COGNITIVE PSYCHOLOGY 14, 143–177 (1982)

Scene Perception: Detecting and Judging Objects Undergoing Relational Violations

IRVING BIEDERMAN, ROBERT J. MEZZANOTTE, AND JAN C. RABINOWITZ
State University of New York at Buffalo

Five classes of relations between an object and its setting can characterize the organization of objects into real-world scenes. The relations are (1) Interposition (objects interrupt their background), (2) Support (objects tend to rest on surfaces), (3) Probability (objects tend to be found in some scenes but not others), (4) Position (given an object is probable in a scene, it often is found in some positions and not others), and (5) familiar Size (objects have a limited set of size relations with other objects). In two experiments subjects viewed brief (150 msec) presentations of slides of scenes in which an object in a cued location in the scene was either in a normal relation to its background or violated from one to three of the relations. Such objects appear to (1) have the background pass through them, (2) float in air, (3) be unlikely in that particular scene, (4) be in an inappropriate position, and (5) be too large or too small relative to the other objects in the scene. In Experiment I, subjects attempted to determine whether the cued object corresponded to a target object which had been specified in advance by name. With the exception of the Interposition violation, violation costs were incurred in that the detection of objects undergoing violations was less accurate and slower than when those same objects were in normal relations to their setting. However, the detection of objects in normal relations to their setting (innocent bystanders) was unaffected by the presence of another object undergoing a violation in that same setting. This indicates that the violation costs were incurred not because of an unsuccessful elicitation of a frame or schema for the scene but because properly formed frames interfered with (or did not facilitate) the perceptibility of objects undergoing violations. As the number of violations increased, target detectability generally decreased. Thus, the relations were accessed from the results of a single fixation and were available sufficiently early during the time course of scene perception to affect the perception of the objects in the scene. Contrary to expectations from a bottom-up account of scene perception, violations of the pervasive physical relations of Support and Interposition were not more disruptive

This research was supported by research grants from the U.S. Army Research Institute for the Behavioral and Social Sciences (Grant MDA903-A-G-0003) and National Institutes of Mental Health (Grant MH33283) to Irving Biederman. The first author also held an NIH Postdoctoral fellowship, MHO7891. We thank Carl M. Francolini and Dana Plude for preparing the slides and running the subjects in Experiment I. Jan C. Rabinowitz is now at the University of Toronto. This manuscript benefitted from the comments and criticism of Frank G. Halasz, Gregory L. Murphy, Richard C. Teitelbaum, James R. Pomerantz, Mark D. Murray, Alinda Friedman, Geoffrey Loftus, and Brian Ross. Requests for reprints should be sent to Irving Biederman, Department of Psychology, State University of New York at Buffalo, 4230 Ridge Lea Road, Amherst, NY 14226.

0010-0285/82/020143-35$05.00/0

on object detection than the semantic violations of Probability, Position and Size. These are termed semantic because they require access to the referential meaning of the object. In Experiment II, subjects attempted to detect the presence of the violations themselves. Violations of the semantic relations were detected more accurately than violations of Interposition and at least as accurately as violations of Support. As the number of violations increased, the detectability of the incongruities between an object and its setting increased. These results provide converging evidence that semantic relations can be accessed from the results of a single fixation. In both experiments information about Position was accessed at least as quickly as information on Probability. Thus in Experiment I, the interference that resulted from placing a fire hydrant in a kitchen was not greater than the interference from placing it on top of a mail box in a street scene. Similarly, violations of Probability in Experiment II were not more detectable than violations of Position. Thus, the semantic relations which were accessed included information about the detailed interactions among the objects—information which is more specific than what can be inferred from the general setting. Access to the semantic relations among the entities in a scene is not deferred until the completion of spatial and depth processing and object identification. Instead, an object's semantic relations are accessed simultaneously with its physical relations as well as with its own identification.

What are the mental events that transpire when our eyes alight upon a novel scene? The comprehension that is achieved is not a simple listing of the creatures and objects. Instead, our mental representation includes a specification of the various relations that exist among these entities.

Some of these relations can be coded solely with reference to physical space. They indicate *where* an object is relative to the other objects in the scene. Such relations can be described verbally by prepositions, such as "on," "in front of," or "in." Other relations, however, require access to the referential meaning of the entities in question. These relations are typically described with verbs or gerunds, such as "eating," "reading," or "playing."

Two questions were of central concern in the present investigation. First, would access to these relations—even those relations dependent upon semantic information—be so fast as to affect the perceptibility of an object? Second, would some kinds of relations be more readily available than others, more detectable or more potent in affecting the perceptibility of objects than other kinds of relations?

The experimental strategy required the construction of scenes in which one or more of the expected relations which typically hold between an object and its setting were violated. In Experiment I, the effects of these violations on the speed and accuracy of the detection of the object were assessed. In Experiment II, subjects judged the presence of the violations themselves. A systematic study of the effects of violating the relations between an object and its setting requires some discussion as to what these relations might be.

TABLE 1
List of Relational Violations and Examples for a Single Object

1. *Support*, e.g., a floating fire hydrant. The object does not appear to be resting on a surface.
2. *Interposition*, e.g., building in the background passing through the hydrant. The background appears to pass through the object.
3. *Probability*, e.g., the hydrant in a kitchen. The object is unlikely to appear in the scene.
4. *Position*, e.g., the fire hydrant on top of a mailbox in a street scene. The object is likely to occur in that scene but it is unlikely to be in that particular position.
5. *Size*, e.g., the fire hydrant appearing larger than a building. The object appears too large or too small relative to the other objects in the scene.

Object Relations and Coherent Scenes

Surprisingly, only five classes of relations may be sufficient to characterize the difference between a display of unrelated objects and a well-formed scene (Biederman 1977, 1981). These are listed in Table 1 and illustrated by examples of their violations in a manner similar to the way in which linguistic relations can be illustrated through their violations. Thus, "the *angry* napkin" illustrates a semantic violation, and "he *smiled* the baby" results in a syntactic violation (since the intransitive verb "smiled" requires a preposition, "at") (cf. Moore, 1972).

The first two relations, Support and Interposition[1], reflect the general physical constraints of gravity (that most objects do not fly or float in air) and that an opaque object will occlude the contours of an object behind it. It should be noted that when an object which is floating in air is designated as an instance of a physical violation of Support, then the designation of this relation as a *violation* is ultimately based on the semantic inappropriateness of the relation, since some objects, e.g. birds, balloons, can normally be unsupported in air. It is possible that the Interposition violation is also ultimately based on semantic inappropriateness, although it is difficult to think of objects whose normal appearance is one where the background passes through an opaque surface. When an object violates the Interposition relation, that object does not merely appear to be transparent. Transparency is itself readily perceived (Metelli, 1974), and only rarely do transparent objects yield equivocal depth relations. Violations of Interposition, however, produce ambiguous coexistence of the object and the background in the same position in depth. The two violations of Interposition and Support are considered together here because the origin

[1] The names of the different classes of relations and their violations will be capitalized to distinguish them from general usage of these words. Ampersands will be used to denote multiple violation conditions.

of the incongruity for these violations could be in an inappropriate assignment of surfaces to bodies during physical parsing of object surfaces prior to any identification of what these bodies might be. This point will be elaborated below when computer vision models of physical parsing are discussed.

While it would be possible to determine that an object was floating in air or did not occlude its background without knowing what the object was, the latter three constraints, Probability, Position, and familiar Size are *semantic* relations, in that they require access to the referential meaning of the objects. (This makes it convenient to consider Interposition and Support as *syntactic* relations.) Probability refers to the likelihood of a given object being in a given scene. Fire hydrants are rarely found in kitchens. The Position relation refers to the fact that objects which are likely to occur in a given scene often occupy specific positions in that scene. Thus, fire hydrants are found on the sidewalk in a street scene, not on top of the mailbox or in the middle of the street. To the extent that objects in a scene are processed independently of the specific positions occupied by other objects, Probability relations will be more readily accessed than Position relations. In Table 1, it should be emphasized that Size refers to the familiar *relative* size of objects, which is achieved through a comparison of an object to other objects in a picture. It does not refer to the visual angle subtended by the object in the scene. Holding the visual angle constant, a kitchen chair could be made to appear smaller than a cup or larger than a refrigerator, depending on whether it was moved toward the background (to make it look larger) or toward the foreground (to make it look smaller) in a picture of a kitchen scene.

Taken individually, the five relations have all been identified at one point or another during the history of perception. Why consider them here? Some of the relations, viz. Interposition, Support, and familiar Size, have been studied only with respect to their role in depth perception (e.g. Gibson, 1950, 1966). That these relations might affect the course of object identification has been overlooked. (Part of the reason is that the psychology of space and depth perception has evolved independently from the psychology of pattern recognition. Research in psychology on problems of pattern recognition has been largely concerned with the perception of print, where issues of spatial and depth relations are not encountered.) Violations of Probability and Position have received some recent study on the tendency of objects undergoing these violations to capture eye fixations (Loftus & Mackworth, 1978; Friedman, 1979). These experiments as well as other studies with Probability and Position violations (e.g. Mandler & Stein, 1974; Hock, Romanski, Galie & Williams, 1978) have concentrated on recall and recognition, rather than on the perception of objects undergoing these violations. Biederman, Glass, and Stacy

(1973), and Palmer (1975), showed that the Probability relation could readily be accessed from a scene, but neither study was designed to directly assess a perceptual effect.

Thus, the five classes of relations have not been systematically studied for their perceptual effects on object recognition. Moreover, it has not been appreciated that these relations might constitute a sufficient set with which to characterize the organization of a real-world scene as distinct from a display of unrelated objects.[2] Proposals for a sixth relation would be welcome, as none have been found to be acceptable in several years of discussions about this research. It should be noted that the relations refer to the arrangement of well-formed objects, rather than to the features of those objects. Thus, for example, the 180° rotation of an object (the most frequently proposed sixth candidate), would be interpretable in terms of altering the object's features, since many of the features used in identifying objects are orientation-specific.

