HUMAN FACTORS, 1981, 23(2), 153—164

Detecting the Unexpected in Photointerpretation

IRVING BIEDERMAN[1], ROBERT J. MEZZANOTTE, JAN C. RABINOWITZ[2], CARL M. FRANCOLINI, *and* DANA PLUDE, *State University of New York at Buffalo*

A methodology for measuring photointerpretation performance from a single glance at a novel scene is presented. Subjects attempted to detect the presence or absence of a target object, specified in advance by the object's name, from a cued position in a 150-ms flash of a line drawing of a real-world scene. The cue, a dot, was presented immediately after the presentation of the scene, so that the subjects were uncertain as to the position of the cued object. Miss rates as a function of distance of the cued position from central fixation, target size, and degree of target camouflage were determined. The experiment also explored how these functions varied with objects that were in unexpected locations, such as a sofa floating in a street scene. Objects in these locations violated the usual constraints which characterize the organization of real-world scenes. Large, uncamouflaged targets in a normal relation to their context suffered only modestly from the effects of increasing distance from fixation. But targets that were small, camouflaged, or undergoing violations of the relational constraints suffered marked increases in miss rates when presented only 3 or 4 deg from central fixation. Since humans can assimilate visual information faster than the eye can move, the maximum rate at which a fixation can be made provides a limit—the saccadic barrier*—of visual information processing. In addition to furnishing general guidelines for scene processing displays, the results of this experiment serve to restrict the breaking of the barrier—with the aid of high-speed, high-capacity display systems—to the extraction of the setting or gist of a scene. Individual targets, unless large, uncamouflaged, and in expected locations, require direct fixation.*

INTRODUCTION

Consider a single fixation made by a photointerpreter at a novel scene. Why, on some fixations, does the interpreter miss the target? The purpose of the present investigation was to explore the role of four variables with respect to their potency in affecting miss rates. These variables were (a) the distance of the target from fixation, (b) the size of the target, (c) the degree of target camouflage (i.e., the degree to which adjacent contours obscured the target's contours), and (d) the spatial and semantic appropriateness of the target to its setting. This last variable will be termed the target's *relational congruity*. Research in classical psychophysics has firmly established the nature of the effects of the first three variables. Thus, it has been established that the further the distance of the target from central fixation, the smaller its size, and the greater the degree of its camouflage, the less accurate its detection.

Although these functions have been studied in the psychophysical laboratory, parametric

[1] Requests for reprints should be sent to Dr. Irving Biederman, Department of Psychology, State University of New York at Buffalo, 4230 Ridge Lea Rd., Amherst, NY 14226.

[2] Now at the University of Toronto.

data on the role of these variables on the detection of single objects in a glance at a scene is lacking. Specifically, how to confidently extrapolate psychophysical functions derived from simple stimulus situations to the processing of an object in the context of a real-world scene is not known. For example, psychophysical research has shown that Vernier acuity is halved at only a 1-deg displacement from central fixation (Alpern, 1971). Given this information, one could not confidently predict that target objects which were 1 deg from central fixation would be detected only half as frequently as targets which were centrally fixated. One purpose of the present investigation was to provide the beginnings of a bridge between psychophysics and photointerpretation by presenting parametric functions for the effects of distance from fixation, size, and camouflage on object detection in real-world scenes.

A second reason for studying these three traditional psychophysical variables in a photointerpretation task was to discover how these variables interact with the target's relational congruity. Biederman (1977; 1981) has identified five classes of relations between an object and its setting which can serve to characterize the difference between a display of unrelated objects and a real-world scene. A list of these five relations and examples of their violations follows:

(1) *Support:* e.g., a floating chair. The object does not appear to be resting on a surface.
(2) *Interposition:* e.g., the background appearing through the back and seat of the chair. The objects undergoing this violation appear to pass through another object.
(3) *Probability:* e.g., the chair in a forest. The object is unlikely to appear in the scene.
(4) *Position:* e.g., the chair on top of a file cabinet in an office scene. The object is likely to occur in that scene, but it is unlikely to be in that particular position.
(5) *Familiar Size:* e.g., the chair appears to be larger than a four-drawer file cabinet or smaller than a telephone. The object appears to be too large or too small relative to the other objects in the scene.

