Object Search in Nonscene Displays

Irving Biederman, Thomas W. Blickle, Richard C. Teitelbaum, and Gary J. Klatsky
State University of New York at Buffalo

When we look at a chair or a giraffe we cannot suppress a semantic interpretation of that image, although we need not name it (e.g., Smith & McGee, 1980). Given that classification of object images is mandatory, is it capacity free? Subjects attempted to detect the presence or absence of a target object, specified by basic-level name, in a 100-ms display of a nonscene (clock face) arrangement of one to six pictures of common objects. There was a sharp monotonic decrease in detectability as a function of the number of objects in the display, indicating that object detection under these conditions is an attention-demanding process. No benefit was observed for targets that were likely to co-occur with the distractors. This latter result is evidence against an account of the perceptual interference found for improbable objects in real-world scenes, which holds that the interference derives from an inventory listing of the objects without regard to their spatial relations.

It is a common, subjective impression that object perception is mandatory; indeed it is. It is impossible to look at a chair or a dog and not identify it. We cannot “turn off” the processes leading to a semantic interpretation of an object, although the elicitation of the object’s name is a relatively slow and effortful step. Evidence supporting these impressions was reported by Smith and McGee (1980) in their demonstration of Stroop-type interference on the superordinate classification of words in the presence of conflicting pictures. For example, the classification reaction times (RTs) of the word shirt as an article of clothing were lengthened when the word was superimposed on a picture of a chair. Additional support for the obligatory processing of pictures derives from the conceptual masking phenomenon investigated by Potter (1976), Intraub (1984), and Loftus and Ginn (1984).

This impression of mandatory processing in object perception extends to the perception of arrangements of objects in real-world scenes. With a scene, more than one object is typically present. Yet at a glance, many of the objects, their relations, and the nature of the setting, can be determined (e.g., Biederman, 1981; Biederman, Mezzanotte, & Rabinowitz, 1982; Biederman, Teitelbaum, & Mezzanotte, 1983; Intraub, 1984). As with individual objects, we cannot seem to suppress the interpretation of the visual arrays that comprise a scene.

Although there is a strong reason to believe that our interpretation of a scene is mandatory in the sense that we cannot suppress the interpretation, it has never been determined whether the perception of multiobjekt arrays is capacity limited. We use the term capacity limited in the operational sense proposed by Shiffrin and Schneider (1977), as a condition under which RTs and error rates for the recognition of a target item (typically a letter or digit) increase as a function of the number of other items (distractors) in the display. Shiffrin and Schneider identified mandatory processing with unlimited capacity, in which no increase in RTs or error rates is observed with increases in the number of distractors. Mandatory processing was observed to be a consequence of the extensive practice, with a consistent stimulus-to-response mapping that led to automaticity.

Our experiment explored the display size effect with pictures of common objects. The pictures were readily identifiable with a basic level name. It could be that pictures function as alphanumerics or conjunctions of analyzable features, showing monotonic increases in RTs or error rates with increasing numbers of distractors unless practice promoting automaticity is accomplished. The monotonic (often linear) effect of the number of distractors in the visual field has been found to hold for a wide range of objects that cannot be distinguished from the distractors on the basis of a single visual feature (Egeth, Jonides, & Wall, 1972; Shiffrin & Schneider, 1977; Treisman & Gelade, 1980).

But there are both perceptual and cognitive reasons why the detection of objects might not show an effect of display size. Perceptually, when targets can be distinguished by a single feature, such as color, shape, or size, target detection latencies have been found to be independent of the number of distractors even when a consistent mapping of target to response has not been extensively practiced (e.g., Treisman & Gelade, 1980).

More relevant to our present interest in object perception is the discovery by Pomerantz, Sager, and Stoever (1977) that there are aspects of the arrangements of contours that might allow objects to be detected on the basis of emergent features. For example, in Pomerantz et al.’s experiment, the detectability of a segment of different slope in a display of segments was made to be independent of the number of distractors by the addition of an L pattern to all the segments. The Ls

This research was supported by the Air Force Office of Scientific Research Grant 86-0106.

We thank Janet Stoeckert for testing many of the subjects, Mary Lalonia for constructing the stimuli, and Robert Bennett and Robert Mezzanotte for their assistance in two preliminary studies. James R. Pomerantz and Marilyn C. Smith provided helpful comments on an earlier version of this article.

Richard C. Teitelbaum is now with the International Business Machine Corporation. Gary J. Klatsky is now with the General Electric Corporation.

Correspondence concerning this article should be addressed to Irving Biederman, who is now at the Department of Psychology, University of Minnesota, Elliott Hall, 75 East River Road, Minneapolis, Minnesota 55455.
converted the discrimination from one of segment slope to one in which a triangle had to be detected among a field of arrows. In this case the more complex stimuli (arrows vs. triangles) not only showed an advantage over the individual segments but also were independent of the number of distractors, quite unlike the identification of stimuli requiring the processing of conjunctions of analyzable attributes such as searching for a red T among blue Ts and red Os (Treisman & Gelade, 1980). This benefit of complexity is also found in the latencies for identifying objects. Objects missing some of their parts require more time for their identification than their complete counterparts (Biederman, 1987b). Yet complex objects, defined as those requiring more than five convex volumes to look complete (e.g., nine in the case of some airplanes), can be identified with shorter latencies than simple objects, such as a cup, that can look complete with only two volumes (Biederman, 1987b).