Accessibility of the Relations

Are some of the relations accessed faster than other relations? One relation could be accessed faster than another relation if it was processed earlier by a sequential processor or required less time for its processing by a parallel processor. (The issue of parallel versus sequential access to the relations will be examined in the discussion of Experiment II.) Under the assumption that faster access (by a serial or parallel processor) to a relation would result in faster detection of its violation, Experiment II was designed to measure the relative accessibility of the different relations by requiring detection of the presence of the violations.

Experiment I provided a somewhat indirect exploration of the accessibility of the different relations by determining whether the violation of one kind of relation resulted in more interference with object detection than violation of another kind of relation. The identification of the relative magnitudes of interference effects with the order of availability of the violations is favored under the assumption that the earlier the arrival of misleading information, the greater the possibility that such information

[2] When the five relations are termed "sufficient" it is only in the sense that if they are not violated, a scene will not look anomalous or disorganized. They are *not* sufficient for conveying the representation of a well-formed scene. Put another way, if none of the objects in two scenes undergo any of the violations, all we would know from the above five relations is that the scenes were not anomalous. The relations by themselves do not convey what could be enormous differences in the meaning of the scenes. This is a problem quite analogous to noting that a given sentence does not violate any of the semantic or syntactic constraints posed by a given linguistic theory. All we would then know is that the sentence is well-formed and not semantically anomalous; we would not know, however, what the sentence meant.

will disrupt object detection. There are at least two reasons why this assumption is plausible. The most obvious reason is that the earlier a violation is accessed, the more likely that it will be present before the object is identified (so that it could have an opportunity to interfere with that object's identification). A second reason stems from general considerations of information combination in complex systems. Early misleading information may often require more time for its correction than later misleading information, in that the system has gone on to do additional processing based on the initial error. Such an analysis was offered by Bruner and Potter (1964) to explain why extremely blurred pictures which were gradually brought into sharper focus required more clarity before they could be identified than pictures that were initially presented with only a moderate degree of blur. The greater deleterious effect of an early (as opposed to late) processing error has been argued by Rumelhart (1977) to explain why garden-path sentences ("The old man the boats.") require longer pauses in their readings than syntactically disambiguated sentences ("The merchants ship their wares.")

We thus have two measures of the relative accessibility of a relation. One is the degree to which the violation of that relation interferes with object detection. The second is the speed and accuracy of the detection of that violation itself.

The accessibility of the relations is central to two general issues in scene perception. The first is whether the physical relations of Interposition and Support are accessed prior to the semantic relations. A bottom-up model that holds that physical relations are processed prior to semantic relations is compatible with theories positing that the processing for depth and space precede the accessing of meaning (e.g. Julesz, 1981; Gibson, 1966). Thus, Julesz (1981) distinguishes the "immediate" depth and contour information from slower "deliberative" processes through which meaning is achieved. Gibson's "direct" perception (1966), through which information is picked up about the distribution of bodies in space, is another account where spatial processing occurs prior to the access of semantic relations.

Perhaps the clearest statement of the bottom-up model is embodied in the scene analysis program of Guzman (Note 1). In principle, Guzman's program could physically parse the input from a line drawing of a collection of objects resembling children's blocks. "Physically parse" means that the various surfaces were assigned to the blocks in a manner identical to the way in which a human observer would assign the surfaces to the blocks. This was achieved through a classification of the vertices formed by the intersection of adjacent rectilinear surfaces. The impressive feature about this result was the claim that "SEE (the name of Guzman's pro-

gram) does not require a preconceived idea of the form of the object which could appear in the scenes. It assumes only that they will be solid objects formed by rectilinear surfaces." (Guzman, Note 1, p. 58). Winston (1975) demonstrated that relations such as Support, i.e. that block A is supported by block B; and Interposition, that block C is in front of block D, could be derived from the kinds of information extracted by Guzman's program. (More recent programs, e.g. Waltz, 1975; Marr, 1978, are able to parse more varied inputs than Guzman's and Winston's models but maintain the same assumptions of bottom-up priority.)

The work of Guzman and Winston demonstrated that it was possible to determine physical relations such as Support and Interposition without identifying the bodies. It is tempting to conjecture that in human perception this information is extracted before relations (viz. Probability, Size, and Position) that are dependent upon object identity. Indeed, the psychology of depth and spatial relations rarely includes any discussion of the semantic content of the scene being viewed. Essentially, this bottom-up view holds that there is an initial processing of the scene by the visual system in which information (features, spatial frequency components, etc.) is extracted. This information is used for the physical parsing of the scene, so that Support and Interposition can be determined. Presumably, the visual information is also used, perhaps simultaneously, to identify the individual objects. The physically parsed scene is then served up to higher levels where the semantic relations among the already identified objects would be specified. (See Biederman, 1981, for a more detailed presentation of such a model.) This bottom-up model proposes that the semantic relations follow, indeed are the result of, physical parsing and object identification, so that physical parsing and object identification proceed independently of the semantic relations among the objects.

The second issue is whether objects are identified prior to the determination of the way in which they interact. If this were true, then the Probability relation, which can be accessed solely from an identification of some of the objects, should be accessed prior to the Position relation, which requires specification of the way in which objects are interacting. Once a sink, stove, and frying pan are identified, fire hydrants become improbable no matter where they are positioned. However, the incongruity of a fire hydrant on top of a mail box in a street scene cannot be determined merely from such inventory listings (Mandler & Johnson, 1976) of the objects in the scene. The specific interaction between the two objects must be perceived. If an early stage in the achievement of a representation (schema, frame) of a scene is information about its general class of settings, e.g. that it is a kitchen, street scene, baseball game, or campsite, and if this information is derived from the identification of some

of the objects independently of the other objects in the scene, then Probability violations would be expected to be accessed faster than Position violations.[3]

It should be noted that these bottom-up expectations of faster access to the physical as compared to the semantic relations, and to Probability as compared to Position, do not depend upon having equivalent scale values for the various violations. If a scene is first physically parsed (for Interposition and Support), and then the objects are identified (so that Probability can be determined), then physical violations should be accessed faster (i.e. be more readily detectable in Experiment II and lead to larger violation costs in Experiment I) than semantic violations, even though the semantic violations were higher on an underlying scale of degree of violation. Similarly, if objects are identified prior to the determination of the specific ways in which they interact, then Probability should be accessed faster than Position, regardless of the respective scale values. Nonetheless, the instances of the various violations were selected so as to produce obvious and subjectively equivalent (i.e. ratings of approximately 9 on a 10-point scale of obviousness of a given violation) violations across the different relations, although there can be no guarantee that the underlying violation scales (whatever they may be) were equivalent. This problem is analogous to the psycholinguistic comparisons of the processing of, for example, a semantic to a syntactic violation. After obvious and equivalent instances of each class are selected, differences in processing are used to infer differences in access to these variations (cf. Moore, 1972; Moore & Biederman, 1979; Rumelhart, 1977, Ch. 3). Actually, in those investigations as well as in the present experiments, the differences in perceptual processing show that the equivalence defined by a ratings task, where the ratings can be made at leisure, do not necessarily reflect differences in temporal access. One likely reason for this is that a ratings task allows exhaustive processing before a response need be made, but a speeded detection task encourages the initiation of a response as soon as sufficient information is available (Moore & Biederman, 1979).

Specific predictions from the bottom-up model about the relative accessibility of Probability versus Size violations are more difficult to make in that they are dependent upon assumptions as to how violations of Size are registered and some of the details of how the processing of the physical relations might affect object identification. Nevertheless, a general pre-

[3] There is a second reason to expect that Probability violations would produce larger violation costs than Position or Size violations. Position violations require that the local region by the cued object be processed, but *any* region of the scene will typically contain sufficient information to produce a Probability violation. Thus even if the subject was not looking at an object undergoing a Probability violation, an inappropriate schema for it would be elicited nonetheless.

bined with the results of Experiment II, provided information as to whether serial or parallel processing models might be more compatible with these data.

Experiment I was also designed to determine if the detection of normally positioned objects would be affected by the presence of violations from other objects. For example, would a sofa floating in a street scene affect the perceptibility of an *innocent bystander* such as a normal-appearing fire hydrant on a sidewalk? The measurement of the effects of one object's violations on the detectability of bystanders provided a way of determining the origin of violation costs. If the costs were incurred because the violations interfered with the elicitation of the appropriate frame for the scene, and the elicitation of the frame would facilitate object detection, then the detectability of bystanders should be reduced. If, however, the costs were incurred because properly formed frames interfered with (or did not facilitate) the perceptibility of the objects undergoing violations, then the perceptibility of bystanders should not be affected by the presence of violations in the scene.

The inclusion of an innocent bystander condition allowed a comparison to be made between the effects of these violations and the scene-jumbling experiments (Biederman, 1972; Biederman, et al., 1974) in which scenes were divided into six sections and five of the six sections (all but the one containing the target) were rearranged so as to destroy the coherency of the target's context. The present experiment explored the minimal case of disrupting a target's context in that only a single other object underwent the relational violations.

In addition to testing the theoretical issues, the experiment also offered parametric data on the major psychophysical variables of object detection in a real-world scene: the effects of distance from fixation, target size, and camouflage.

METHOD

Scenes

Two hundred forty-seven scenes were composed by superimposing one or two clear acetate overlays, each with one of 42 objects drawn on them, over one of 17 background drawings. The backgrounds were of a variety of different settings; e.g. kitchen, downtown street, farm, living room, classroom, picnic. Each object, e.g. man, book, car, frying pan, was in a normal location in at least one of the slides but appeared in one to five slides where it underwent a violation. The background and overlays were then photocopied together to produce a scene with the object or objects in it and a slide was made of the photocopy. The slides were produced by direct positive development of Kodak Panatomic X film.