Violating one (or several) of these relations produces an incongruity between the object and the scene. Consider, for example, the detection of a "standard" office chair. The chair could be in its normal position, resting on the floor by a desk in an office scene. Or it could violate one or several of the relations listed above. Thus, the background might appear to pass through its solid surfaces, such as its back or seat, to violate Interposition (or occlusion). If it were floating in air, it would violate Support. If it were on top of a file cabinet in the office, it would violate Position. If it appeared to be smaller than a telephone or larger than a four-drawer file cabinet, it would violate Size. If it were in the middle of a forest, it would violate Probability.

The primary interest in this article is in how the presence of any violation—the "unexpected"—affects detection performance, rather than in the detailed theoretical implications of the comparison of the various kinds of violations. In particular, the research was directed toward discovering (a) how the psychophysical variables of a target's distance from fixation, size, and camouflage interact with these violations and (b) the implications for such interactions for understanding what can be seen from a single glance at a scene. (See Biederman, 1981, and Biederman, Mezzanotte, and Rabinowitz, in press, for a theoretical discussion and detailed comparison of the different kinds of violations.)

The Saccadic Barrier

The study of performance from only a single glance was selected for both theoretical and applied reasons. With respect to theory, the information assimilated from a single fixation would appear to be the fundamental unit by which to model visual scanning performance of real-world scenes. But most of

the research on scene perception has not been concerned with what was perceived during a single fixation. Instead, primary concern has been devoted to the determinants of the order of fixations or how information was integrated from one fixation to the next (e.g., Yarbus, 1967). This research always begged the question as to what was acquired during any one fixation. The importance of studying the information assimilated from a single fixation can be appreciated once it is realized that a single fixation is all that is needed to comprehend most scenes (Biederman, 1972, 1981; Biederman, Rabinowitz, Glass, and Stacy, 1974). It is known that the second fixation at a scene will bring the eye to a region that is interesting or important (Loftus and Mackworth, 1978; Yarbus, 1967) but how, on the very first fixation, did the eye know what was interesting and important? It is this question that motivates the single-fixation research described in this article.

On a more practical level, the advent of sophisticated subject-controlled systems for rapidly displaying visual information has brought us to the point where the *saccadic barrier* may be breached. This barrier (or limit) on visual acquisition is a result of limitations in the rate—about 3/s—at which a human can make voluntary eye fixations. The minimum dwell time for these fixations is approximately 300 ms. But a number of recent experiments have suggested that visual information processing may often be completed within the first 100 to 150 ms (Sperling, Budiansky, Spivak, and Johnson, 1971; Biederman, et al., 1974; and Di Lollo, 1977). That is, the brain may have the capacity for processing visual information two to three times faster than the eye can feed it! One way to increase the rate of assimilation of visual information is to present information at extremely rapid rates (cf. Sperling et al., 1971). Under such conditions, the eye naturally remains immobile (other than the high-frequency nystagmus tremors). Thus, a system for presenting information frame-by-frame at rates higher than the eye can provide may be both practical and useful. (Obviously, such systems would need a capacity for reducing presentation rates so that the interpreter could fixate upon a region where greater acuity was required.) How can one start to study such performance? How should the human factors guidelines for such a system—or any system where a human is required to perceive scenes—be conceptualized? These were the practical questions that the present experiment addressed.

METHOD

Scenes

Two hundred and forty-seven slides of 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 from one to five slides where it underwent a violation. The background and overlays were then xeroxed together to produce a scene with the object or objects in it, and a slide was made of the xerox. The slides were produced by direct positive development of Kodak Panatomic X film.

When an object was not in its normal (or Base) condition, 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, Figure 2 is an example of an Interposition violation, and Figure 3 is an example of a triple violation of Size, Probability, and Support.