Cognitive considerations also support the possibility of capacity-free identification of objects. There is a consistent mapping of an arrangement of the components of an object to meaning categories. Thus, every time we see a chair, we tend to think of it as a chair—never as a banana, an elephant, or a car—although we rarely say “chair.” This may be part of the reason why the elicitation of meaning is mandatory from an object but not its name (Smith & McGee, 1980). The potential importance of the consistent mapping of an object’s contours to a semantic category (but not necessarily its name) remains even if the automaticity-controlled distinction is regarded as the extrema of a continuum of response competition (Cheng, 1985). Consequently, we might expect to find unlimited capacity (or no response competition) when detecting objects.

The Probability Violation Effect

A secondary goal of the present experiment was to assess partially the origins of the probability-violation effect (PVE). An object that is unlikely to occur in a given real-world scene is less perceptible (has a lower interpretability) than when it is in a normal relation to its context (Biederman, Mezzanotte, & Rabinowitz, 1982; Biederman et al., 1983; Klatsky, 1983; Malcus, 1983; Murphy & Wisniewski, 1987). For example, a fire hydrant is less detectable when pictured in a kitchen than when on a sidewalk in a street scene. The PVEs occur at scene presentation durations so brief that only a single visual fixation is possible.

How is this quick access to the probabilistic relations between an object and its setting achieved? A straightforward possibility, described by Biederman (1981), is that the subject first identifies one (or more) of the more discriminable (larger, more central, less camouflaged, more familiar, etc.) objects in a scene. Although this processing of the first object(s) would be independent of the processing of other objects, it could bias the processing of the remaining objects, independent of their relations to the other objects. This could occur either through an inference as to the general class of settings (If I am seeing a refrigerator then it is likely that I am looking at a kitchen and object x must be a stove and object y a pot) or directly from one object to another (If I am seeing a refrigerator then x must be a stove and y a pot). In either case, this route to the probabilistic information about a scene assumes that once an object is identified, we quickly know the kind of company it keeps. There is no doubt that people can make such inferences. Indeed they can do it from a list of object names, such as those shown in the Appendix. The issue is whether such inferences are responsible for the PVE in scenes. Note that this route posits that probabilistic information can be accessed without any determination of the spatial relations among the objects. In the previous characterizations, nothing was said about where the blobs were with respect to other objects in the scene. For example, it was not specified whether y, a pot, was on x, the stove. All that need be specified according to this mode of accessing probabilistic information is an inventory listing (Mandler & Stein, 1974) of the objects (or an object) in the setting. This access mode implies that variations in the probabilistic relations among a set of objects should produce effects on object perceptibility even when the displays do not form coherent scenes.

An alternative to the inventory listing account of the PVE is a mode by which probability effects (and all spatial relation effects) on scene perception derive from the utilization of stored knowledge about the likely relations among objects (Biederman, in press, 1987a; Biederman et al., 1982). This mode of access will be described in the discussion section.

One goal of this experiment was to determine whether the effect of violations of probability on object perception was a consequence of an inventory listing. The strategy capitalized on the use of the nonsense arrangements of objects in the present experiment where scene-like spatial relations were not present. However, probability (or consistency) could be defined on the basis of an inventory listing of the distractors. If the perceptibility of objects suffered as a consequence of their
improbability, then the credibility of inventory listings as an account of the PVE would be increased.

In the present experiment, subjects were provided with the name of a target object (e.g., tea kettle) and attempted to determine whether it was present in a briefly flashed display of objects arranged in a non-scene-like fashion around an imaginary clock face. The objects in the display would either be ones from a setting that would be likely to contain the target (consistent) or, on other trials, ones where the object would be improbable (inconsistent). As noted earlier, when the arrangement of objects does form a scene, improbable objects are less detectable than probable objects. If the same processes are at work with a display of unrelated objects, then inconsistent objects would be expected to be less detectable (i.e., have lower d’s) than consistent objects. Such an effect will be termed assimilation. However, if the perceptual effect from violating probabilistic relations requires a scene-like arrangement among the objects, then no difference would be expected in the perceptibility of probable and improbable objects.

Egeth et al.’s (1972) and Jonides and Gleitman’s (1972) discovery of “pop-out” effects in letter-digit detection raises the third possibility: Improbable objects might be more readily detected than probable objects. These investigators reported that the detectability of a digit was unaffected by increases in the number of letters in the display but was directly related to the number of nontarget digits. Phenomenally, the digit appeared to pop out from a field of letters. (If pop out is an apt description, then the Egeth et al., 1972, and Jonides and Gleitman, 1972, experiments show that subjects can monitor for the absence of a pop out in that the advantage of different category items was obtained on negative trials. Whatever the status of the pop-out effect with letters and digits, would such an effect occur with an object that was not likely to be found in the same setting as other objects in the field? Would a tractor pop out from a display with a stove, frying pan, toaster, spice rack, and salt shaker? Such a result would suggest the possibility that pop-out context effects might arise whenever the usual structure—scenes for objects, words for letters—was not present. (The absence of a pop-out effect would suggest that such effects are restricted to alphanumeric stimuli.)