As mentioned above, each of the 42 objects appeared in at least one scene in which it was in a normal (or Base) condition relative to its setting. In the remaining 205 slides, the object was not in a Base condition; instead, it was displaced to various sections of the scene or imported to other scenes to violate one or several of the five constraints. Figure 1 is an example of a Position violation, Fig. 2 is an example of an Interposition violation, and Fig. 3 is an example of a triple violation of Size & Probability & Support.

diction can be advanced, based upon a comparison of the information needed to discern a violation of the two relations. The registration of Size violations requires not only the identification of at least one other object in the setting, but also the processing of the specific spatial and depth relations between the target and this other object (e.g. to know that the fire hydrant is too large, sufficient depth processing is required to perceive that it must be too large compared to the car which is next to it). In contrast, violations of Probability require only object identification and should, therefore, have faster access than violations of Size. The only way that Size could be accessed as readily as Probability is if the spatial relations were processed no later than the Probability relations in the overall course of scene processing, and such spatial processing did not impose any additional load on the capacities for processing scenes.

The results of the two experiments reported here confirm none of the bottom-up expectations that physical relations should have faster access than semantic relations, and that Probability should be accessed prior to Position and Size. Instead, it appears that information about the detailed semantic relations among the objects in a scene is accessed at least as quickly as information about the physical relations of Support and Interposition; quickly enough, in fact, to affect object identification.

EXPERIMENT I: OBJECT DETECTION

The major purpose of this experiment was to determine if the presence of a violation in the relation between an object and its setting would affect that object's perceptibility. Violations of all five relations were produced to allow comparison of the magnitudes of the various violations. As discussed above, the assumption by the bottom-up model of faster access to physical parsing and object identification as compared to the semantic relations leads to several predictions as to the relative magnitudes of the violation costs. First, since the semantic relations, viz. Size, Probability, and Position, would be derived only following object identification, the bottom-up model predicts that their violations should not affect object identification. To the extent that the Support and Interposition relations were accessed prior to object identification, then the bottom-up model predicts that violations of these relations would be expected to affect object identification. In particular, a violation of Interposition would be most disruptive to figure–ground segregation and would defeat a physical parsing program such as Guzman's (Note 1).

By simultaneously violating two or three of the relations between an object and its setting, the effects of multiple violations could be compared to single violations. Would any anomaly produce a constant effect or would two or three violations produce a greater effect than a single violation? The effects of variation in the number of violations, when com-

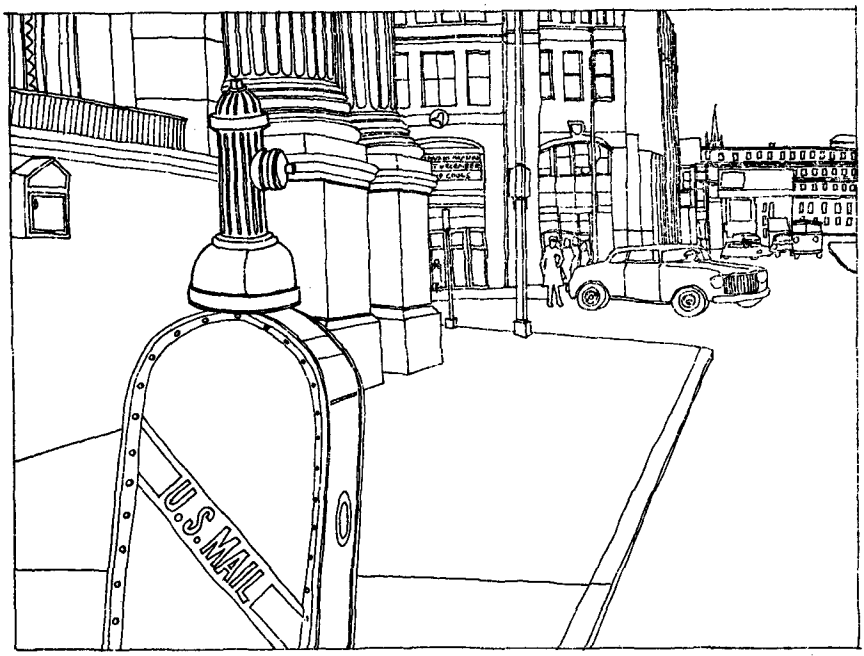

FIG. 1. An example of a Position violation for the fire hydrant. The camouflage rating for the fire hydrant was 5.5.

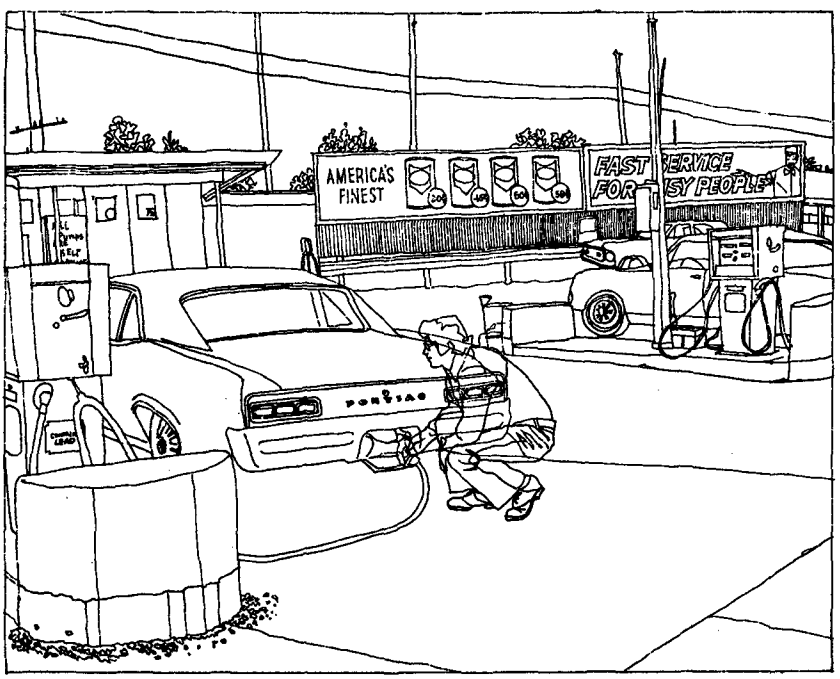

FIG. 2. An example of an Interposition violation for the man pumping gas. His camouflage rating was 8.0.

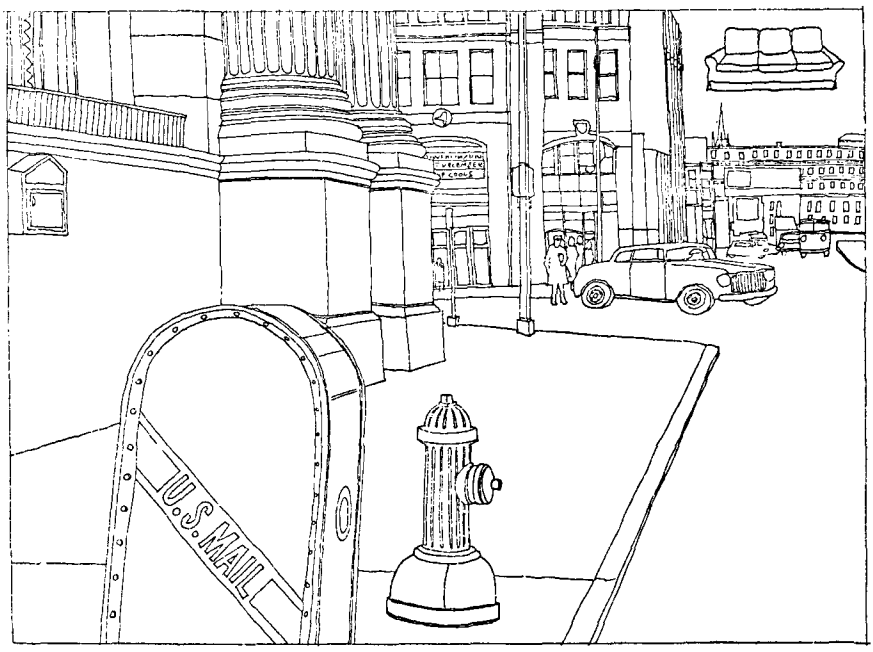

FIG. 3. "Goodyear Sofa." An example of a triple violation of Probability & Size & Support. The fire hydrant, which is shown in a Position violation in Fig. 1, would be an innocent bystander in this scene. If the sofa was not present, then the hydrant would be in a Base Condition. The camouflage rating for both the sofa and fire hydrant was 2.0.

The cued object in each scene was either in a Base condition (no violations), or was undergoing 1 of 10 Violation conditions listed in Table 2. In 5 Violation conditions the target violated only a single relation, in 4 conditions 2 relations were violated, and in one condition 3 relations were violated.

The selection of multiple violation conditions for inclusion in this experiment was subject to the restriction that not all combinations of violations were deemed possible. Specifically, an object violating either (or both) Support or Probability, could not also violate Position. The assumption motivating this exclusion was that if an object was unlikely to be in a scene or was inappropriately floating in air, then *any* position would be inappropirate. (A criticism of this assumption is discussed under Violation Specifications.) Another restriction was that we wanted to be able to evaluate the three pairwise violations of the relations that comprised the triple violation of Size & Probability & Support, so that the number of violations could be varied over a fixed set of 3 violations. The remaining multiple violation, Size & Position, was included because it was readily varied with this set of stimuli. Current experiments are exploring other combinations of multiple violations.

Physical Specifications of Scenes and Cued Objects.

The various conditions and their specifications for distance from fixation, size, and camouflage of their cued objects are shown in Table 2. An attempt was made to select instances so as to maintain approximately equivalent values for these measures across the various conditions.

