractors by transforming objects of differing thematic importance. The failures to notice new or missing expected objects in the present experiment are similar to the paraphrase confusions more commonly obtained when people are asked to remember prose; and they validate the hypothesis that expected objects are not normally episodically tagged. Subjects clearly did not have "copies" of the target pictures in their heads, nor did they have anything resembling an accurate verbal description. Instead, they seemed to have represented the important variations from six particular themes (i.e., a set of difference structures that represented the differences between the global frames and the actual pictures).

The theoretical distinction between feature detection and feature analysis, and the necessity for both, is a reasonable and parsimonious heuristic to use for explaining the observed differences in the way that expected and unexpected objects were identified, as well as the differences in the amount and type of information that was stored or retained about them. In particular, it should be clear that not all things can be comprehended in a predominantly top-down fashion, since if this were true, it would be difficult to understand why first fixations to unexpected objects were almost twice as long as first fixations to expected objects. Similarly, if we assume that everything was comprehended in a predominantly interactive mode, the fixation data could be explained by the fact that the meaning of unexpected objects simply took longer to "construct"; but it would be difficult to explain the recognition data without postulating mechanisms in addition to perceptual mechanisms that could account for such memory differences. That this is particularly cumbersome to do, in the absence of any well-articulated theory of context effects, is indicated by the fact that many theorists have been equivocal on the relationship between knowledge, comprehension, and memory. For example, it has sometimes been hypothesized that events that are difficult to comprehend will also be difficult to remember (Bransford & Johnson, 1972). Indeed, ambiguous paragraphs about washing clothes or flying kites or bizarre serenades were in fact incomprehensible and poorly remembered

Table 5
Relationship Between Rated Likelihood, Proportion of Correct Rejections Given an Original Fixation P(CR|F), and Viewing Time (People and Writing Excluded)

Correlation	Transformation[a] Token (n = 16)	Rearrange (n = 15)	Delete (n = 16)	Type (n = 45)
Total time × P(CR|F)·rating	−.038	.327	−.310	064
First fixation × P(CR|F)·rating	−.542	.372	−.628	−.425
Rating × P(CR|F)·total time	.591	.344	.692	.381
Rating × P(CR|F)·first fixation	.754	.134	.814	.568
Rating × P(CR|F)·total time and first fixation	.720	.028	.811	.538

[a] Objects in the token, rearrangement, and deletion transformations have the same rated likelihoods in both the target and distractor versions of the pictures. For the type transformations, the ratings used in the correlations were the sum of the ratings given to the original objects and the objects that replaced them, since the likelihoods of both the original and the new objects had been hypothesized to exert an influence on recognition accuracy.

when subjects were unaware of what the paragraphs were about. In contrast to those results, if the fixation durations in the present study are indicative of comprehensibility, then unexpected objects were at least initially *less* comprehensible than expected objects; however, they were clearly *more* memorable, both episodically (e.g., deletions) and in detail (e.g., token changes). The present notion that unexpected events will be well remembered because they are generally the sole recipients of relatively interactive processes and because they are represented in a separate difference structure thus represents a considerable modification of one side of the constructive approach.

There have been, however, instances in which theorists have argued for exactly the point of view just described. For example, Bransford and McCarrell (1974, pp. 208–209) claim that sentences that are difficult to comprehend will require more elaboration than those that are easy, and hence the former should be remembered better than the latter. In the present view, this sort of theoretical flip-flop derives from a failure to distinguish between internal and external contexts, their uses in comprehension, and the different consequences of predominantly top-down and predominantly interactive perception for event memory. In addition, as already noted, sentence or text comprehension may involve different procedures than object recognition, and it may therefore be dangerous to generalize from such experiments to perception as a whole. Further, even in experiments which use text, some of the results cited as evidence for highly interactive processes in comprehension may not have been due to comprehensibility per se. For example, in one study (Bransford & Johnson, 1973), subjects did not remember an ambiguous sentence in the context of a paragraph whose title disposed them to interpret the paragraph as being about a parade, but they did remember it when they thought the paragraph pertained to a Martian landing. Rather than being a difference in the degree to which the sentence was comprehensible in the different contexts, however, this memory difference may only mean that the sentence in question was thematically irrelevant to the parade paragraph, and so was not included on an "event list" (Schank & Abelson, 1975). In addition, the sentence was not unusual enough, in terms of the events described in the paragraph, to warrant its inclusion on a weird list or in a difference structure. In short, a major shortcoming of an approach in which perception is normally presumed to be predominantly interactive is its inability to go much beyond saying *that* knowledge or context may be related to comprehensibility and memory.

The present approach places a great deal of stress on the primacy of abstract, invariant information for the apprehension of meaning. One of its advantages is that it allows the a priori specification of the types of objects, actions, or events within a given combination of internal and external contexts whose comprehension will be either relatively automatic or relatively interactive. In addition, it allows one to specify the memorial consequences for those objects, actions, or events of having been perceived via one or the other means. Adopting the validity of the distinction between feature detection and feature analysis, and the consequences that this distinction has for event memory, thus allows a fairly precise specification of the conditions under which two events might be indistinguishable from each other, and hence makes a step toward resolving some aspects of the stimulus equivalence phenomenon.