Our interest in the PVE in the context of the present experiment was in obtaining some information on one of the possible routes through which the semantic information in a scene might be accessed. Consequently, the category boundaries were defined in terms of the kinds of objects that are likely to be found in a given setting. Obviously, another parallel between the different categories of letters and digits and object stimuli could be drawn. Objects could have been drawn from the same semantic category (e.g., weapons). But items from such categories (e.g., a machine gun and a crossbow) are not necessarily probable within the same scene.

Method

Subjects

The subjects were 48 State University of New York at Buffalo undergraduates who participated in the experiment as part of their introductory psychology course requirement. All were native English speakers and had normal or corrected to normal vision.

Stimuli

The displays consisted of one to six line drawings of objects generated from a set of 84 objects. Each object was centered on the perimeter of an imaginary circle, producing a clock-face arrangement of the objects as shown in Figures 1 and 2. Direct positive Kodak Panatomic X slides were made of the arrangement.

The 84 objects could be grouped into 12 settings of 7 objects each (see Appendix). The settings were battlefield, kitchen, baseball field, living room, office, city street, child’s nursery, backyard, campsite, farm, bathroom, and orchestra. The objects were single entities, with readily available basic level names (Roch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). The entities were familiar and readily named by all the subjects, with the exception of the bazooka; another did not know the name for the garden hose. Both understood the objects once they were named by the experimenter. These objects never served as targets.

The sizes of the displayed objects were kept fairly uniform so that an object’s size could not be used as a cue to its identity (e.g., the canoe was not necessarily larger than the baseball bat). The range of maximum extents of the various objects was in a 2:1 ratio, with their displayed sizes uncorrelated with their real-world sizes. Object size was measured as the length times the width of the longest prominent dimensions. The objects averaged approximately 1 1/2” in height (range of 6” to 21 1/4”) and 1 1/2” in width (range of 29” to 22”). The overall projected diameter of the displays containing six objects averaged 5 1/4”. The distance of the center of an object from the central fixation point (i.e., the radius of the imaginary circle) was 21 1/4” for all objects. As illustrated in Figures 1 and 2, the objects never touched and did not appear crowded. The mean distance between the closest pair of objects in the six-object displays (i.e., the separation of the closest adjacent objects) was 3 1/4” (range 8” to 1 1/3”).

Two preliminary reports with these stimuli (Biederman, 1981; Biederman et al., 1979) had evidenced strong physical similarity effects in that high false-alarm rates resulted when an object similar in shape to the target was in the display (e.g., a broom for the target fork). To study the effect of the number of distractors over and above the effects of similarity, only distractor objects that were relatively dissimilar to the target were included in a display for a given target. (Physical similarity was partially responsible for a high false-alarm rate for the consistent conditions in the earlier experiments.) The possible distractor objects for a given target were subjectively ranked by their physical similarity to the target, according to criteria of axes structure, elongation, and component similarity.

Display size (number of objects in the display) was varied by adding these objects in decreasing order of their similarity. Thus, when there were two objects in the display on a present trial, the distractor object would be of maximum similarity to the target. The variation from five to six objects in a display was made by adding the object least similar to the target. This meant that there was a decreasing amount of target-distractor similarity (as measured by the presence of the most similar object to the target) as display size increased. Consequently, increases in RTs or error rates with increasing display sizes could not be attributable to the increased likelihood of including a highly similar distractor with the larger display sizes. Consistent and

---

3 Under some conditions, Francolini and Egeth, 1979, have shown that this phenomenon does not occur. Cardosi (1986) has argued that the letter-digit category effect in visual search is a consequence of reduced physical similarity of letters and digits compared to within category similarity values.
OBJECT SEARCH

inconsistent targets had equivalent levels of similarity to the distractors.

Procedure

The subject's task was to determine if a given target object, specified by name, was present in the display. Each trial was initiated when the subject read the name of a target object from a display terminal. When ready, the subject pressed a key on the terminal with the nonpreferred hand. One-half second later, a central fixation point would be presented for 500 ms, followed by the presentation of the stimulus display for 100 ms. The fixation point was at the center of the clock-face display. Immediately after the offset of the display, a mask of random-appearing contours (straight and curved lines from object fragments) was presented for 500 ms. If the named object was present in the display, the subject was to press a microswitch marked yes with the index finger of the preferred hand. Otherwise, a microswitch labeled no was to be pressed with the middle finger. The subjects were encouraged to respond "as quickly and as accurately as possible."