TABLE 2
Mean Distance from Fixation, Size, and Camouflage of the Cued Objects in Each of the Slide Conditions

Condition	No. of slides	Distance from fixation (deg)	Size (length × width) (deg)	Camouflage rating
Zero violations				
Base	42	3.26	4.60	3.9
One violation				
Interposition	23	3.23	5.44	7.5
Support	27	4.09	4.18	4.3
Size	21	4.57	4.73	4.0
Probability	14	4.02	5.31	2.9
Position	22	3.90	4.32	4.3
Mean, one violation		3.95	4.74	4.7
Two violations				
Size & Position	22	3.45	2.97	4.4
Size & Support	16	3.44	4.20	3.9
Probability & Support	18	3.80	4.23	4.0
Probability & Size	21	3.21	5.16	3.4
Mean, two violations		3.46	4.12	3.9
Three violations				
Probability & Size & Support	21	4.28	5.89	4.0
Overall	Total = 247	Mean = 3.70	Mean = 4.62	Mean = 4.25

The distance from fixation was the difference in degrees between the fixation point and the judged center of the cued object. The target objects averaged approximately 2° in height and 1.6° in width. The scenes were 14° in width and 11° in height. The mean distance of a cued object from central fixation was 3.34° (SD = 1.40°).

Object size was measured as the length × width of the longest prominent dimension of a target.

The 10 Violation conditions and one Base condition were approximately equivalent with respect to ratings of their targets' degree of camouflage. The one exception to the equivalence in camouflage ratings across conditions, as shown in Table 2, was the Interposition condition, which had a higher average degree of camouflage. (However, this served to strengthen the result of a lack of an effect of Interposition violations.) Degree of camouflage was defined as the rated degree of masking of a target's critical features by the adjacent contours. Two judges made the ratings on a 10-point scale, from 1 (no camouflage) to 10 (target extremely obscured by adjacent contours). The captions to Fig. 1–3 present some representative values. The raters were encouraged to use the complete scale and they did. The mean (and SD) for Rater 1 was 4.53 (2.25); for Rater 2, 4.01 (2.18). The mean camouflage rating was 4.27. The interrater correlation was .793 (df = 245, $p < .001$). (Reasonably high correlations for camouflage ratings were also obtained in another experiment where, on a somewhat different set of 287 scenes, the interrater correlations among three different raters averaged .70 and the test–retest correlations with a second rating two weeks later averaged .81, $p < .001$ for both r's.) Thus, these ratings of camouflage were reliable. The raters were also instructed to judge camouflage independent of target size by considering the *proportion* of a target's significant contours which was obscured by adjacent contours. They were successful in doing this: the correlation between camouflage and target size was small, −.146 (though significant, $p < .05$, df = 245).

Violation Specifications

Two judges rated the degree to which a given target violated the various relations—from extremely obvious (10) to not present (1). Scenes were selected for the various Violation conditions to produce obvious (mean rating of 8.9) and subjectively equivalent degrees of violation. As described above, if a target was judged to violate Probability or Support, with a rating of 5 or more for that violation, then no rating was entered for Position. (In retrospect, it might have been better to include the position ratings anyway. Even if an object is improbable in a scene, it might be expected to occur in some positions more than others. Similarly, perhaps a floating object is more likely to be floating over some areas of a scene than other areas.)

The violation ratings were highly reliable. Interrater correlations were .873 for Size, .928 for Support, .950 for Interposition and Probability, and .970 for Position.

The largest violation rating for each scene was noted as well as the sum of the violation ratings of all relations. For example, a given scene in the Support & Probability Violation condition might have had a 9 rating on Support, and an 8.5 rating on Probability, and a 1 rating on Size and Interposition. The largest rating would be 9 and the sum would be 19.5. The means for the largest violation rating and the sum of the violation ratings for the scenes in each of the Violation conditions are shown in Table 3.

Ratings for Support and Interposition were strongly determined by physical variations which could be measured on the screen. Height (distance on the screen) above a possible supporting surface was an important factor in the ratings of Support. Interposition ratings were heavily influenced by the amount of contour which appeared through an object.

Procedure

The sequence of events on a single trial in the object detection task is illustrated in Fig. 4. The subject first read the name of the target object from a card in a deck of target cards and,

TABLE 3
Violation Ratings for the 10 Violation Conditions of the Object Detection Experiment

Violation condition	No. of slides	Largest single violation rating (mean)	Sum of all violation ratings (mean)
Single			
Interposition	23	8.8	13.6
Support	27	8.7	12.8
Size	21	8.9	13.9
Probability	14	8.8	13.2
Position	22	9.0	14.8
Double			
Size & Position	22	9.2	21.6
Size & Support	16	9.4	19.3
Probability & Support	18	9.2	19.3
Probability & Size	21	9.7	21.5
Triple			
Probability & Size & Support	21	9.7	27.2
	Total = 205	Mean = 9.1	Mean = 17.6

when ready, initiated the trial by pressing a switch. A fixation point was then presented on a screen for 500 msec and followed immediately by a 150-msec flash of a slide of the scene. The 150-msec presentation duration of the scene was selected so as to be long enough to allow as much processing as possible within a single fixation but brief enough so that the subject could not make a second eye fixation at the scene. The scene was immediately followed by a cue (a dot) embedded in a mask of random-appearing lines. The position of the cue varied from trial to trial but it always appeared at a position at which an object had been centered in the scene. On half the trials, the cue pointed to the object that corresponded to the target name. For example, if the subject was given the target name "fire hydrant" then the cue on such a trial would point to a position on the screen at which there had been a fire hydrant in the scene. The fire hydrant could be in a normal (Base condition) location or undergoing one or more of the violations (Violation conditions). On such a trial, the subject was to say "yes" into a voicekey. On the other half of the trials, the cue pointed to a position

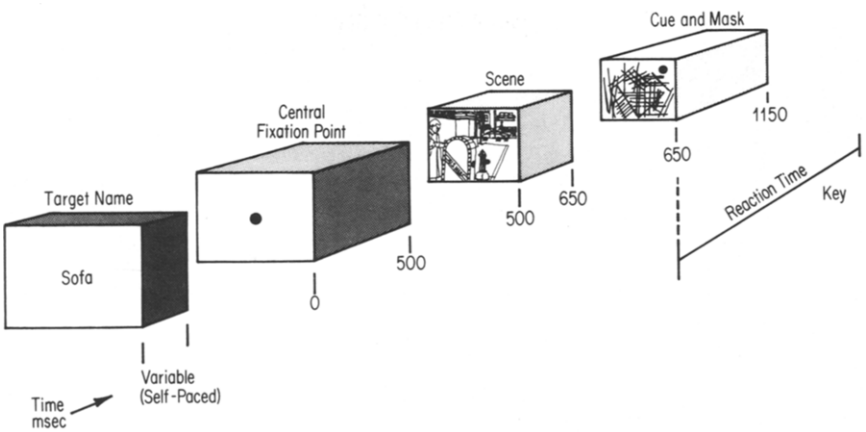

FIG. 4. Sequence of events in the Object Detection task.

at which a different object had occurred in the scene (e.g. a mailbox). On such a trial, the subject was to say "no."

Subjects and Design

Ninety-six subjects, all college students, viewed all 247 slides grouped into 12 blocks of 18 to 22 scenes. The violation conditions, objects, and background scenes were distributed homogeneously across the 12 blocks so that each Violation condition, background scene, and half the target objects, would appear at least once in each block.

For each violation slide there were two possible cues: one designating a base object in the scene and the other designating the violated object. For each cue there were two target–object labels: one naming the cued object (for a correct "yes" response) and the other naming a different object not present in the scene (for a correct "no" response). By including a condition in which an object was in a normal relation to its background while another object in the scene was undergoing a violation, the effects of the presence of violations on an innocent bystander were assessed. Thus, for each violation slide there were four conditions: yes–violation, no–violation, yes–innocent bystander, no–innocent bystander. For each Base slide (where there were no objects in violation) there were two conditions: yes and no. Four decks of target–object cards were made to produce the various conditions. Each base slide–response combination appeared in two of the four decks so as to match the frequency of the four violation scene conditions.

As shown in Table 2, Probability violations were present, either singly or in combination with other violations, in 30% of the slides (74 of the 247 slides). Thus, on 30% of the yes trials, the target label named an object that was improbable in the scene that was to be presented. To eliminate any possible benefits to a strategy where yes responses would be selected once a Probability violation was detected, labels on no trials were selected so as to match the improbable–probable proportions on yes trials. Thus on 30% of the no trials, for both Violation and Base slides, the label named an object that was highly unlikely to occur in that particular scene. For example, the label might have been "fire hydrant," the scene that of a kitchen, and a fire hydrant would not be present (hence some other object would have been cued). Such trials were designated as Improbable–no trials. On the remaining 70 percent of the no trials, the label named an object that would be likely to occur in that setting, e.g. a "frying pan" in a kitchen scene, but which, of course, was not present. Such trials were designated as Possible–no's.

The sequence of the blocks was balanced across subjects and the four decks by two Latin Squares. Half the subjects took the blocks according to one Latin Square, the other half of the subjects by the second Latin Square. Within each Latin Square, one-fourth of the subjects (i.e. 12 subjects) had each of the 4 decks of targets. Half the subjects within each counterbalancing cell took the slides in forward order; the other half viewed the slides in the reverse order. Thus all scenes had the same mean serial position (123.5). Each subject also had 12 practice trials of Violated and Base scenes which were not used in the experiment proper. The task was self-paced; after subjects read the name of the object, they pressed a switch (with the nonpreferred hand) to initiate the trial. Subjects were fully instructed as to the nature of the scenes and violations. They were encouraged to respond "as fast and as accurately as possible."

Slides were presented by four Kodak Carousel projectors fitted with Gerbrands Electronic Tachistoscope shutters. One projector was used for a central fixation point, one for the scene, one for the cue-dot, and one for the mask. Subjects responded verbally into a microphone (Philmore model GM60), which was connected to an audio threshold detector. The signal from the detector stopped a Hewlett–Packard clock from which RTs were recorded.