Remembering the theme or gist of an event is tantamount to remembering the types of things which were *most likely* to have occurred; the undetected type-high to high changes and deletions attest to the fact that commonplace (i.e., thematic) episodic information was not especially noted or tagged. The distinction between feature detection and feature analysis thus allows the distinction between semantic and episodic memory to become simultaneously more precise and more fluid: The episodic information that will be remembered about an event is the difference between that event and its prototypical frame representation in memory. Thus, remembered episodic information will tend to be those objects, actions, features, and events which distinguish a particular situation from others of the same general class. That this is so has much adaptive validity. In general, we do not need to notice or remember the parts of an event that we expect or that will not be useful in the future. In a recent example of this principle, Keenan, MacWhinney, and Mayhew (1977) showed that people engaging in spoken discourse in a natural setting have remarkable verbatim memory for some aspects of it, even as long as 48 hours afterwards. The portions of the discourse that were especially well remembered were those having high "interactional content," that is, those portions of the conversation which could be of some future use to the participants. These results are entirely in accord with the present point of view.

The consequences of predominantly top-down perception for event memory are thus twofold: First, automatic perception of expected objects or events frees resources that can then be used to process the unexpected or novel; these more interactively perceived objects will have representations that are separate, and at least initially more detailed, than the representations of expected objects. Second, paraphrase confusions can be predicted for expected objects, again as an accident of the way in which they were perceived. We have seen, however, that the descriptive details of even the unexpected objects were not perfectly remembered within the relatively short time span of the present experiment. In all likelihood, they would have been forgotten had there been a longer delay between original viewing and the recognition test. Thus, the eventual memorial distinction between expected and unexpected objects may be primarily episodic.

Current theorizing about the representation of knowledge is of course meant to include perceptual knowledge, but is nevertheless primarily based upon what we know about comprehension and memory of language. The present experiment points out the usefulness of a frame theory approach to the representation of both perceptually and linguistically derived knowledge. The data from the present experiment are clearly not definitive, however. For example, they may be nonreplicable in situations where target pictures contain only a few objects, or distractors vary the size and shape of the new objects in addition to their rated likelihood, or the original or new expected objects have details falling outside the range specified by default knowledge, or the subjects are not primed, or other sorts of frame violations are made to either the targets or the distractors (e.g., moving a stove into the middle of the air). In addition, it may be that in the present experiment, the presence of unchanged unexpected objects biased subjects to falsely recognize distractors in which expected objects were different. Given the average length of time for a false recognition, however, this seems unlikely. Still, further experiments are needed to test all of these possibilities. It does seem, however, that subjects in the present experiment used their more general world knowledge to perceive the meaning of some of the objects in the target pictures, and that the "unused" part of this knowledge (i.e., arguments for nondisplayed expected objects) became part and parcel of their memory representations. Subjects were therefore able to spend less time looking at expected objects in order to perceive their meaning, but at the cost of becoming confused when new objects that they were required to reject corresponded to their default knowledge.

Finally, it should be noted that adopting a frame theory approach to perception and memory is by no means an endorsement of a computer metaphor for the mind. However, adopting a frame theory does allow us to say that the best paraphrase of a picture preserves the semantic interpretation of the original either by changing only its descriptive or surface structure details, or else by making additions or deletions that allow the new picture to qualify as an instance of the same frame as the original—that is, by making changes that do not require the use of any additional frames in order for the picture to be comprehended. Thus, since two pictures are paraphrases if they instantiate the same frame, it seems likely that one frame is worth 1,000 pictures.

Reference Notes

1. Turvey, M. T. *Perspectives in vision: Conception or perception?* (Status Report on Speech Research No. SR-42/43). Storrs, Conn.: Haskins Laboratories, 1975.
2. Marr, D., & Nishihara, H. *Representation and recognition of the spatial organization of three dimensional shapes* (Memo No. 377). Cambridge: Massachusetts Institute of Technology, Artificial Intelligence Laboratory, 1976.
3. Minsky, M., & Papert, S. *Progress report on artificial intelligence* (Memo No. 252). Cambridge: Massachusetts Institute of Technology, Artificial Intelligence Laboratory, 1972.
4. Anderson, R., Reynolds, R., Schallert, D., & Goetz, E. *Frameworks for comprehending discourse* (Tech. Rep. No. 12). Champaign, Ill.: Laboratory for Cognitive Studies in Education, 1976.
5. van Dijk, T. *Recalling and summarizing complex discourse* (Tech. Rep.). Amsterdam: University of Amsterdam, 1975.
6. Parker, R. *Picture recognition processes as revealed by eye fixations and brief exposures of test pictures.* Paper presented at the Second Annual Interdisciplinary Conference on Psychology, Jackson Hole, Wyoming, 1977.
7. Scinto, L. F. *Relation of eye fixations to old-new information of texts.* Paper presented at the Conference on Psychological Processes and Eye Movements, Monterey, California, 1977.

References

Abelson, R. P. Concepts for representing mundane reality in plans. In D. G. Bobrow & A. Collins (Eds.), *Repersentation and understanding: Studies in cognitive science.* New York: Academic Press, 1975. (a)

Abelson, R. P. Does a story understander need a point of view? In R. C. Schank & B. Nash-Webber (Eds.), *Theoretical issues in natural language understanding.* Arlington, Va.: Tinlap Press, 1975. (b)

Anderson, R., & Pichert, J. Recall of previously unrecallable information following a shift in perspective. *Journal of Verbal Learning and Verbal Behavior*, 1978, *17*, 1–12.

Antes, J. The time course of picture viewing. *Journal of Experimental Psychology*, 1974, *103*, 62–70.

Antes, J. Recognizing and localizing features in brief picture presentations. *Memory & Cognition*, 1977, *5*, 155–161.