Slides were presented by three Kodak Carousel projectors fitted with Gerbrands Electronic Tachistoscope shutters, one each for the central fixation point, object display, and mask. The projectors were controlled by a microcomputer, which also timed and stored the response and controlled feedback information.

Design

For half of the trials the target object was present; on the other half of the trials it was not. On half of each of these kinds of trials, the target label was consistent in that it named an object that would be likely to occur in a setting which contained the other objects in the display (e.g., tea kettle among kitchen objects). On the other half of the trials, the target label was inconsistent in that it named an object that would be unlikely to be found in a setting that contained the other objects in the display (e.g., file cabinet among objects that might appear in a street as shown in Figure 2). On all consistent trials and those on which an inconsistent target was absent in the display, the objects in the display were all from the same kind of setting.

Twelve objects, one from each setting, were designated as targets for all subjects. The subject was always trying to detect just one of these 12 objects on a given trial. These 12 objects were selected so as to be minimally confusable in shape with the 72 nontarget objects in the experimental displays. Each subject viewed 264 experimental slides along with 16 buffer slides. For half of the slides the target was present; for the other half it was absent. There were an equal number of consistent and inconsistent trials for display sizes of two through six. But an inconsistent target with only a single object in the display could not be defined. Consequently, there were 12 more consistent-present trials (one for each target) than inconsistent-present trials. Across subjects, the variation in the assignment of target labels to slides on which all of the objects were from the same setting (i.e., there were no inconsistent objects present in the display) served to produce three of the four conditions with the following number of slides per subject: consistent-present, N = 72; consistent-absent, N = 66, and inconsistent-absent, N = 66. Those slides that contained an inconsistent object (N = 60) always were inconsistent-present trials.

The composition of the experimental slides and the subjects' target sequences produced a 2 (present-absent) x 2 (consistent-inconsistent) x 5 (display size, from two to six objects) x 12 (targets) x 24 (subjects) design. An additional 24 experimental slides contained only a single object, 2 slides for each of the 12 targets. Half of these trials were yes trials; the other half no trials. Because the consistency factor could not be defined for these 24 slides, they were not included in the factorial analysis but are included in the description of the results.

A given target object, when present, occupied the identical position in the clock face on consistent and inconsistent trials for a given display size. For example, if a hatchet was a consistent target and it was present at the two o'clock position when there were three objects in the campground setting, then when it was an inconsistent target in the three-object kitchen setting, it was also present at the two o'clock position. However, for a different set size, it would occupy a different clock position. With a display size of two, objects were at opposite ends of a diameter of the clock face. At larger display sizes the positions were determined to minimize the clumping of objects around the clock face. Even though there were at most six objects in a display, a total of seven consistent objects were needed for each kind of setting because when there were six objects in the display and

Figure 1. Sample stimulus display (city street) with five objects. (For this display, TRAFFIC LIGHT would be the target on a target-present, consistent trial and on a target-absent, consistent trial, but on the latter the traffic light would be replaced by a street lamp. On a target-absent, inconsistent trial, this same display would be shown but the target would be FILE CABINET.)

Figure 2. Sample stimulus display with four objects on a target-present, inconsistent trial. (Target would be FILE CABINET.)
a consistent target was not present (viz., an absence trial), an additional object was needed to produce the sixth display object. In the Appendix these objects are listed as the consistent target plus six distractors. For example, with a display size of six, when the grenade was the target on an absent trial, the bazooka was included so that there would be six objects.

Each subject viewed 264 experimental slides grouped into four blocks of 66 slides each. Within each block, there were an equal number of settings, present and absent trials, and targets. Stimuli were presented in two counterbalanced orders. One order was randomly determined subject to the constraints that (a) each sequence of 6 slides contained one instance of each of the six set sizes, and (b) each sequence of 12 slides contained all 12 settings. The second order was the reverse of the first. This was done by rotating the carousels in forward order for half of the subjects and reverse order for the other half. Within each of the directions of rotation, the order of the blocks was balanced across subjects by Latin square. These features of the design produced mean serial positions of each slide and each slide-condition combination that were identical and of equal frequency of occurrence in all quarters of the slide sequences. Following each response, RT and error feedback was displayed on the subject's terminal. At the end of each block of trials, the mean error rate for that block was displayed.

In addition to the 264 experimental slides, 2 slides (of objects from a circus setting) were used at the beginning and end of each block as buffers for warm up or end-of-block-anticipation effects. The data from these slides were omitted from the analyses.

To familiarize the subjects with the individual objects, each object was shown for 3 s prior to the practice trials. The subject named these objects as they were shown. The only difficulties encountered, as mentioned earlier, were with the nontarget objects bazooka and garden hose. All subjects understood what these objects were after the experimenter named them. Each subject was given instructions and 22 practice trials with the objects from the circus scene (the same objects used as buffers) prior to the onset of the experiment. The practice slides were initially presented for 1,600 ms and reduced by halves every fourth slide to the experimental duration of 100 ms.

Results

The overall error rate was 21.28%, and the mean correct RT was 580 ms. Although performance improved over the four trial blocks as shown in Figures 3 and 4, the effects of the major experimental variables of consistency and response were generally independent of practice and consequently the data were collapsed over trial blocks.