RESULTS AND DISCUSSION: EXPERIMENT I

The effects of the balancing variables, viz. target-label deck, order, and Latin Squares were negligible. Although there was an overall decrease in error rates and reaction times (RTs) with practice over the 12 blocks, this effect was relatively constant over the experimental conditions described below. Consequently, the data from the different blocks, decks, and Latin Squares were combined to produce mean values for the major variables of interest. (The learning that is evident in this kind of task is of a nonspecific nature in that it shows almost complete transfer to a different set of objects and scene backgrounds. (Teitelbaum & Biederman, 1979). The major results described below were apparent if only the very first occurrence of an object and a background were included in the data analysis.)

The overall error rate was 31.2%, with the miss rate (saying no when the target was cued) far higher than the false alarm rate (saying yes when the target was not cued), 43.2−19.2%, respectively. Mean correct reaction times (RTs) were 999 msec.

Violation Costs

Violation costs were evident in the miss rates in that a target which violated a relation was more likely to be missed than the same target in a Base position (Fig. 5).[4]

The miss rate for the Violation conditions averaged 45.0% as compared to 24.9% in the Base condition, with increased violations (from zero to three) producing higher miss rates $F(3,276) = 72.71$, $p < .001$. As to whether the violation cost on miss rates represented a criterion shift, i.e. responding no if a violation was detected, it is important to note that false alarm rates were also higher, albeit slightly, by 2.7%, when the cued object was undergoing a violation compared to when it was in a base position. As shown in Fig. 5, false alarm rates increased with an increase in the number of violations, $F(3,276) = 5.64, p < .002$. Thus, there was a consistent decline in d' from 0 to 3 violations, 1.62, 1.14, .78, and .54, respectively.

There was no effect, on either miss or false alarm rates, of the presence of a violation on the detection of other objects not undergoing violations (bystanders). This indicates that the violation costs were not due to interference in the elicitation of a frame for the scene but rather because appropriately formed frames either interfered with or did not facilitate the

[4] Although false alarm rates and d's will be presented, the miss rates are emphasized because the effect of the violations was primarily to cause the subject to miss the target. Also, for many of the applications for this research (e.g. to photointerpretation), a miss is a more critical error in that the observer often has sufficient time to take a second look to correct a false alarm. A missed target, however, may not draw a second look.

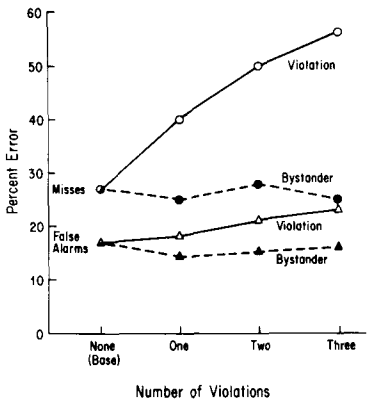

FIG. 5. Mean percentage misses (responding NO when the cued object was the target) and false alarms (responding YES when the cued object was not the target) as a function of the number of violations in the scene and the condition of the cued object. The functions labeled "violation" are the data for cued objects undergoing violations; the functions labeled "bystander" are the data when the cued object was in a normal position but some other object in the scene was undergoing a violation.

perceptibility of objects undergoing violations. In an object detection task which included a condition where objects were presented alone (without any context), as well as in Base and Violation conditions, Klatsky, Teitelbaum, Mezzanotte, and Biederman (Note 2) found evidence for both interference effects (viz. violation costs), as well as slight facilitation effects from the Base condition. The evidence for facilitation was that objects that appeared in a Base Condition, when uncamouflaged, could be more readily detected than objects that appeared alone.

False alarm rates were lower when the target object (not the cued object) was improbable in the scene compared to when it was probable, as shown in Fig. 6. This result replicates a finding reported by Biederman, Glass, and Stacy (1973), in a search task with photographs of scenes. Thus, subjects were less likely to false alarm with "truck" as a target object in a kitchen scene than when the target object was "frying pan."

Violations Costs and the Effects of Physical Parameters

The expected psychophysical effects held, in that the further an object was from fixation, the smaller its size, or the greater its camouflage, the more likely it was to be missed. The Pearson r's ($df = 245$) between miss rates and distance was .398 ($p < .001$), miss rates and Size (length $\times$ width) was $-.497$ ($p < .001$) and miss rates and Camouflage Rating was .151 ($p < .001$). The multiple R was .605 ($p < .001$) between these three

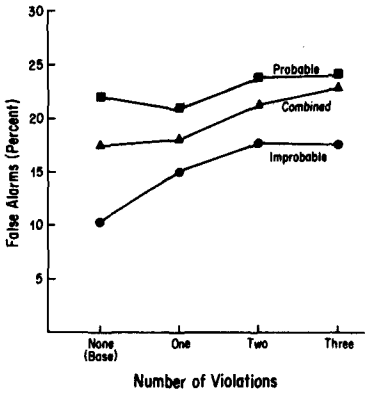

FIG. 6. Mean false alarm rate as a function of the number of violations and target likelihood.

variables taken together and miss rates. The violation costs along with the effects of distance from fixation and target size are shown in Fig. 7.

It is evident that the violation costs were incurred for large as well as small targets, and—importantly—were incurred when subjects were looking at the cued objects as well as when the cued objects were several

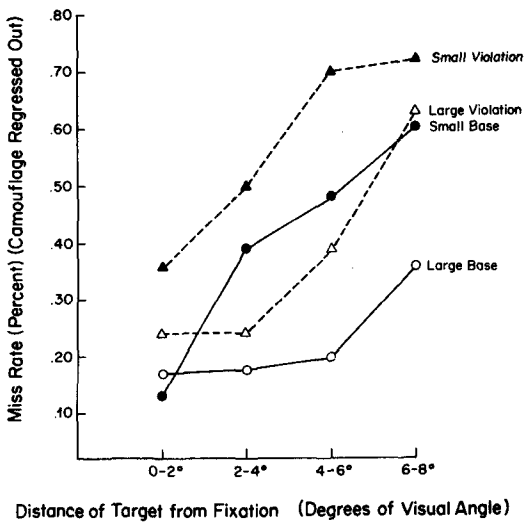

FIG. 7. The effects of distance of cued object from fixation, size of cued object, and Violation Condition on miss rates with camouflage regressed out.

degrees removed from fixation. The Klatsky et al experiment (Note 2), in which the cue served as the fixation point by being presented prior to the scene, confirmed the existence of a violation cost even when subjects were looking directly at the cued object.

Individual Violation Conditions

The miss rates for the Base and 10 Violation conditions are shown in Fig. 8. Although there was considerable variability among the slides in these conditions, with the exception of the Interposition violation, violation costs were evident for all the relations. A considerable portion of the variability can be reduced by correcting these data for the effects of distance from fixation, size, and camouflage, as well as for the specific objects used in the various conditions. The correction for size, distance, and camouflage was done by performing a regression analysis with these variables as predictor variables and then using the residuals as the corrected scores. To remove effects such as prototypicality and quality of the depiction of the individual objects that comprised a given violation condition, differences between Violation and Base residual miss rates were calculated. Thus, for example, if a truck was one of the objects in the Support violation condition, then the residual miss rate for the truck when it was in the Base condition was subtracted from the residual miss rate when the truck was undergoing the Support violation. This was done for all the objects in all the violation conditions. The results from this analysis, the mean residual difference scores, are shown in Fig. 9.

Although regression effects led to some shrinkage of the violation costs,

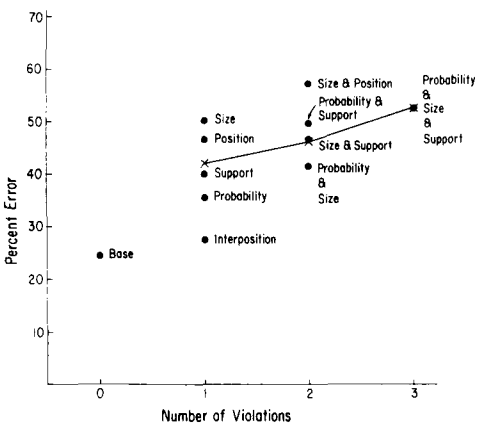

FIG. 8. Mean percentage misses in detecting a cued target as a function of the number and kind of violated relations. The X's and the line connecting them are the mean miss rates for the Size, Support, and Probability conditions. These conditions were run under all three levels of violation.

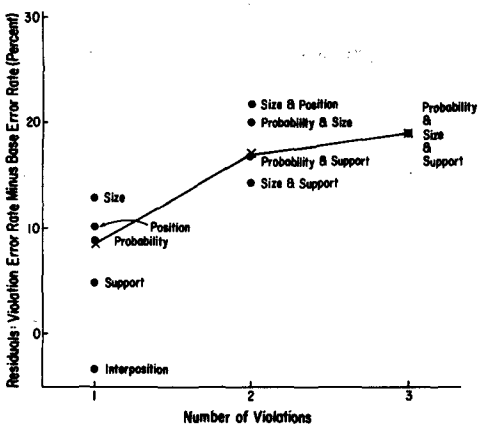

FIG. 9. Differences (violation minus base) in the residual miss rates. Distance from fixation, size, and camouflage of cued object have been regressed out. Positive values indicate that the objects in a given Violation Condition had a higher residual miss rate than those same objects when they were in a Base Condition.

the general picture of these data was highly similar to the uncorrected data presented in Fig. 8, and the 12.6% greater miss rate for the violation scenes remained significant; $t(205) = 7.13$, $p < .001$. (Over the various experimental conditions, the standard errors of the differences in residual miss rates ranged from 3.4% for Interposition to 6.8% for Size & Position.) From Figs. 8 and 9, it is evident that violations of the physical relations of Support and Interposition were *not* more disruptive than violations of the semantic relations. In fact, there was no violation cost from violations of Interposition. Also, violations of Position and Probability yielded nearly equivalent costs. It should also be noted that the addition of a violation of a semantic relation of Probability or Size to the violation of Support resulted in an increase in miss rates compared to the single violation of the Support relation.