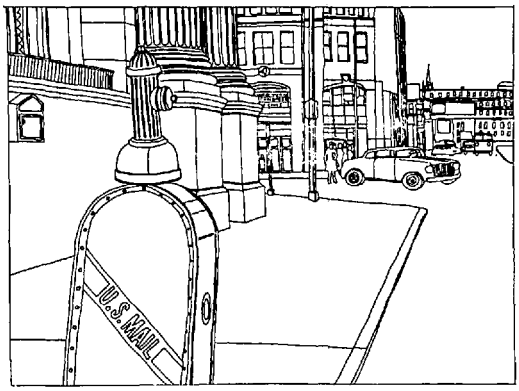

Figure 1. *An example of a Position violation for the fire hydrant. The camouflage rating for the fire hydrant was 5.5. The van on the right would be an innocent bystander. Its camouflage rating was 9.0.*

The cued object in each scene was either in a base condition (no violations) or was undergoing 1 of 10 violation conditions. In five of these violation conditions, the target violated only a single relation; in four conditions two relations were violated; and in one condition three relations were violated. Table 1 shows the various conditions and their specifications for distance from fixation, size, and camouflage of their cued objects. An attempt was made to equate the mean values for these parameters across the various conditions.

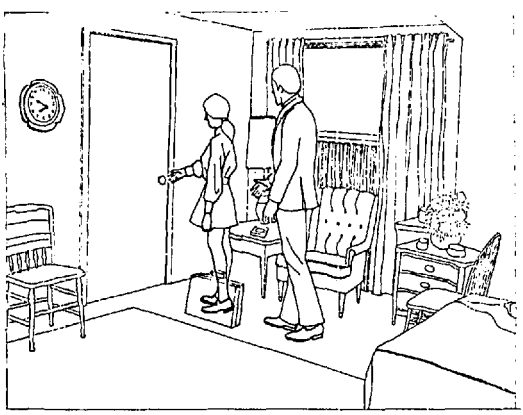

Figure 2. *An example of an Interposition violation for the attache case.*

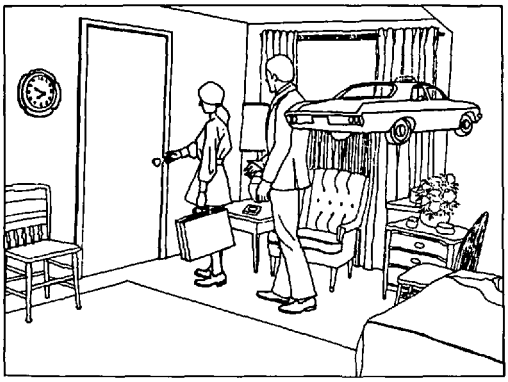

Figure 3. *An example of a triple violation. The taxi is violating the Probability, Support, and Size relations.*

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 deg in height and 1.6 deg in width. (One screen inch [2.54 cm] subtended a visual angle of 0.5 deg.) The scene backgrounds were 14 deg in width and 11 deg in height. The overall mean distance of a cued object from central fixation was 3.34 deg (S.D. = 1.40 deg).

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

The 10 violation conditions and 1 base condition (42 scenes, one for each object) 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 1, was the interposition condition, which had a higher average degree of camouflage. 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 caption to Figure 1 shows some representative values. The raters were encouraged to use the complete scale, and they

TABLE 1

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

Condition	Number of Slides	Distance from Fixation (Degrees)	Size (Length × Width) Degrees²	Camouflage Rating
Zero Violations				
Base	42	3.26	4.60	3.9
One Violation				
Position	22	3.90	4.32	4.3
Support	27	4.09	4.18	4.3
Size	21	4.57	4.73	4.0
Probability	14	4.02	5.31	2.9
Interposition	23	3.23	5.44	7.5
Mean 1 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 2 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

did. The mean (and standard deviation) for Rater 1 was 4.53 (2.25); for Rater 2 it was 4.01 (2.18). Therefore, the mean camouflage rating was 4.27. The interrater correlation was 0.793 ($df = 245$, $p < 0.001$). (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 raters averaged 0.70, and their test-retest correlations with a second rating 2 weeks later averaged 0.81. Both *r*s were significant at the 0.001 level for the 287 scenes used in that study.) 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, −0.146, though significant ($p < 0.05$, with $df = 245$).