$F'$ statistics were constructed (because targets was a random variable) by combining subjects, from 48 to 24 with the error analysis, and reducing denominator $df$s to produce balanced subsets of conditions. With the analysis of correct RTs, the existence of errors in responding to some objects meant that these objects did not have an RT value for a given subject. The 48 subjects had to be reduced to 4 balanced pseudosubjects before a balanced set of data could be established without missing cells. In neither case did this conservative procedure

![Figure 3](image-url)  
*Figure 3. Error rates as a function of trial block, consistency (consistent-inconsistent), and response (presence-absence).*
(because it ignores between-subjects variation within a pseudosubject) noticeably affect the description of the results.

The presence of distractors in the display resulted in marked declines in detectability. Figure 5 shows the effects of consistency, display size, and response (presence-absence) for error rates; Figure 6 shows these effects for mean correct RTs. As the number of objects in the display increased, error rates and RTs increased, \( F'(4, 40) = 16.06, p < .001 \), for errors, and \( F'(4, 23) = 9.54, p < .001 \) for RTs. The increase in RTs and false alarm rates were linear. (As noted in the Method section, the data from trials where there was only one object in the display were omitted from the analysis of variance because the consistency factor was not defined at this level.)

Overall error rates were lower, though nonsignificantly, for inconsistent compared with consistent trials (23.37\% vs. 19.21\%, respectively), \( F'(1, 11) = 3.15, .05 < p < .10 \). The effect of consistency was significant for RTs, however, with the inconsistent trials 33 ms shorter, overall, than the consistent trials. \( F'(1, 10) = 6.29, p < .05 \). For both measures, all the advantage of the inconsistent trials derived from absent trials, producing a strong Response \( \times \) Consistency interaction. Target objects that were inconsistent with the other objects in the display had much lower false alarm rates (on absent trials) but slightly higher miss rates (on present trials) than consistent targets, \( F'(1, 11) = 18.49, p < .01 \). With RTs, inconsistent targets had shorter RTs on absent trials but were equivalent to consistent RTs on present trials, \( F'(1, 8) = 15.17, p < .01 \).

Curiously, this interaction was not noticeably magnified with increases in display size. Consistency effects might have been expected to increase in magnitude with increasing display sizes in that a larger number of objects from a given setting would define more strongly the nature of that setting and hence consistency-inconsistency with it. But although the data were in that direction (e.g., false alarm rates on consistent trials showed a precipitous increase compared with the inconsistent absent trials), the Consistency \( \times \) Response \( \times \) Display Size interaction fell short of significance for errors, \( F'(4, 36) = 1.95 \), and was totally nonexistent for RTs, \( F'(4, 22) < 1.00 \).

Figure 7 shows \( d' \)'s as a function of display sizes separately for consistent and inconsistent trials. Inconsistent trials produced higher \( d' \)'s than consistent trials throughout the range of display size studied. The \( d' \)'s decreased linearly as the number of objects in the display increased, with the magni-
tude of the advantage for the inconsistent trials unaffected by display size.

Table 1 shows beta values as a function of display size and consistency. Consistent trials had betas that were lower than inconsistent trials. Put simply, subjects were prone to respond “yes” if the target was consistent with the display and “no” if it was inconsistent. For both kinds of trials, mean beta declined with increasing set size. Although this was expected for the consistent trials, it is somewhat surprising that there was such a decline on the inconsistent trials. A strategy that could give rise to such an effect might go something like this: “Even though the target is unlikely in this kind of setting, because there are more objects in the display, the target is more likely to be one of them.”

Probability of detection calculations from a high threshold, correction-for-guessing model showed a linear decline in the probability of target detection with increasing display sizes. The slope of this function was .034 per item for both consistent and inconsistent conditions. The high threshold model requires the unlikely assumption that no information is accrued about the target on trials on which the high threshold is not exceeded. This has never been found to be the case in experiments like the present one (R. D. Sorkin, personal communication, 1986).

Discussion

Effects of Display Size

With these nonscene displays, the likelihood of detecting a target object declined monotonically—almost linearly—with the number of other objects in the display. The slope of this function was independent of the consistency of the target with the items in the display. The monotonic decline over set size suggests that object recognition is a capacity limited process (Shiffrin & Schneider, 1977; Treisman & Gelade, 1980). That aspect of the design that increased size by adding objects in order of declining similarity allows the conclusion that the effect is one of the number of objects rather than one of similarity, inadvertently confounded with set size.

Given the apparent inability to suppress an interpretation of an object that is viewed, the evidence for capacity limitations in handling multiple objects supports the conjecture that performance can sometimes be mandatory but not necessarily capacity free.