Antes, J., & Stone, L. Multidimensional scaling of pictorial informativeness. *Perceptual & Motor Skills*, 1975, *40*, 887–893.

Bahrick, H., Bahrick, P., & Wittlinger, R. Fifty years of memory for names and faces: A cross-sectional approach. *Journal of Experimental Psychology: General*, 1975, *104*, 54–75.

Banks, W., & Flora, J. Semantic and perceptual processes in symbolic comparisons. *Journal of Experimental Psychology: Human Perception and Performance*, 1977, *3*, 278–290.

Barclay, J. R. The role of comprehension in remembering sentences. *Cognitive Psychology*, 1973, *4*, 229–254.

Bartlett, F. C. *Remembering.* London: Cambridge University Press, 1932.

Biederman, I. Perceiving real-world scenes. *Science*, 1972, *177*, 77–79.

Biederman, I., Glass, A., & Stacy, E. Searching for objects in real-world scenes. *Journal of Experimental Psychology*, 1973, *97*, 22–27.

Biederman, I., Rabinowitz, J., Glass, A., & Stacy, E. On the information extracted from a glance at a scene. *Journal of Experimental Psychology*, 1974, *103*, 597–600.

Bishop, P. O. Beginning of form vision and binocular depth discrimination in cortex. In F. O. Schmitt (Ed.), *The neurosciences: Second study program.* New York: Rockfeller University Press, 1970.

Bobrow, D. Dimensions of representation. In D. G. Bobrow & A. Collins (Eds.), *Representation and understanding: Studies in cognitive science.* New York: Academic Press, 1975.

Bobrow, D. G., & Collins, A. (Eds.). *Representation and understanding: Studies in cognitive science.* New York: Academic Press, 1975.

Bobrow, D. G., & Norman, D. A. Some principles of memory schemata. In D. G. Bobrow & A. Collins (Eds.), *Representation and understanding: Studies in cognitive science.* New York: Academic Press, 1975.

Bower, G., & Glass, A. Structural units and the redintegrative power of picture fragments. *Journal of Experimental Psychology: Human Learning and Memory,* 1976, *2,* 456–466.

Bransford, J. D., Barclay, J. R., & Franks, J. J. Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology,* 1972, *3,* 193–209.

Bransford, J. D., & Johnson, M. Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior,* 1972, *11,* 717–726.

Bransford, J. D., & Johnson, M. 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: Some thoughts about understanding what it means to comprehend. In W. B. Weimer & D. S. Palermo (Eds.), *Cognition and the symbolic processes.* Hillsdale, N. J.: Erlbaum, 1974.

Buswell, G. T. *How people look at pictures.* Chicago: University of Chicago Press, 1935.

Charniak, E. Organization and inference in a frame-like system of common knowledge. In R. C. Schank & B. Nash-Webber (Eds.), *Theoretical issues in natural language processing.* Arlington, Va.: Tinlap Press, 1975.

Clark, H. H. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior,* 1973, *12,* 335–359.

Cooper, L. A., & Shepard, R. N. Chronometric studies of the rotation of mental images. In W. G. Chase (Ed.), *Visual information processing.* New York: Academic Press, 1973.

Durham, R., Nunnally, J., & Lemond, L. The effects of levels of information conflict on visual selection. *Perception & Psychophysics,* 1971, *10,* 93–96.

Frederiksen, C. H. Effects of context-induced processing operations on semantic information acquired from discourse. *Cognitive Psychology,* 1975, *7,* 139–166.

Friedman, A. Memorial comparisons without the "mind's eye." *Journal of Verbal Learning and Verbal Behavior,* 1978, *17,* 427–444.

Gibson, J. J. *The senses considered as perceptual systems.* Boston: Houghton Mifflin, 1966.

Gibson, J. J. The theory of affordances. In R. Shaw & J. D. Bransford (Eds.), *Perceiving, acting, and knowing.* Hillsdale, N. J.: Erlbaum, 1977.

Gross, C., Bender, D., & Rocha-Miranda, C. Inferotemporal cortex: A single unit analysis. In F. O. Schmitt & F. G. Worden (Eds.), *The neurosciences: A third study program.* Cambridge: MIT Press, 1973.

Haviland, S. E., & Clark, H. H. What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior,* 1974, *13,* 512–521.

Hewitt, C. Stereotypes as an actor approach towards solving the problem of procedural attachment in frame theories. In R. C. Schank & B. Nash-Webber (Eds.), *Theoretical issues in natural language processing.* Arlington, Va.: Tinlap Press, 1975.

Hock, H., Gordon, G., & Whitehurst, R. Contextual relations: The influence of familiarity, physical plausibility, and belongingness. *Perception & Psychophysics,* 1974, *16,* 4–8.

Hubel, D. H., & Weisel, T. N. Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *Journal of Physiology,* 1962, *160,* 106–154.

Jenkins, J., Wald, J., & Pittenger, J. Apprehending pictorial events: An instance of psychological cohesion. In C. W. Savage (Ed.), *Minnesota studies in the philosophy of science* (Vol. 9). Minneapolis: University of Minnesota Press, 1977.

Johnson, M. K., Bransford, J. D., & Solomon, S. Memory for tacit implications of sentences. *Journal of Experimental Psychology,* 1973, *97,* 203–205.

Just, M., & Carpenter, P. Eye fixations and cognitive processes. *Cognitive Psychology,* 1976, *8,* 441–480.

Keenan, J., MacWhinney, B., & Mayhew, D. Pragmatics in memory: A study of natural conversation. *Journal of Verbal Learning and Verbal Behavior,* 1977, *16,* 549–560.