Procedure

The sequence of events of a single trial is illustrated in Figure 4. The subject read the name of the target object from a card in a deck of target cards and, when ready, initiated the trial by pressing a switch. A fixation point was then presented on a screen for 500 ms, immediately followed by a 150-ms flash of a slide of the scene. The 150-ms 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, in turn, immediately followed by a cue (a dot) embedded in a mask of random

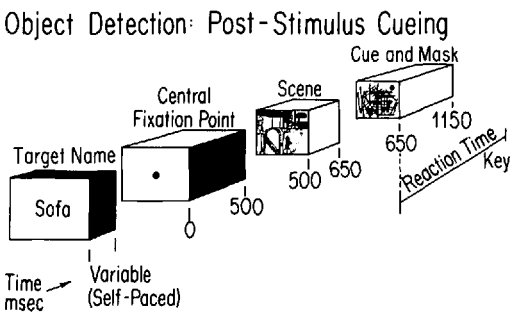

Figure 4. *Sequence of events on a trial in the Object Detection task.*

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 located 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) position 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 target object would not be present in the scene. The cue would point to a position 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. An attempt was made to distribute the violation conditions, objects, and background scenes homogeneously across the 12 blocks so that each violation condition background scene and half the objects would appear at least once in each block.

For each violation slide there were two possible cues: one designating an object in a normal relation to its setting, 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 (for a correct "NO" response). By including a condition in which an object was in a normal relation to its background in a scene where another object was undergoing a violation, the effects of the presence of violations on an "innocent bystander"—an object not undergoing a violation—was 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. Four decks of target-object cards were made to produce the various conditions. The decks were balanced across subjects so that 24 subjects used each deck.

The sequence of blocks was balanced across subjects by two Latin squares. Half the subjects took the blocks according to one Latin square; the other half of the subjects according to the second Latin square. Within each Latin square, one-fourth of the subjects (i.e., 12 subjects) had each of the four decks of targets. Half of the subjects within each of these subgroups 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 (127.5). 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 would press a switch (with the nonpreferred hand) to initiate the trial. Subjects were fully instructed as to the nature of the scenes and violations.

Slides were presented by four projectors fitted with electronic tachistoscope shutters. One projector was used for a central fixation point, one for the scene, one for the cue (a dot), and one for the mask. The voicekey was a microphone connected to an audio threshold detector. The signal from the detector stopped a timer from which response times (RTs) were recorded.

RESULTS

None of the balancing variables (viz., target-label deck and Latin Square) yielded significant effects. There was an overall decrease in error rates with practice over the 12 blocks, but this effect was relatively constant across the major experimental conditions described below. Consequently, the data from the different blocks were combined to produce mean values for the major variables of interest. (See Teitelbaum and Biederman, 1979, for a full discussion of learning effects in this kind of task.) 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% to 19.2%, respectively. Mean correct RTs were 999 ms.

Figure 5 shows the miss and false alarm rates as a function of the number of violations for both violations and innocent bystander (and base) cued objects.

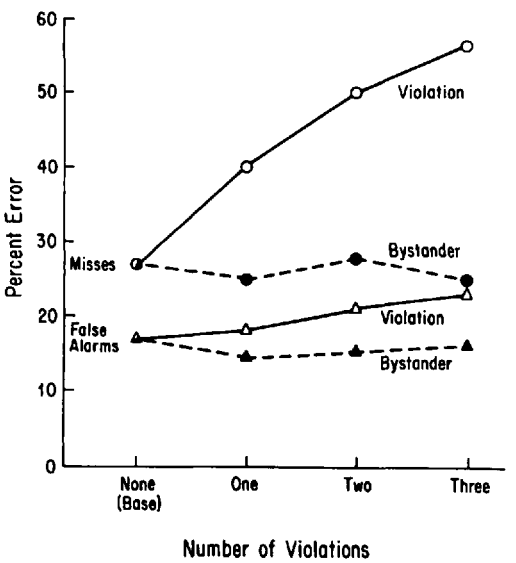

Figure 5. *The effects of number of violations on miss and false alarm rates for both violation and bystander cued objects.*

Since the individual Violation conditions were not of primary concern to this investigation, the 11 conditions were combined into four levels of a number of violations variable by combining the data from the five single Violation conditions into one level (one violation) and the four double Violation conditions into another level (two violations). The Base condition constituted the zero violation level, and the triple Violation condition the three violations level. The term *violation cost* will be used to refer to the increase in error rate or RT in the detection of a target undergoing a violation compared to its error rate or RT when that target was in the Base condition. A target which violated a relation was more likely to be missed than when it was in a Base position. The miss rates increased from 24.9% with no violations (Base condition) to 40%, 51%, and 58% for one, two, and three violations, respectively, $F(3,276) = 72.71$, $p < 0.001$. Thus, violation costs were incurred in the miss rates, and these costs increased with the number of violations.