Effects of Consistency

Error rates, d’s, and RTs were in a direction favoring an advantage for inconsistent over consistent trials. (Response bias effects will be discussed later.) It would be somewhat inappropriate to term this result a pop-out effect in that all the advantage for inconsistent targets was confined to the target-absent trials. We have found this result—a modest overall advantage for inconsistent targets completely attributable to target absent trials—in four prior experiments (Biederman, 1981, 1982; Biederman, Teitelbaum, Klatsky, et al., 1979) with differing stimulus sets and instructional conditions in these nonscene displays.

Even the weaker conclusion, that there was no advantage for consistent trials (rather than an advantage of the inconsistent trials), differs from what has been repeatedly found in scenes. In real-world scenes, there is an advantage (higher d’s) for consistent objects (i.e., those in a base or normal position; Biederman et al., 1982, 1983). The present results are thus evidence against an explanation of the PVE that assumes that the interference arises from an inventory list of the objects in a scene.

But why would performance for inconsistent targets be better than consistent targets, and then only on target-absent trials? The explanation we favor requires a detailed examination of what might be required to determine that a given display object was not the target. Quite likely this decision was performed at a conceptual level—following activation of its representation in memory—rather than at the level of shape primitives. It might have even been performed on a name level. The targets were specified by name and the positions and identities of possible distractors were large and uncertain. Even with prefamiliarization it was unlikely that a diagnostic piece of contour could be readily employed to classify the target even if one could be found. Thus, if one

---

4 Some caution must be exercised with this conclusion in that absent trials in the scene perception experiments (e.g., Biederman, Mezzanotte, & Rabinowitz, 1982) differed in what might have been two critical features from those in the present investigation: (a) In the current experiment, as soon as an inconsistent object could be detected in the display, a “yes” response that would necessarily be corrected could be emitted without checking against the target label, because inconsistent objects never appeared in a display on “no” trials (although a subject was not told of this in the instructions). This could have had the effect of facilitating performance in the inconsistent trials. Cued objects in Biederman et al.’s (1982) experiment were either high or low in probability on both presence and absence trials. (b) To produce a calculation of d’ in the probability violation (scene condition) that is comparable to the inconsistent condition in the present experiment, the false alarm rates from the base condition with low probability distractor targets should be used because, as noted in the foregoing point, inconsistent absent trials never had an inconsistent object in the display. When calculated in this way, the d’ of 1.69 for the probability violation condition is higher (slightly) than the d’ of 1.48 for the base condition (where the distractors were of high probability). This difference is solely a function of comparing d’s from low probability distractors (with a lower false alarm rate) in the probability violation condition against high probability distractors (with a higher false alarm rate) in the base condition. In Biederman et al.’s (1982) experiment, as in the present experiment, false alarm rates were lower for low probability targets than they were for high probability targets. Even with these calculations for d’s in the scene experiments, the relative advantage for the inconsistent condition was marginally greater in the present experiment. At a set size of six, d’s for consistent and inconsistent conditions were 1.12 and 1.42, respectively, a difference of .30 as opposed to a difference of .21 for the scene calculations (1.69 and 1.48, for probability and base d’s). It should also be noted that for the violations of relations other than probability—namely, support, position, and size—where both base and violations had high probability targets, d’s were clearly lower in the violation conditions.
were looking for a traffic light, there would be at least one object that contained straight lines and was compact and vertically elongated (e.g., a file cabinet), so that it could not be rejected on the basis of such simple shape descriptors. It is possible that when an object does activate a representation in memory, its scene membership is also activated. For example, the perception of a traffic light might activate the knowledge that it is a street object. Such knowledge can also be activated from the target name. When one is attempting to detect the presence of a traffic light among distractors that are also street objects, the distractors will share the same scene-membership association, so they cannot be rejected on that basis. However, an object that does not share that association, say an office object, can be readily rejected from these associations on that basis. To the extent that the rejection of an object as a nontarget is performed at a conceptual level on the basis of scene-membership associations, the inconsistent targets will show an advantage in the target absent conditions, as we have found. Essentially, the foregoing explanation is similar to the accounts of fast “no” responses on Sternberg memory scanning tasks to nontargets that are numerically remote from the target set (e.g., Morin, DeRosa, & Stultz, 1967). When a positive set is defined by the digits 1, 2, and 3, RTs for responding “no” to a probe of 4 are longer than to a probe of 8.

From this perspective, our finding that consistency effects were present to their maximum extent with the presence of only a single distractor was a likely consequence of that distractor’s high visual similarity to the target. This would have required the subject to identify the distractor and consequently afford an opportunity to make a fast rejection based on a scene membership association. Such fast rejections of inconsistent objects have been observed during object search in scenes. If inconsistent (or improbable) targets were never present, a correct “no” response for an inconsistent target, for example, FIRE HYDRANT in a kitchen scene, was made with
much shorter latencies and higher accuracies than a correct rejection of toaster in that same scene (Biederman, Glass, & Stacy, 1973). 3

**Response Bias Effects**

In addition to the effects of consistency on detectability, there was a pronounced effect of consistency on response bias. The marked interaction between consistency and response, producing the large effect of consistency on beta, also indicates that sufficient information could be extracted from the 100-ms display presentations to determine the probabilistic relations and that these relations affected responses. This interaction became evident with as few as two objects in the display. Even though the term guessing has often been used to describe these response effects, these responses were not produced at the end of a long and deliberative process: Mean correct RTs were only 580 ms.