Kieras, D. Beyond pictures and words: Alternative information-processing models for imagery effects in verbal memory. *Psychological Bulletin,* 1978, *85,* 532–554.

Kintsch, W. *The representation of meaning in memory.* Hillsdale, N.J.: Erlbaum, 1974.

Kintsch, W. On comprehending stories. In M. Just & P. Carpenter (Eds.), *Cognitive processes in comprehension.* Hillsdale, N. J.: Erlbaum, 1977.

Kintsch, W., & Keenan, J. Reading rate and retention as a function of the number of propositions in the base structure of sentences. *Cognitive Psychology,* 1973, *5,* 257–274.

Kolers, P. Memorial consequences of automatized encoding. *Journal of Experimental Psychology: Human Learning and Memory*, 1975, *1* 689–701.

Kosslyn, S. Information representation in visual images. *Cognitive Psychology*, 1975, *7*, 341–370.

Kosslyn, S., & Pomerantz, J. Imagery, propositions, and the form of internal representations. *Cognitive Psychology*, 1977, *9*, 52–76.

Kuipers, B. A frame for frames: Representing knowledge for recognition. In D. G. Bobrow & A. Collins (Eds.), *Representation and understanding: Studies in cognitive science*. New York: Academic Press, 1975.

Lea, G. Chronometric analysis of the method of loci. *Journal of Experimental Psychology: Human Perception and Performance*, 1975, *1*, 95–104.

Lefton, L., & Fisher, D. Information extraction during visual search: A developmental progression. *Journal of Experimental Child Psychology*, 1976, *22*, 346–361.

Loftus, G. R. Eye fixations and recognition memory for pictures. *Cognitive Psychology*, 1972, *3*, 525–551.

Loftus, G. R. A framework for a theory of picture recognition. In R. A. Monty & J. W. Senders Eds.), *Eye movements and psychological processes*. Hillsdale, N. J.: Erlbaum, 1976.

Loftus, G. R., & Mackworth, N. H. Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance*, 1978, *4*, 565–572.

Mackworth, N. H., & Morandi, A. J. The gaze selects informative details within pictures. *Perception & Psychophysics*, 1967, *2*, 547–552.

Mandler, J., & Johnson, N. Some of the thousand words a picture is worth. *Journal of Experimental Psychology: Human Learning and Memory*, 1976, *2*, 529–540.

Mandler, J., & Johnson, N. Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, 1977, *9*, 111–151.

Mandler, J., & Parker, R. E. Memory for descriptive and spatial information in complex pictures. *Journal of Experimental Psychology: Human Learning and Memory*, 1976, *2*, 38–48.

Mandler, J., & Ritchey, G. Long-term memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, 1977, *3*, 386–396.

Mandler, J., & Stein, N. Recall and recognition of pictures by children as a function of organization and distractor similarity. *Journal of Experimental Psychology*, 1974, *102*, 657–669.

Meyer, D., Schvaneveldt, R., & Ruddy, M. Loci of contextual effects on visual word recognition. In P. Rabbitt & S. Dornic (Eds.), *Attention and performance V*. New York: Academic Press, 1975.

Mial, R., Smith, P., Doherty, M., & Smith, O. The effect of memory color on form identification. *Perception & Psychophysics*, 1974, *16*, 1–3.

Minsky, M. A framework for representing knowledge. In P. H. Winston (Ed.), *The psychology of computer vision*. New York: McGraw-Hill, 1975.

Mitterer, J., & Rowland, G. Picture memory: Role of figure-ground organization. *Perceptual & Motor Skills*, 1975, *40*, 753–754.

Moyer, R. S. Comparing objects in memory: Evidence suggesting an internal psychophysics. *Perception & Psychophysics*, 1973, *13*, 180–184.

Moyer, R. S., & Bayer, R. Mental comparison and the symbolic distance effect. *Cognitive Psychology*, 1976, *8*, 228–246.

Myers, J. L. *Fundamentals of experimental design* (2nd ed.). Boston: Allyn & Bacon, 1972.

Nash-Webber, B. The role of semantics in automatic speech understanding. In D. G. Bobrow & A. Collins (Eds.), *Representation and understanding: Studies in cognitive science*. New York: Academic Press, 1975.

Nickerson, R. S. Short term memory for complex meaningful visual material. *Canadian Journal of Psychology*, 1965, *19*, 155–160.

Nickerson, R. S. A note on long term recognition memory for pictorial material. *Psychonomic Science*, 1968, *11*, 58.

Norman, D. A., & Rumelhart, D. E. (Eds.). *Explorations in cognition*. San Francisco: Freeman, 1975. (a)

Norman, D. A., & Rumelhart, D. E. Memory and knowledge. In D. A. Norman & D. E. Rumelhart (Eds.), *Explorations in cognition*. San Francisco, Freeman, 1975. (b)

Norman, D. A., & Rumelhart, D. E. Reference and comprehension. In D. A. Norman & D. E. Rumelhart (Eds.), *Explorations in cognition*. San Francisco: Freeman, 1975. (c)

Oldfield, R., & Wingfield, A. Response latencies in naming objects. *Quarterly Journal of Experimental Psychology*, 1965, *17*, 273–281.

Paivio, A. Perceptual comparisons through the mind's eye. *Memory & Cognition*, 1975, *3*, 635–647.