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% overall, when the cued object was undergoing a violation compared to when it was in a Base position. As with the *miss* rates, Figure 5 shows that the false alarm rates also increased consistently with an increase from zero to four violations, $F(3,276) = 5.64$, $p < 0.002$. Thus, there was a consistent decline in d' from zero to four violations, 1.62, 1.14, 0.78, and 0.54, respectively.

Figure 5 also shows that the detection of innocent bystanders was completely unaffected by the presence of a violation.

Because the error rates were so high, a number of the slides had too few correct RTs for inclusion in the RT analysis. As a criterion

for inclusion, a slide was required to have at least six correct RTs. By this criterion, 36 slides—all violations—were excluded from the analysis because so many errors were made on them that fewer than six correct RTs were recorded. When this was done, the mean correct detections of objects undergoing violations averaged 31 ms longer than the detection of those objects in the base condition. Thus, a speed-for-accuracy tradeoff would only have increased the violation costs for errors.

Violation Costs and the Effects of Physical Parameters

Figure 6 shows the miss rates along with the effects of distance from fixation and target size with the effects of camouflage removed by a regression analysis. Not surprisingly, the further an object from fixation or the smaller its size or the greater its camouflage (not shown in Figure 6), the more likely it would be missed. The Pearson r (df = 245) between miss rates and distance was 0.398 ($p < 0.001$); between miss rates and size

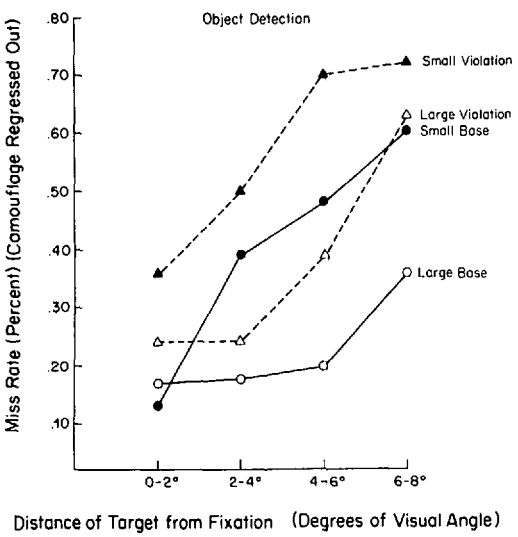

Figure 6. *The effects of distance of cued object from fixation, size of cued object, and Violation condition on miss rates with camouflage regressed out.*

(maximum length × maximum width) was −0.407 ($p < 0.001$); and between miss rates and camouflage rating was 0.151 ($p < 0.001$). The multiple R was 0.605 ($p < 0.001$) between these three variables taken together and miss rates. For this analysis, the sizes (length × width of most prominent dimensions) averaged 8.15 deg² for the large bases, 7.86 deg² for the large violations, 1.03 deg² for the small bases, and 1.41 deg² for the small violations. A symmetrical, large object would average about 3 deg, and a small object would average about 1.5 deg in extent.

It is evident that violations interfered with the detection of large as well as small targets. Most importantly, the violations adversely affected target detection even when the targets were within foveal vision (as well as several degrees removed from foveal vision). A recent experiment (Klatsky, Teitelbaum, Mezzanotte, and Biederman, 1980) 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.

Figure 7 shows the effects of camouflage and visual angle on miss rates, and Table 2 presents the statistical analyses of these functions. For this analysis, the cued objects were partitioned into high versus low camouflage subsets. The low camouflage targets averaged 2.55 on the camouflage rating scale. The high camouflage targets averaged 5.67 on the camouflage rating scale. The effects of camouflage were relatively modest within the area of central fixation. However, once targets were located beyond 2.75 deg, a considerably greater effect of camouflage was found. With targets which were 5.75 deg removed from central fixation, miss rates climbed precipitously from 39 to 64%, an increase of 25%.