The presence of a response bias effect as a function of target-setting consistency in these nonscene displays has a parallel in the processing of scenes. Biederman et al. (1982) reported that when the target was absent for a scene, subjects were more likely to false alarm if the target was consistent than if it was inconsistent. Similar effects had been observed by Biederman et al. (1973) and Palmer (1975). For example, if the target was toaster and a frying pan was cued in a kitchen scene, then subjects were more likely to respond with a false alarm (by responding “yes” [target present]) than if fire hydrant was the target. However, this response bias effect was insufficient to account for the increased miss rate when an inconsistent object was the cued target in the Biederman et al. (1982, 1983), Klatsky (1983), and Malcus (1983) scene perception experiments. In those experiments, inco-

---

3 As noted in Footnote 4, the detection of cued objects in real-world scenes also reveals faster rejection of improbable targets over probable targets in that lower false alarm rates have been found for target names that are improbable in a scene compared with target names that are of high probability (Biederman, Mezzanotte, & Rabinowitz, 1982).
consistent targets had lower \( d' \)'s as well as higher criteria (\( \beta \)); in the present experiment, inconsistent targets had higher criteria and higher \( d' \)'s.

**Implications for the Probability Violation Effect**

Although an inventory list account is insufficient for the PVE, the foregoing explanation of the consistency effect actually salvages one assumption of the inventory list hypothesis: that scene membership associations are elicited from objects in nonscene displays sufficiently early to affect performance (but not object perception).

That an inventory list is insufficient to account for perceptual effects is supported by three results from prior experiments. The first is the equivalent magnitude of the interference effects on object perception from violations of the Position and Probability relations. An object undergoing a Position violation (e.g., a hydrant on a mailbox in a street scene) is probable in the scene. The incongruity derives from the object's inappropriate spatial relation(s) to other objects. Violations of Probability can be determined merely from the identification of the objects—their spatial relations need not be accessed. But violations of Position do require that relations be determined—no other information is available for the incongruity. If objects are identified prior to their relations, then violations of Probability should be more disruptive than violations of Position. But, as previously noted, this has not been found to be the case. Position violations are at least as

---

**Table 1**

Values of \( d' \) and \( \beta \) as a Function of Consistency and Display Size

<table>
<thead>
<tr>
<th>Display size</th>
<th>Consistent</th>
<th>Inconsistent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( d' )</td>
<td>( \beta )</td>
</tr>
<tr>
<td>1</td>
<td>2.46</td>
<td>88</td>
</tr>
<tr>
<td>2</td>
<td>2.18</td>
<td>78</td>
</tr>
<tr>
<td>3</td>
<td>1.66</td>
<td>78</td>
</tr>
<tr>
<td>4</td>
<td>1.46</td>
<td>72</td>
</tr>
<tr>
<td>5</td>
<td>1.34</td>
<td>69</td>
</tr>
<tr>
<td>6</td>
<td>1.12</td>
<td>58</td>
</tr>
<tr>
<td>( M ) (2–5)</td>
<td>1.55</td>
<td>69</td>
</tr>
</tbody>
</table>

* The consistency variable was not defined for displays of one item.
disruptive to object identification as Probability violations (Biederman et al., 1982, 1983; Klatsky, 1983; Mezzanotte, 1981). This result suggests that the disruptive effect of Probability violations arises from the specific relations among the objects in the scene, not their identities.

A second piece of evidence against the object-then-relation hypothesis derives from Klatsky's (1983) direct test of the role of identification in producing the violation effect. Klatsky prepared scenes in which a large, prominent object was degraded through contour deletion. The degraded object was either one that would be highly diagnostic to that scene's identity (e.g., a stove in a kitchen) or nondiagnostic (e.g., a person in the kitchen). If the quick access to a semantic representation of a scene (as assessed by the presence of violation effects) was dependent upon the identification of diagnostic objects for that scene, then the degradation of a prominent diagnostic object should have reduced or eliminated the violation effects more than such degradation applied to a nondiagnostic object. This did not happen. Klatsky found that degrading prominent diagnostic objects did not affect the magnitude of the violation effect for either the probability or the position relations. (The degradation of a diagnostic object interfered with the classification of the scene itself [e.g., a picture of a kitchen was more difficult to verify as a kitchen], more than did the degradation of a nondiagnostic object. This finding confirmed his classification of what was and was not diagnostic and the potency of the manipulation.)

The third and strongest piece of evidence against the inventory list account of the PVE derives from Mezzanotte's demonstrations whereby an identifiable scene emerges from objects that have been degraded to the point where, presented individually, they are unidentifiable (described in Biederman, 1981). Mezzanotte (1981) showed that violation effects on nondegraded targets from such scenes were comparable with the effects on nondegraded scenes.