Palmer, S. E. The effects of contextual scenes on the identification of objects. *Memory & Cognition*, 1975, *3*, 519–526. (a)

Palmer, S. E. The nature of perceptual representation: An examination of the analog/propositional controversy. In R. C. Schank & B. Nash-Webber (Eds.), *Theoretical issues in natural language processing*. Arlington, Va.: Tinlap Press, 1975. (b)

Palmer, S. E. Visual perception and world knowledge: Notes on a model of sensory-cognitive interaction. In D. A. Norman & D. E. Rumelhart (Eds.), *Explorations in cognition*. San Francisco: Freeman, 1975. (c)

Palmer, S. E. Hierarchical structure in perceptual representation. *Cognitive Psychology*, 1977, *9*, 441–475.

Pollack, I., & Spence, D. Subjective pictorial information and visual search. *Perception & Psychophysics*, 1968, *3*, 41–44.

Potter, M. Meaning in visual search. *Science*, 1975, *198*, 965–966.

Potter, M. Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, 1976, *2*, 509–522.

Pylyshyn, Z. W. What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin*, 1973, *80*, 1–24.

Pylyshyn, Z. W. Do we need images and analogues? In R. C. Schank & B. Nash-Webber (Eds.), *Theoretical issues in natural language processing*. Arlington, Va.: Tinlap Press, 1975.

Raphael, B. *The thinking computer*. San Francisco: Freeman, 1976.

Riesbeck, C. Computational understanding. In R. C. Schank & B. Nash-Webber (Eds.), *Theoretical issues in natural language processing*. Arlington, Va.: Tinlap Press, 1975.

Rocha-Miranda, C., Bender, D., Gross, C., & Mishkin, M. Visual activation of neurons in inferotemporal cortex depends on striate cortex and forebrain commissures. *Journal of Neurophysiology*, 1975, *38*, 475–491.

Rosch, E. Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 1975, *104*, 192–233.

Rumelhart, D. E. Understanding and summarizing brief stories. In D. LaBerge & S. J. Samuels (Eds.), *Basic processes in reading: Perception and comprehension*. Hillsdale, N. J.: Erlbaum, 1977.

Rumelhart, D. E., & Norman, D. A. The active structural network. In D. A. Norman & D. E. Rumelhart (Eds.), *Explorations in cognition*. San Francisco: Freeman, 1975. (a)

Rumelhart, D. E., & Norman, D. A. The computer implementation. In D. A. Norman & D. E. Rumelhart (Eds.), *Explorations in cognition*. San Francisco: Freeman, 1975. (b)

Ryan, T., & Schwartz, C. Speed of perception as a function of mode of representation. *American Journal of Psychology*, 1956, *69*, 60–69.

Schank, R. C. Using knowledge to understand. In R. C. Schank & B. Nash-Webber (Eds.), *Theoretical issues in natural language processing*. Arlington, Va.: Tinlap Press, 1975.

Schank, R. C., & Abelson R. Scripts, plans, and knowledge. *Proceedings of the Fourth International Joint Conference on Artificial Intelligence*, Tbilisi, USSR, 1975.

Schank, R. C., & Nash-Webber, B. (Eds.). *Theoretical issues in natural language processing*. Arlington, Va.: Tinlap Press, 1975.

Selfridge, O. G., & Neisser, U. Pattern recognition by machine. *Scientific American*, 1960, *203*, 60–68.

Shepard, R. N. Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, 1967, *6*, 156–163.

Standing, L., Conezio, J., & Haber, R. Perception and memory for pictures: Single trial learning of 2500 visual stimuli. *Psychonomic Science*, 1970, *19*, 73–74.

Sulin, R., & Dooling, D. Intrusion of a thematic idea in retention of prose. *Journal of Experimental Psychology*, 1974, *103*, 255–262.

Turvey, M. T. Contrasting orientations to the theory of visual information processing. *Psychological Review*, 1977, *84*, 67–88.

van Dijk, T. Semantic macro-structures and knowledge frames in discourse comprehension. In M. Just & P. Carpenter (Eds.), *Cognitive processes in comprehension*. Hillsdale, N. J.: Erlbaum, 1977.

Weisstein, N., & Harris, C. S. Visual detection of line segments: An object-superiority effect. *Science*, 1974, *186*, 752–755.

Wingfield, A. Effects of frequency on identification and naming of objects. *American Journal of Psychology*, 1968, *81*, 226–234.

Winograd, T. Understanding natural language. *Cognitive Psychology*, 1972, *3*, 1–191.

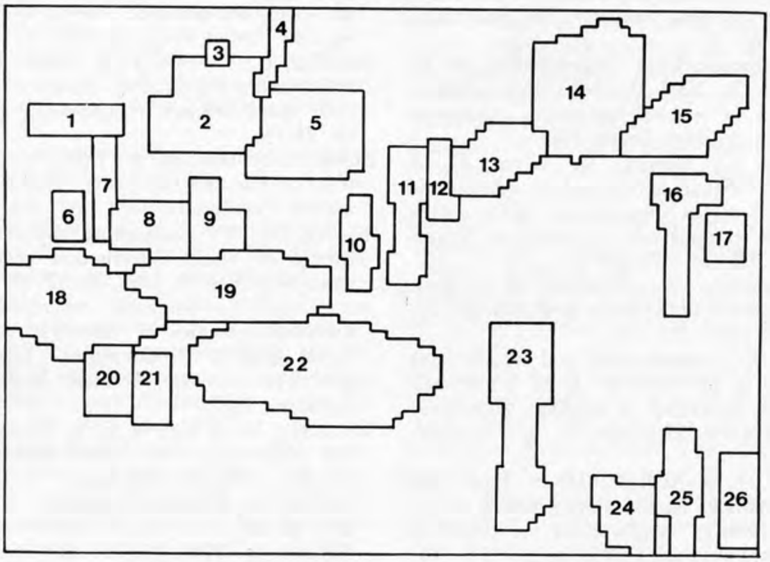
Winograd, T. Frame representations and the declarative-procedural controversy. In D. G. Bobrow & A. Collins (Eds.), *Representation and understaning: Studies in cognitive science*. New York: Academic Press, 1975.