Closer inspection of Figure 7 reveals a dramatic interaction in the effects of distance of a cued object from central fixation; degree of

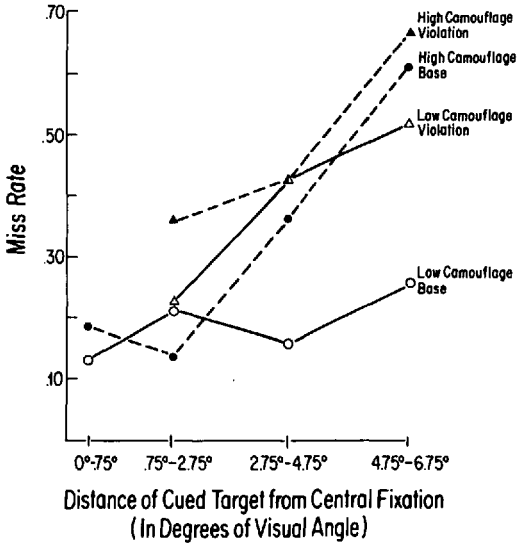

Figure 7. *Miss rates as a function of the distance of the cued object from fixation, camouflage, and Violation condition.*

camouflage; and the presence of a violation. Objects that were highly camouflaged, whether in a base or violation condition, and low camouflaged objects undergoing violations all suffered dramatically from their removal from central fixation. However, low camouflaged base objects showed only a slight, nonsignificant effect of distance from fixation. As indicated in Table 2, the Newman-Keuls linear tests were highly significant for the first three groups. The effect of distance from fixation for the low-camouflage Base objects was not significant. For none of the four groups were the quadratic or cubic trends significant.

DISCUSSION

Distance from fixation, target size, degree of camouflage, and the violations of the relational constraints all affected target detectability. Although there were not enough objects for a presentation of the four-way interaction, the overall picture seems reasonably clear. Large Base objects or low-camouflaged Base objects are relatively resistant to a 4- or 5-deg removal from central fixation. A 3-deg object in a normal relation to its context was detected on approximately 85% of the trials, with little decline in accuracy over a 5-deg displacement from central fixation. It is as if the effective visual field was reduced for the unexpected (that is, the violations), the camouflaged, and the small, which suffered miss rates up to 70% with only a 3- or 4-deg removal from central fixation.

Thus, even within a single fixation, an interpreter of a picture is affected by violation of the relations listed in Table 1. That both miss and false alarm rates incurred violation costs suggests that the violations produced perceptual (d') effects. Given the large number of slides with violations present, it might have been the case that the semantic processing entailed by the relational constraints could have been shortcircuited with a little practice. But this was not the case. Vio-

TABLE 2

Miss Rates as a Function of Distance from Fixation, Camouflage, and Violation Condition

Violation Condition	*Camouflage*	*Distance from Fixation*				*Newman-Keuls Linear Trend*		
		0.38	*1.75*	*3.75*	*5.75*	*df*	F	p
Base	Low	0.13	0.21	0.16	0.26	(1,13)	<1	ns
	High	0.18	0.14	0.36	0.61	(1,13)	21.91	<0.001
Violation	Low	—	0.23	0.42	0.52	(1,95)	12.66	<0.001
	High	—	0.36	0.42	0.67	(1,93)	15.74	<0.001

Note: Newman-Keuls tests were performed over seven levels of distance from fixation for the base objects and six levels for objects undergoing violations. The data shown above are grouped into larger intervals.

lation costs were of approximately the same magnitude early in practice as they were on later trial blocks.

The lack of any effect of the violations on the detection of innocent bystanders suggests that the photointerpreter cannot rely on an impression that, if a violation is present, he will sense that there is "something wrong" and know when to take another look. This conclusion is dependent upon the assumption that if the presence of violations were automatically signaled to the observer, that signal would interfere with the processing of other parts of the scene. Thus, violations of the relations for a single object will affect the detection of that object but will not affect the processing of other parts of the scene. (As more and more objects undergo violations, however, the scene will start to appear jumbled, and the detection of other objects will suffer [Biederman, 1972; 1981; Biederman, et al., 1974.]) The effects of violations of the relational constraints on object detection and the constancy of these effects throughout practice suggest that the relations are not the products of leisurely and intellectual reflection but comprise part of the (subjectively) instantaneous and obligatory processing that results when our eyes alight upon a scene. Thus, photointerpreters are affected by their knowledge of the visual world even when presentation rates are considerably shorter than the time required to make a second fixation.