This picture does not substantially change when the false alarm rates are included in the calculation of violation costs. The mean d' values calculated from the residual difference scores from the miss and false alarm rates are presented in Table 4. (The mean hit and false alarm rates from the Base condition were added to all scores to maintain the original performance levels.) The physical violations were not more disruptive on object detection than the semantic violations; the mean d' value for Support and Interposition was 1.48, as compared to .98 for the 3 semantic violations. Violations of Probability were less disruptive than violations of Position, with d' values of 1.42 and .98, respectively. Particularly striking was the extremely low detectability of objects undergoing Size violations; such objects had a d' value of only .61, the lowest of any condition.

TABLE 4
Mean False Alarm Rates and d' Values for Each of the Experimental Conditions

Condition	FAR	d'
Zero violations		
Base	.16	1.70
One violation		
Interposition	.13	1.77
Support	.16	1.24
Size	.27	.61
Probability	.15	1.42
Position	.17	1.05
Mean, one violation	.18	1.22
Two violations		
Size & Support	.23	.87
Size & Position	.17	.77
Probability & Support	.18	.96
Probability & Size	.19	1.11
Mean, two violations	.19	.93
Three violations		
Probability & Size & Support	.21	.76

Although Figs. 8 and 9 reveal a general trend for multiple violations to yield higher violation costs than single violations, departures from monotonicity were apparent. Thus, both figures show that the error rate for the triple violation condition was approximately equivalent to the error rate for the double violation condition of Size & Position. Also, the mean d' value for the individual violation of Size was the lowest of all the violation conditions, including the triple violation condition. Some of these departures from a monotonic relation between error rates and the number of violations might be reduced if a speed-for-accuracy trade-off was operative. As will be shown in the next section, RTs for the Size & Position condition were markedly shorter than the other double violation conditions and the RTs for the triple violation condition were slower than any of the other violation conditions.

Reaction Times

The mean correct RTs for the individual experimental conditions are shown in Fig. 10. Data were included only from those scenes in which at least six correct RTs were recorded. Thirty-six scenes, all violations, were eliminated by this criterion from the remaining analysis.

The time required for the correct detection of objects undergoing a violation was 31 msec longer than that required for the detection of ob-

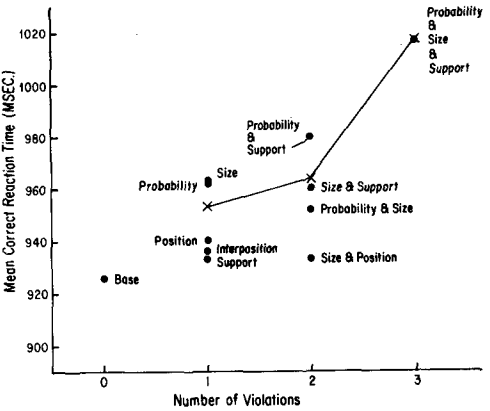

FIG. 10. Mean correct reaction times for detecting a cued target (i.e., RTs for Hits) as a function of the number and kind of violated relations. The solid line connecting the X's are the mean miss rates for the Size, Support, and Probability conditions only.

jects in a Base condition. This underestimated the violation cost, in that the 36 scenes which were removed were violations which tended to be of greater difficulty (as defined by the predictor variables) than the 211 scenes not excluded. As with the miss rates, some of the within-condition variability can be eliminated through a regression analysis (correcting for differences in fixation distance, size, and camouflage), and by presenting the data in terms of difference scores, as shown in Fig. 11.

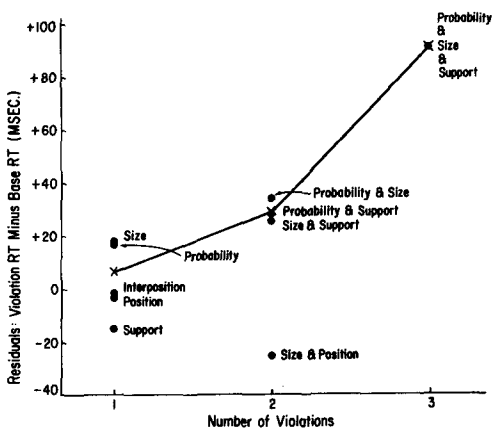

FIG. 11. Differences (violation minus base) in residual RTs. See the legend to Fig. 10 and text for details of the regression analysis.

In general, the RT data revealed a similar ordering of conditions as the error data, thus precluding a general speed-for-accuracy trade-off account for the miss rates. Thus, violations of the physical relations did not yield larger violation costs (on RTs) than violations of the semantic relations and there was an apparent increase in the violation costs as the number of violations increased. The one exception to this was the relatively fast RTs to the Size & Position violation condition. However, as shown in Figs. 8 and 9, that condition had a relatively high miss rate. Thus, a speed-for-accuracy trade-off correction would have brought the miss rate for the Size & Position violation condition more in line with the other double violation conditions.

A full accounting of why an increase in the number of violations lead to greater violation costs is beyond the scope of this experiment. At least two nonexclusive possibilities suggest themselves. With more violations present, the more likely one misleading relation will be registered before an object is fully identified. This could occur with either serial or parallel processing of the different relations. The misleading relation could serve to decrease the plausibility of the target or it could propose an incorrect candidate object. The second possibility is that when more than one violation is registered, more incongruity is produced and, perhaps, the plausibility of the cued object is reduced below what it would be with only a single violation. In addition, stronger or more incorrect candidates could be proposed with more violations. By either account, information about an object's relations to its setting is held to be available before the target is identified.

EXPERIMENT II: VIOLATION DETECTION

What are the relative detection speeds of the violations themselves? Experiment II employed an acceptability judgment task in which subjects judged whether a given target object was undergoing any of the violations. The bottom-up model implies that violations of Interposition and Support would be detected more rapidly than violations of Probability, Position, or Size. Furthermore, the addition of a semantic violation to a physical violation should not render the combined violation detectable any faster than the physical violation by itself. It would imply, for example, that the detection of the incongruity of a fire hydrant floating in a kitchen would not be any faster than the detection of that same hydrant when it was floating in a street.

The detection of the Position violation in this experiment provides an intuitively acceptable criterion of scene comprehension. For one to judge accurately that a hydrant does not belong on top of a mailbox requires not only that the various objects be identified but that our knowledge about the acceptable relations of those objects in that setting be accessed. Thus,

if judgments of Position violations can be accurately made from a single fixation at a scene, then the above definition of scene comprehension would imply that the scene was comprehended from the results of a single fixation.

Method

The sequence of events in a trial on the Violation Detection task, as shown in Fig. 12, was similar to that of the object detection task, except that the positional cue preceded the scene and the object cued *always* corresponded to the target name. So, if the target name was "fire hydrant," the object that was cued would always be a fire hydrant. Thus the subject knew where to look and what object was to be judged before the scene was presented.

The subject responded with one microswitch finger key, marked "normal," if the target object was in a base setting and another key, "violation," if it was violating any or several of the five relations. Subjects were instructed as to the nature of the relational violations and shown several examples of each type. Forty-eight subjects each viewed 277 scenes, only 246 of which were included in the data analysis. The 31 extra scenes were used to provide more base objects to increase the proportion of normal responses.

Target labels were presented on a Sorac display terminal controlled by an Automatic Data Systems 1800E minicomputer. The computer also stored the response data and provided speed and accuracy feedback to the subject after each trial. The cues, scenes, and mask were presented by a slide projector as in Experiment I. After reading the target label the subject would look up at the screen for the presentation of the cue, scene, and mask.

Results and Discussion

With the exception of the Interposition violation, subjects were able to detect the violations within a single glance. The overall hit rate (detecting the presence of a violation) was 88%. The false alarm rate (responses with the violation key when no violation was present) was 10.3%. Correct RTs averaged 851 msec. The fact that subjects can do this task so well given

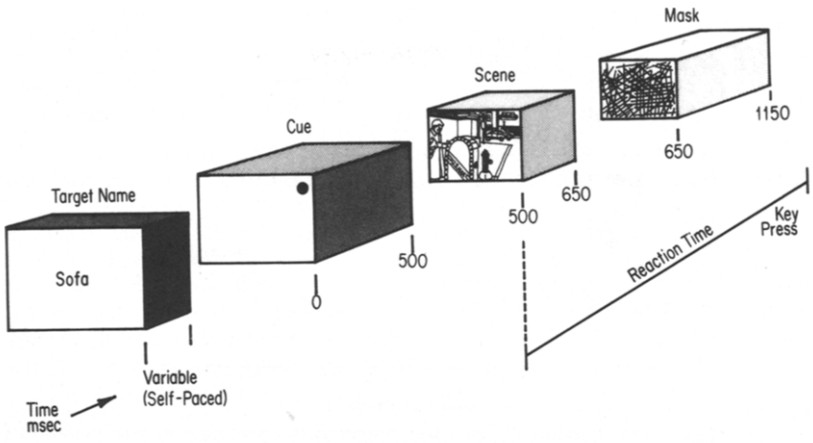

FIG. 12. Sequence of events in the Violation Detection task.

only a 150 msec presentation of a picture of a scene clearly demonstrates that semantic relations can be accessed from a single glance at a scene.

Figure 13 shows the RTs and Fig. 14 the miss rates for the Base and 10 Violation conditions. Camouflage was the only physical variable that significantly correlated with RTs and errors, r's = .318 and .319, respectively, $p < .001$, $df = 244$. With precuing, the actual size of the target and its distance from the center of the screen were uncorrelated with either RTs or errors. The residual differences in RTs and errors are shown in Figs. 15 and 16. These data are corrected for camouflage and the target objects across the Violation conditions. As in Experiment I on object detection, for both the original and corrected data no evidence was found for a consistent advantage in the accessibility of the Support and Interposition relations over the Size, Position, and Probability relations. In fact, the Interposition violation had a much higher miss rate than the other violations. Also, Probability violations were not more readily detected than Position violations.