(*Appendices start on next page.*)

Appendix A

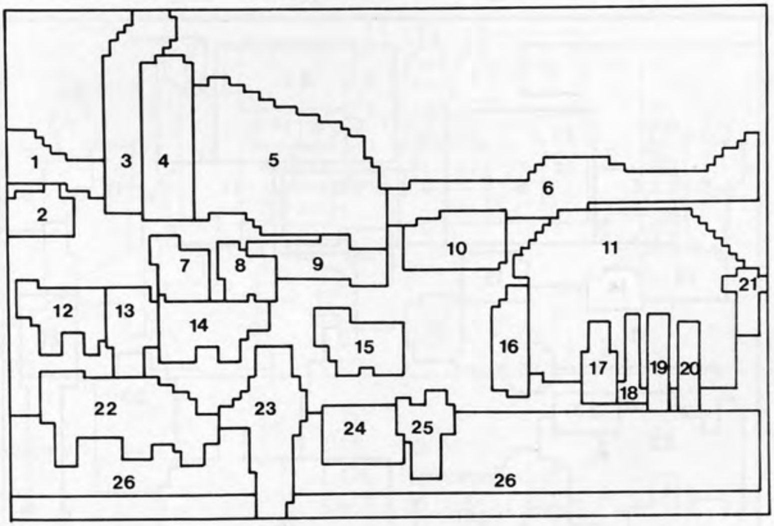
Schematic Representations of the Target Pictures and the Rated Likelihoods of the Objects

Figure A1. City:



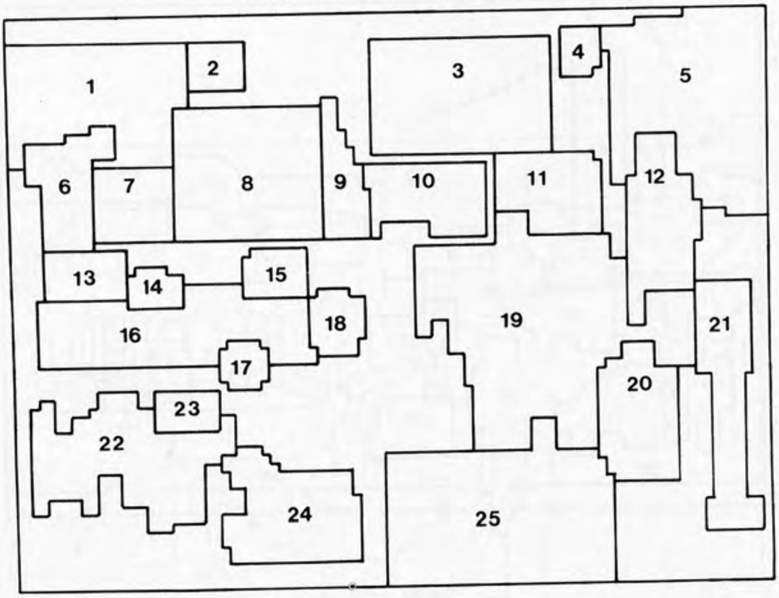
1. sign on a bank	2.235	14. people and parking meters	1.778
2. balcony	3.353	15. car	1.020
3. pigeon	3.608	16. one-way sign	1.176
4. balloon	5.588	17. wire trash basket	3.471
5. canopy	3.255	18. police car	2.392
6. fire hydrant	1.333	19. car	1.020
7. bus stop sign	1.765	20. old-fashioned standing radio	6.059
8. bench at bus stop	2.039	21. mailbox	3.098
9. woman and child	1.039	22. taxicab	1.863
10. construction worker	2.961	23. traffic light	1.235
11. traffic light	1.235	24. dog	3.255
12. man	1.039	25. man	1.039
13. vegetable cart	5.020	26. fire alarm box	1.882

Figure A2. Farm:



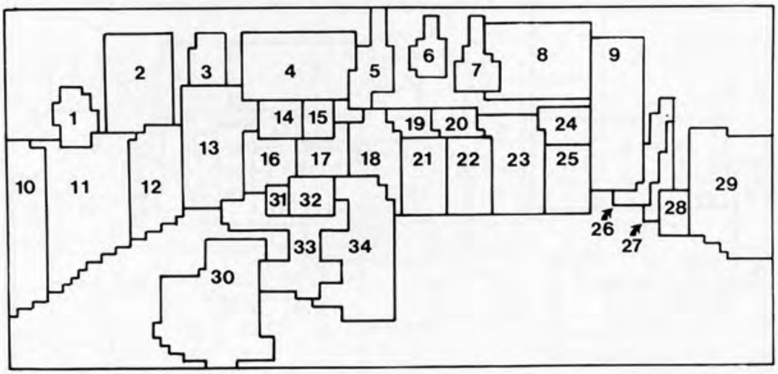
1. trees	1.922	14. cow	2.039
2. steamroller	6.490	15. sheep	3.275
3. factory chimney stack	6.510	16. worker	3.196
4. silo	3.400	17. butter churn	4.333
5. barn	1.373	18. hoe	1.549
6. trees	1.922	19. shovel	1.627
7. horse	2.176	20. broom	1.725
8. pony	3.275	21. scarecrow	3.824
9. pickup truck	1.784	22. cow	2.039
10. oil truck	6.100	23. farmer	1.098
11. tool shed	2.260	24. horse blanket	4.760
12. cow	2.039	25. saddle	4.608
13. cow	2.039	26. fence	2.588