If an inventory list account is insufficient to account for the PVE, let alone the effects of the other scene relations, what might account for the real-time activation of a semantic representation of a scene? Biederman (in press, 1987a) has proposed rapid activation of a scene's representation might be mediated by geon clusters. A geon cluster is an arrangement of the convex volumetric primitives (geons) from different objects that preserve the relative size and aspect ratio and relations of the largest visible geon(s) of each object. For example, a vertical slab when occluded by a larger brick often resembles a table occluding a chair with only the chair back in view. In such cases, the individual geon (e.g., the chair back) may be insufficient to allow identification of the object. However, just as an arrangement of two or three geons almost always allows identification of an object, an arrangement of two or more geons from different objects may produce a recognizable combination. The cluster acts very much as a large object.

If this account is true, fast scene perception should only be possible in scenes in which such familiar object clusters are present. These clusters may be the basis by which the visual system, when confronted with real-world scenes, overcomes the costly effects of display size so apparent in the present experiment.

References


### Appendix

#### Targets and Distractors for the Twelve Settings

<table>
<thead>
<tr>
<th>Condition</th>
<th>Battlefield</th>
<th>Bathroom</th>
<th>Kitchen</th>
<th>Campgrounds</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TARGETS</strong></td>
<td>GRENADE</td>
<td>TOILET</td>
<td>TEA KETTLE</td>
<td>HATCHET</td>
</tr>
<tr>
<td>CONSISTENT</td>
<td>BASEBALL</td>
<td>BARBECUE GRILL</td>
<td>TRACTOR</td>
<td>BABY RATTLE</td>
</tr>
<tr>
<td>INCONSISTENT</td>
<td>TANK</td>
<td>BATHTUB</td>
<td>TOASTER</td>
<td>BACKPACK</td>
</tr>
<tr>
<td>DISTRACTORS</td>
<td>CANNON</td>
<td>HAIRDRYER</td>
<td>FRYING PAN</td>
<td>SLEEPING BAG</td>
</tr>
<tr>
<td>FIRST</td>
<td>MILITARY HELICOPTER</td>
<td>SHAVING CREAM CAN</td>
<td>SPICE RACK</td>
<td>TENT</td>
</tr>
<tr>
<td>SECOND</td>
<td>MILITARY HELICOPTER</td>
<td>TOOTHPASTE</td>
<td>SALT SHAKER</td>
<td>CANOE</td>
</tr>
<tr>
<td>THIRD</td>
<td>MILITARY HELICOPTER</td>
<td>RAZOR</td>
<td>FORK</td>
<td>COOLER</td>
</tr>
<tr>
<td>FOURTH</td>
<td>MILITARY HELICOPTER</td>
<td>TOOTHBRUSH</td>
<td></td>
<td>DEER</td>
</tr>
<tr>
<td>FIFTH</td>
<td>MILITARY HELICOPTER</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SIXTH</td>
<td>MILITARY HELICOPTER</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| **OFFICE**      | FILE CABINET                                     | TRACTOR                       | BABY RATTLE                  | HARP               |
| CONSISTENT      | TRAFFIC LIGHT                                    |                              | TEA KETTLE                   | GRENade            |
| INCONSISTENT    | IN/OUT TRAYS                                     |                              |                              |                    |
| DISTRACTORS     | DESK                                             | EAR OF CORN                   | BABY BOTTLE                  |                    |
| FIRST           | PENCIL SHARPENER                                 | Barn                          | WALKER                       |                    |
| SECOND          | PENCIL SHARPENER                                 | PSG                           | BASSINET                     |                    |
| THIRD           | PENCIL SHARPENER                                 | ROOSTER                       | MOBILE                       |                    |
| FOURTH          | PENCIL SHARPENER                                 | FARMER                        | BABY                         |                    |
| FIFTH           | PENCIL SHARPENER                                 | PITCHFORK                     | CRIB                         |                    |
| SIXTH           | PENCIL SHARPENER                                 |                              |                              |                    |

| **BACKYARD**    | BARBECUE GRILL                                   | LAMP                          | BASEBALL                      | STREET LIGHT       |
| CONSISTENT      | TOILET                                           | HATCHET                       | TRAFFIC LIGHT                | STOP SIGN          |
| INCONSISTENT    | PATIO TABLE                                      | CANDLE STICKS                 | BASEBALL CAP                 |                    |
| DISTRACTORS     | LAWNMOWER                                        | EASY CHAIR                    | BALL GLOVE                   |                    |
| FIRST           | LOUNGE CHAIR                                     | GRANDFATHER CLOCK            | PITCHER                      |                    |
| SECOND          | WATERING CAN                                     | SOFA                          | BATTER                       |                    |
| THIRD           | WATERING CAN                                     | TELEVISION                    | BAT                           |                    |
| FOURTH          | WATERING CAN                                     | FIREPLACE                     | FIELDER                      |                    |
| FIFTH           | GARDEN HOSE                                      |                              |                              |                    |
| SIXTH           | BIRDHOUSE                                        |                              |                              |                    |

Received March 17, 1986
Revision received May 26, 1987
Accepted June 22, 1987