As the number of violations increased, there was a suggestion of a redundancy gain (Biederman & Checkosky, 1970), in that the speed and accuracy of violation detection generally increased.

Redundancy gains can be used in determining whether several

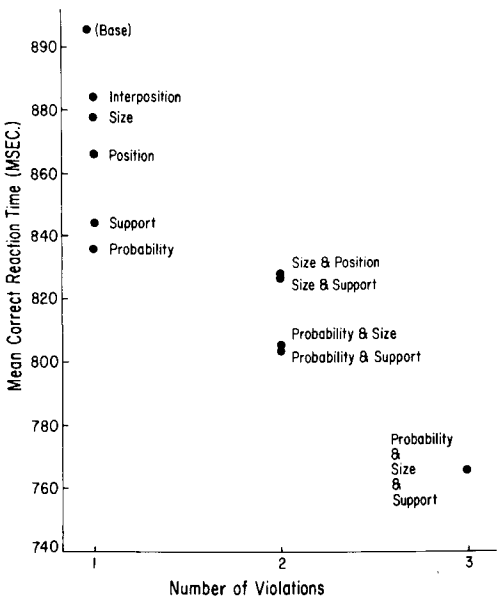

FIG. 13. Mean correct reaction times for detecting the presence of any violation as a function of the number and kind of violated relations. The "Base" condition is for the correct responses when a target object did not violate any relations.

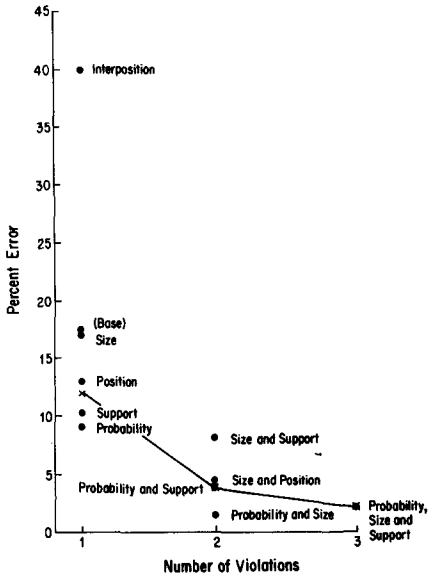

FIG. 14. Mean percentage error for detecting the presence of any violation as a function of the number and kind of violated relations. The "Base" condition is for the errors when a target object did not violate any relations. The solid line connecting the X's are the means for the Size, Support, and Probability conditions only.

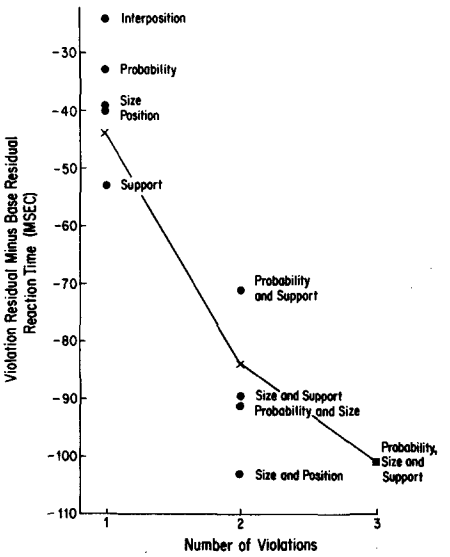

FIG. 15. Mean correct residual RT differences (violation minus base), with camouflage regressed out, in the Violation Detection task. Negative values indicate that the RTs for detectable violations for a given set of objects in a given Violation Condition were faster than the RTs to judge those same objects when in a Base Condition. The more negative the values, the faster the RTs relative to the Base Condition.

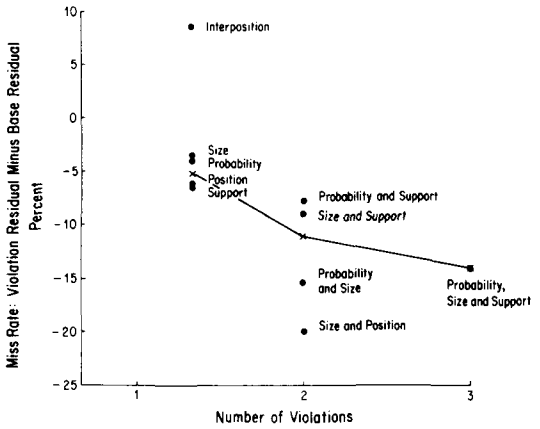

FIG. 16. Mean residual miss rates (violation minus base), with camouflage regressed out, in the Violation Detection task. See legend to Fig. 15 and text for additional interpretation.

components—violations, in the present case—are processed sequentially or simultaneously (in parallel). For instance, a redundancy gain can rule out a sequential fixed order model—the bottom-up model—which holds that Support and Interposition are processed before Probability, Position, and Size. Such a model would imply no gain in speed when a violation of a semantic relation, say Probability or Size, was added to a violation of Support. But the results show that violations of Probability & Support together and Size & Support together were detected faster (and more accurately) than violations of Support alone.

A redundancy gain is compatible with both a varying order sequential detection model and a parallel detection model. The sequential model predicts a redundancy gain by holding that the greater the number of relational violations in a scene, the more likely that one of those would be processed first (or earlier) in the sequence. The parallel model predicts a redundancy gain under the assumption that the different relations are processed concurrently with times that are not perfectly correlated. If there is overlap in the distribution of times, then the greater the number of violations actually present, the more likely that, on a given trial, a single one would be quickly detected. This parallel model can be likened to a horse race in which all the components (horses) start simultaneously but the greater the number of the horses, the more likely by chance alone that one will have a fast race. This experiment was not designed to distinguish between these models, but it should be noted that neither of them posits greater accessibility of the physical over the semantic relations. Accessibility, according to the varying-order, sequential model, would be the earlier processing of a relation. If Support were

always processed prior to the semantic relations, then the varying-order, sequential model reduces to the fixed-order, sequential model and no redundancy gain is predicted. Greater accessibility in the parallel model would result from one of the components being faster than the others. But if Support always won the race, then no redundancy gain would have resulted.

GENERAL DISCUSSION

Implications of the specific results of these experiments will be considered first. The status of the relational violations will then be discussed.

Violation Costs

Objects undergoing a violation were harder to see than objects in a Base condition. A response bias or sophisticated guessing explanation could not readily account for these data. Such an explanation might hold that subjects tended to respond or guess no when they could not detect a target but did, somehow, detect a violation at the cued location. That is, they would respond no when they realized that a given blob, if it were a fire hydrant, would be improbable in the scene they were looking at, or would be floating, etc. Such a strategy would, indeed, produce a higher miss rate for targets undergoing a violation but this guessing explanation also would predict that the presence of a violation should *reduce* the false alarm rate. However, false alarm rates were slightly higher for objects undergoing a violation than they were when the objects were in the Base condition.

The reduced perceptibility of targets undergoing a violation does not necessarily stand in contradiction to those studies showing earlier eye fixations during free scanning of pictures to targets placed in low probability contexts (Loftus & Mackworth, 1978; Friedman, 1979). Targets undergoing violations are harder to perceive but once perceived they are likely to be what is interesting about a scene. Longer visual dwell times and better recall would then be expected.

Schema Activation

The cued objects in this experiment were readily identifiable without context (Klatsky et al., Note 2). Moreover, since the cued objects were often undergoing violations, it might have been in the subjects' best interests to simply ignore the context. That under these conditions a violation cost was obtained underscores the rapid—perhaps obligatory—activation of a schema, as evidenced by the effects of the semantic relations. The absence of violation effects on innocent bystanders indicates that the elicitation of schemata for these scenes were not disrupted by the presence of an object undergoing violations. Instead, the violation costs were

incurred by well-formed schemata interfering with (or not facilitating) the identification of objects.

Independent evidence for schema activation came from an analysis of the effects of target probability on false alarm rates (Fig. 6). Subjects were less likely to false alarm to targets which were improbable in the scene than targets which were probable. This result is similar to one reported by Biederman et al. (1973) in a visual search task. Unlike the perceptual effect of the violation cost discussed in the preceding paragraph, however, this particular effect of target (not cued object) probability on false alarm rates could be a response bias (sophisticated guessing) effect.

A schema for a scene, defined by interacting relations among the objects, would be expected to yield effects across the visual field. That violation costs were obtained both at the area of central fixation as well as several degrees removed from fixation documents this aspect of a schema. It is possible that some of the subjective intelligibility of peripheral vision is due to the extended nature of a schematic representation. That is, the schema can bias the interpretation of objects outside the area of central fixation.

Physical vs Semantic Violations

Violations of semantic relations were at least as disruptive and at least as detectable as violations of Support and Interposition. Moreover, the addition of a violation of a semantic relation to a violation of Support tended to result in a greater violation cost in object detection and better violation detection than just the Support violation by itself. Neither of these results is readily compatible with a model of scene perception which would hold that physical parsing precedes the interpretation of semantic relations. Instead, semantic relations appear to be accessed at least as quickly as relations which can be defined by physical parameters, viz. height (on the screen) above a supporting surface or the amount of inappropriate background detail in an Interposition violation. Actually, the lack of a violation cost for Interposition and its lack of detectability suggest that it becomes available relatively late in processing. Instead of a 3D parse being the initial step, the pattern recognition of the contours and the access to semantic relations appear to be the primary stages. In this respect, the detection of violations of Support may simply be a special case of the detection of violations of Position in real-world scenes.