Figure A3. Kindergarten:



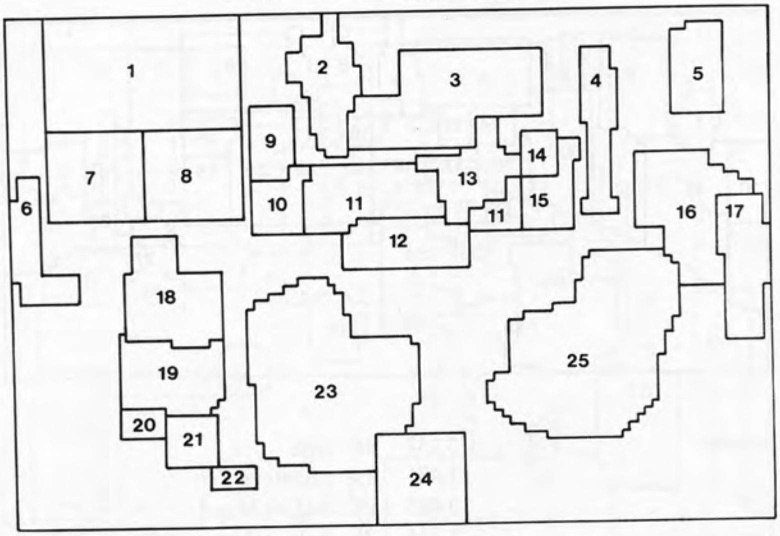
1. window	1.412	14. teddy bear	3.314
2. clock	1.569	15. television	4.843
3. blackboard	1.118	16. bookcases	1.804
4. fire bell	2.078	17. basketball	4.353
5. movie screen	4.118	18. globe	2.922
6. telescope	5.588	19. children, table, and chairs	1.140
7. cupboard	2.373	20. child	1.098
8. piano and stool	3.510	21. traffic light	6.431
9. cello	5.980	22. table and chairs	1.211
10. children's desk	1.725	23. radio	5.039
11. bookcase	1.804	24. hobbyhorse	3.400
12. teacher	2.627	25. desk, blotter, apple, and pencil cup	1.098
13. aquarium	4.098		

Figure A4. Kitchen:



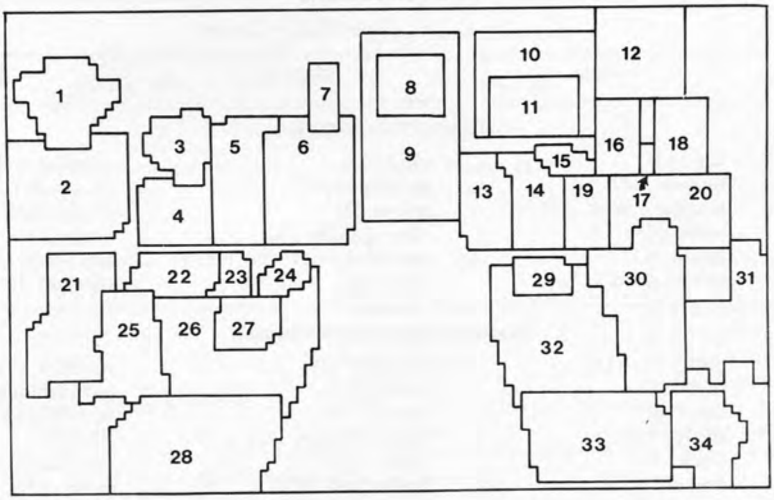
1. plant on counter	3.137	18. sink	1.039
2. window	1.471	19. canisters	2.039
3. lamp	5.863	20. loaf of bread	1.490
4. cabinets	1.314	21. dishwasher	3.137
5. large hanging plant	4.776	22. cabinet	1.314
6. small hanging plant	3.760	23. stove and pots	1.941
7. small hanging plant	3.760	24. rotisserie	4.176
8. cabinets	1.314	25. cabinet	1.314
9. door	1.471	26. hockey stick	6.745
10. cabinet	1.314	27. broom	2.235
11. fireplace	6.647	28. pail	4.420
12. bookcase	5.627	29. old-fashioned steam radiator	5.480
13. refrigerator	1.020	30. rocking chair	6.490
14. toaster	1.922	31. salt and pepper shakers	2.275
15. coffeepot	1.784	32. bowl of fruit	3.480
16. cabinet	1.314	33. dining table	3.667
17. cabinet	1.314	34. dining chairs	3.667

Figure A5. Living Room:



1. bookcases	3.549	14. vase	2.216
2. hanging plant	3.600	15. end table and plant	1.861
3. group of paintings	2.529	16. desk	4.235
4. floor lamp	2.549	17. large floor plant	3.431
5. painting	2.529	18. portable television	4.451
6. vacuum cleaner	6.040	19. television stand	4.451
7. bookcase	3.549	20. roller skate	6.102
8. cabinet	4.686	21. toy truck	6.102
9. table lamp	1.157	22. toy car	6.102
10. end table and ash tray	1.794	23. stuffed chair	2.647
11. couch and pillows	1.618	24. drum	6.490
12. coffee table	2.255	25. recliner	3.216
13. man	2.118		