Probability vs Position and Size Violations

It is of some interest that violations of Probability were not more disruptive than violations of Position or Size. As described previously, violations of Probability require only inventory information; once a kitchen, a stove and a frying pan are identified, a fire hydrant could suffer from its

improbable inclusion in such a setting. Its position relative to the other objects need not be determined. If, however, Position relations could only be determined after objects were identified, then violations of Position might be expected to be less disruptive and less detectable than violations of Probability. That this did not occur is further evidence that an object's semantic relations to other objects are processed simultaneously with its own identification. The large cost associated with Size violations reinforces this point.

Innocent Bystanders

No innocent bystander effect was obtained. This result is somewhat inconsistent with the disruptive effects on object identification reported by Biederman (1972), and Biederman et al. (1974). Our guess is that the resolution to this apparent discrepancy will be with the number of objects undergoing violations in a scene. With the jumbling operation, a large proportion of the objects in the scene would be undergoing Support, Position, and Size violations. In the current experiments, violations were applied to only a single object. These scenes remained well formed, with the objects undergoing violations interpretable by visual metaphor such as Goodyear sofa for Fig. 3. Schema-plus-correction and weird list are terms that have been used to describe such anomalies. But as the number (or proportion) of objects undergoing violations in a scene increases, at some point metaphor fails and the scene no longer appears to be integrated. It no longer is a scene, but instead resembles a display of unrelated objects (Biederman, 1981). Perhaps it would be at this point that innocent bystander effects are obtained. We are currently exploring this possibility.

There are some scenes, such as junkyards or some store display windows, where many of the relations are typically violated. By the preceding account, such scenes should be more difficult to process, perhaps behaving like jumbled scenes. In a sense, they are analogous to sentences such as "The words that I used in my memory experiment were: ashtray, justice, tree, shallow, glove, tuna, chalk, train, fatigue, newspaper, mission, rapid."

The lack of an innocent bystander effect poses obstacles to attempts at using recognition memory as a measure of the encoding of a scene. Thus, Friedman (1979) demonstrated that an improbable object in a scene is readily remembered—to the detriment of the other objects in the scene. This recognition memory result would, superficially at least, appear to be inconsistent with the detection results of Experiment I, where improbable objects were themselves more difficult to detect but their presence did not affect the detectability of other objects in the scene. The resolution of this apparent inconsistency is relatively straightforward. Improbable objects, while difficult to detect initially, once perceived can be what is interesting

about a scene. Consequently, such objects will tend to attract thought and memory elaboration in Von Restorf fashion.

A similar problem exists in the attempt to infer scene encoding from the duration of eye fixations (Friedman, 1979). Although part of the increase in fixation duration for low probability objects may reflect increased encoding difficulty, the greater portion of the increase would appear to reflect, again, interest. Thus, Friedman (1979) reported that initial fixations onto improbable objects averaged 650 msec—350 msec longer than when an object was rated as being likely in the scene. By that estimate, an additional 350 msec would be required to see an improbable object. Although improbable objects require more time to perceive, 350 msec would appear to be an overestimate of the magnitude of this effect. Similarly, 300 msec—the duration of a fixation onto a likely object in Friedman's study—overestimates the amount of time required to detect an object in a normal position. If, instead of the subject controlling the duration of viewing a display through eye fixations, scenes are presented for brief periods of time, improbable objects can be detected at presentation durations far briefer than the estimates derived from fixation dwell times. Thus, Klatsky et al. (Note 2) found that when positions of to-be-detected objects were precued (by presenting the dot prior to the scene), objects could be detected at an accuracy rate of 90% from a presentation duration of only 100 msec. Even when not looking directly at an object, Fig. 8 shows that improbable objects can be detected on 75% of the trials from a 150 msec presentation of a scene. More likely, much of the increased dwell time on improbable objects in Friedman's (1979) experiment reflects processes which are occurring *after* the object is identified. The difficulty of disentangling the effects of these postperceptual processes from perception presents formidable obstacles to the use of fixation dwell times and recognition memory as measures of scene perception.

Plausibility and the Violations

When evaluating the psychological reality of the five relations, the status of plausibility should be considered. This is the possibility that the violations of the five relations affect some mechanism which evaluates the plausibility of an arrangement of objects in perceiving the scene. For example, violations of Size or Interpositions or Probability would all reduce the plausibility of an object's relation to its setting and thus, perhaps, interfere with its perceptibility. There is no fundamental incompatibility between the relations and plausibility, but plausibility would introduce—perhaps unparsimoniously—another stage of processing. Given that a plausibility generator was shown to be a needed stage, then the five constraints might be ways of conveniently summarizing ways in which plausibility might be affected. A major research question would then be

how such disparate methods for producing violations affect a given mechanism (the plausibility generator) in the same way.

Actually, the plausibility generator may explain very little. Given that the scenes presented to the retina are novel, general mechanisms are still needed with which to generate plausibility values from novel inputs. The previous linguistic examples, "He smiled the baby" and "The angry napkin" will furnish a convenient analogy with which to appreciate this issue. Subjects can readily detect the syntactic violation of transitivity in the former sentence or the semantic violation of agreement in animatedness in the latter (Moore, 1972; Moore & Biederman, 1979). It might be the case that these violations affect some sentence plausibility generator in identical ways, despite the obvious differences in the ways in which the violations were produced. Again, as it was for scenes, the critical research question must then be directed toward the plausibility generator. If plausibility does play a role, it might be that the plausibility values are determined by the various constraints which characterize the organization of scenes (or sentences). Rather than plausibility being the explanatory basis, the plausibility values themselves might be explainable in terms of volations of the relational constraints.

GENERAL CONCLUSIONS

These results allow a sketch of an alternative model to the bottom-up, depth-then-object-identification-then-semantic-relations model. This model would be one where the processing of contours proceeds at least simultaneously with the processing of depth relations. But the contours are not only informative about the objects. They are also informative about the possible interactions that these objects might enjoy with other objects. Support for this notion comes from the increased perceptibility found for degraded objects when those objects are brought together to form a scene (Biederman, 1981).

Although there can be no doubt that movement of the eye, observer, or elements of the scene can be critically important in conveying information about contour, the evidence for rapid comprehension of the scenes in the present experiments where such movement was impossible, calls into question those accounts which assign primacy in perception to the role of movement (Gibson, 1979). We simply do too well in the absence of movement for it to be the fundamental principle upon which scene perception depends.

On an empirical level, the results of these experiments show that semantic relations, defined by the specific ways in which objects typically interact in the visual world, are accessed from a 150 msec presentation of a picture of a novel scene. So rapid and efficient is this access that not only can the violations be detected readily from a single glance, but

the perceptibility of objects undergoing violations of these relations will be impaired.

REFERENCES

Biederman, I. Perceiving real-world scenes. *Science*, 1972, **177**, 77–80.

Biederman, I. On processing information from a glance at a scene. Some implications for a syntax and semantics of visual processing. In S. Treu (Ed.), *User-oriented design of interactive graphic systems*. New York: ACM, 1977.

Biederman, I. On the semantics of a glance at a scene. In M. Kubovy & J. R. Pomerantz (Eds.), *Perceptual organization*. Hillsdale, NJ: Erlbaum, 1981.

Biederman, I., & Checkosky, S. F. Processing redundant information. *Journal of Experimental Psychology*, 1970, **83**, 486–490.

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

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

Bruner, J. S., & Potter, M. C. Interference in visual recognition. *Science*, 1964, **144**, 424–425.

Friedman, A. Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General*, 1979, **108**, 316–355.

Gibson, J. J. *The perception of the visual world*. Boston: Houghton Mifflin, 1950.

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

Gibson, J. J. *The ecological approach to visual perception*. Boston: Houghton Mifflin, 1979.

Hock, H. S., Romanski, L., Galie, A., & Williams, C. S. Real-world schemata and scene recognition in adults and children. *Memory & Cognition*, 1978, **6**, 423–431.

Julesz, B. Figure and ground perception in briefly presented iso-dipole textures. In M. Kubovy, & J. R. Pomerantz (Eds.), *Perceptual organization*. Hillsdale, NJ: Erlbaum, 1981.

Loftus, G. R., & Mackworth, N. H. Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance*, 1978, **4**, 565–576.

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., & 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.

Marr, D. Representing visual information. In E. M. Riseman & A. R. Hanson (Eds.), *Computer vision systems*. New York: Academic Press, 1978.

Metelli, F. The perception of transparency. *Scientific American*, 1974, **230**, 90–99.

Moore, T. E. Speeded recognition of ungrammaticality. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 550–560.

Moore, T. E., & Biederman, I. Speeded recognition of ungrammaticality: Double violations. *Cognition*, 1979, **7**, 285–299.

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

Rumelhart, D. C. *Introduction to human information processing*. New York: Wiley, 1977.

Teitelbaum, R. C., & Biederman, I. Perceiving real-world scenes: The role of a prior glance. *Proceedings of the Human Factors Society*, 1979, **23**, 456–460.

Waltz, D. Understanding line drawings of scenes with shadows. In P. H. Winston (Ed.), *The psychology of computer vision*. New York: McGraw–Hill, 1975.

Winston, P. H. Learning structural descriptions from examples. In P. H. Winston (Ed.), *The psychology of computer vision.* New York: McGraw–Hill, 1975.

REFERENCE NOTES

1. Guzman, A. *Computer recognition of three-dimensional objects in a visual scene* (Project MAC Tech. Rep. 59) Cambridge, MA: MIT Artificial Intelligence Laboratory, December 1968.
2. Klatsky, G. J., Teitelbaum, R. C., Mezzanotte, R. J., & Biederman, I. *Evidence for mandatory processing of contextual information in real-world scenes.* Paper presented at the Meetings of Eastern Psychological Association, Hartford, CT, April 1980
3. Biederman, I. *Background depth gradients and object detection.* Unpublished manuscript, State University of New York at Buffalo, 1980.

(Accepted April 1, 1980)