Figure A6. Office:



1. plant	3.765	18. water cooler	3.098
2. safe	4.000	19. file cabinet	1.740
3. plant	3.765	20. storage cabinets	2.612
4. file cabinets	1.740	21. chair	1.740
5. chest of drawers	6.549	22. typewriter	1.549
6. file cabinets	1.740	23. index card file	2.080
7. sculpture	5.353	24. calculator	2.620
8. glass door with writing	1.720	25. swivel chair	2.863
9. door	1.720	26. desk	1.451
10. window	1.720	27. in–out baskets	2.220
11. air conditioner	3.333	28. stereo	5.100
12. calendar	1.667	29. telephone	1.235
13. file cabinet	1.740	30. secretary, typewriter	1.628
14. storage cabinet	2.612	31. chair	1.740
15. cups	2.900	32. desk and ashtray	1.569
16. coffee urn	2.900	33. steamer trunk	6.617
17. cups	2.900	34. magazine rack	3.196

Appendix B

Table B1
Objects Used in the Distractor Transformations

Picture	Likelihood of object: High	Medium	Low
	Token transformation		
city	car (15)	dog (24)	vegetable cart (13)
farm	farmer (23)	scarecrow (21)	oil truck (10)
kindergarten	teacher's desk (25)	piano (8)	traffic light (21)
kitchen	stove (23)	dining table (33)	lamp (3)
living room	couch (11)	portable television (18)	toy truck (21)
office	swivel chair (25)	plant (3)	chest of drawers (5)
	Rearrangement transformation		
city	police car (18)	balcony (2)	mailbox (21)
	taxi (22)	canopy (5)	old-fashioned radio (20)
farm	hoe (18)	saddle (25)	factory chimney stack (3)
	shovel (19)	horse blanket (24)	silo (4)
kindergarten	clock (2)	globe (18)	television (15)
	firebell (4)	basketball (17)	radio (23)
kitchen	sink (18)	small hanging plant (5)	bookcase (12)
	cabinet (22)	large hanging plant (7)	fireplace (11)
living room	vase (14)	bookcase (7)	toy car (22)
	table lamp (9)	cabinet (8)	rollerskate (20)
office	telephone (29)	water cooler (18)	stereo (28)
	calculator (24)	coffee urn (16)	steamer trunk (33)
	Deletion transformation		
city	fire hydrant (6)	dog (24)	balloon (4)
farm	farmer (23)	sheep (15)	steamroller (2)
kindergarten	child (20)	aquarium (13)	radio (23)
kitchen	loaf of bread (20)	bowl of fruit (32)	steam radiator (29)
living room	vase (14)	large floor plant (17)	vacuum cleaner (6)
office	in–out baskets (27)	plant (3)	sculpture (7)
	Type-high transformation[a]		
city	one-way sign (16)	fire alarm box (26)	bench at bus stop (8)
	street sign	mailbox on a post	sofa
farm	broom (20)	pickup truck (9)	cow (14)
	pitchfork	flatbed truck	hippopotamus
kindergarten	bookcase (11)	blackboard (3)	clock (2)
	cupboard	wall map	dartboard
kitchen	coffeepot (15)	toaster (14)	window (2)
	paper towels	radio	oil painting
living room	man (13)	floor lamp (4)	coffee table (12)
	woman	large floor plant	steamer trunk
office	storage cabinet (14)	calendar (12)	typewriter (22)
	file cabinet	painting	tape deck

Table B1 (*continued*)

Picture	Likelihood of object: High	Medium	Low
	Type-medium transformation[a]		
city	construction worker (10)	wire trash basket (17)	pigeon (3)
	businessman	cement planter	chicken
farm	butter churn (17)	sheep (15)	pony (8)
	bucket	goat	deer
kindergarten	movie screen (5)	teddy bear (14)	hobbyhorse (24)
	blackboard	baby doll	dog
kitchen	pail (28)	rotisserie (24)	dishwasher (21)
	garbage can	portable television	washing machine
living room	desk (16)	desk (16)	recliner (25)
	stereo	cabinet	barber's chair
office	magazine rack (34)	air conditioner (11)	safe (2)
	briefcase	window fan	cigarette machine
	Type-low transformation[a]		
city	vegetable cart (13)	old-fashioned radio (20)	old-fashioned radio (20)
	subway	mailbox	jukebox
farm	steamroller (2)	factory chimney stack (3)	oil truck (10)
	tractor	silo	covered wagon
kindergarten	traffic light (21)	cello (9)	telescope (6)
	coat rack	guitar	machine gun
kitchen	steam radiator (29)	hockey stick (26)	rocking chair (30)
	cabinet	mop	stuffed chair
living room	drum (24)	drum (24)	vacuum cleaner (6)
	end table	hassock	lawn mower
office	chest of drawers (5)	steamer trunk (33)	chest of drawers (5)
	filing cabinet	coffee table	refrigerator

Note. Numbers in parentheses refer to the schematic diagrams of the scenes in Appendix A.
[a] The first row of objects for each picture gives the original target objects; the nomenclature for the type transformations refers to the rated likelihood of these original objects. The second row of objects for each picture gives the new objects, which had either high, medium, or low rated likelihoods. For example, in the type-high transformation, for the City picture, a bench at a bus stop was the original object and it had a high rated likelihood; a sofa was the new object and it had a low rated likelihood.

Received May 25, 1978 ■