That violations of the relations have true perceptual (d') effects raises an important problem with respect to the design of high-speed photointerpretation display systems. The most obvious system would be one which allows variable presentation speeds. When the interpretor wants additional "looks" he would be able to slow down the system, much as with a Moviola. The problem is that when a target is in an unexpected relation to its context, it is often not seen. Nor does the observer know that he needs another look from a "something wrong" signal. Instead it seems that one automatically "fills in" the scene with what is expected at that location in a manner that may be analogous to the filling in of the blind spot. Since this problem is exacerbated the greater the distance of the target from fixation, when specific targets are to be detected, displays greater than 10 deg (where targets could be more than 5 deg from fixation) should be used only when time is available for additional fixations. Put another way, the *effective* extent of the visual field for detecting 3-deg targets in scenes 85% of the time is approximately 10 deg.

Although any proposed display system would have to be evaluated in its own right, these results can furnish a guideline for initial specifications of performance operating for visual angles, target size, and camouflage. The correlation between camouflage ratings and miss rates documents the validity of camouflage judgments as performance predictors. Moreover, it may be possible to generalize the overall results to (a) shorter presentation durations and (b) colored slides. With respect to (a), in other experiments with these stimuli (e.g., Teitelbaum and Biederman, 1979), the authors found that a reduction of exposure duration to 100 ms yielded approximately the same performance levels as obtained in this experiment. With respect to (b), it is tempting to believe that higher performance rates would be feasible with colored photography. Obviously, there will be some cases where this will be true. Thus, if the interpreter is searching for a day-glo target on a blue background, the color cue will be an enormous help. Unfortunately for the interpreter, color is not always a reliable cue. The authors have not systematically compared the processing of information from these line drawings with colored photography, but in one study which allowed a rough comparison (but in which color could not be used as an a priori cue) performance

levels were not dramatically different from what was observed with the black and white line drawings. There is no reason to expect that the general form of the data reported here would be any different had color photography been used. But this question certainly needs additional examination.

The issue of the effects of color is complicated by several factors. Serving to facilitate detection performance is the powerful cue that can be furnished by a color difference. Also, because it is so much more pleasant to look at color photography than black and white pictures (after all, people do pay an extra $300 for color television sets), performance is likely to be sustained at high levels when viewing colored scenes. On the negative side, large color differences might serve as distracting features when the critical target detection requires the fine discrimination of contour differences. Also, shading, which is present in color photography, often becomes an irrelevant cue to contour. Consequently, detection and recognition performance is often found to be more accurate when subjects are furnished with a line drawing than when they are given a photograph (e.g., Ryan and Schwartz, 1956).

There is a task variation that would have resulted in a dramatic improvement in performance. If subjects were simply required to select the topic or "gist" of a scene, it would have been possible to present the scenes at durations much shorter than 100 to 150 ms, e.g., 50 ms, yet performance would have been much *more* accurate. Thus, Biederman, et al. (1974) found that subjects were 93% accurate in selecting between two dissimilar labels (e.g., kitchen versus street scene), one of which correctly described a scene, from only a 100-ms presentation duration of that scene. With a 50-ms stimulus exposure, subjects were able to accurately select the correct descriptor 72% of the time. Other research (Biederman, 1977; 1981) has shown that when subjects were looking directly at a target that may or may not have been undergoing a violation, they were 90% accurate in detecting when a violation was present. If detecting the presence of a Position violation is taken as a measure of scene comprehension (e.g., to know that a fire hydrant is in an inappropriate position when it is placed on top of a mailbox requires that one understand something about fire hydrants, mailboxes, and street scenes), then it is clear that very good comprehension can be shown from a 50-ms flash of a scene. Taken in conjunction with the current results, these studies suggest a system design in which the saccadic barrier is breached (i.e., brief scene presentation durations used) when only the extraction of the "gist" or "topic" of scenes is required, or large, uncamouflaged targets in expected locations are to be detected. However, slower rates will be necessary when specific targets are to be detected.

CONCLUSIONS

Objects subtending a visual angle of 3 deg and lying within 5 deg from fixation can be detected 85% of the time from a 150-ms exposure of a scene. When an object is small, camouflaged, or in an unexpected relation to its setting, the effective visual field is much smaller, and miss rates climb precipitously to 70% as objects are removed by 3 or 4 deg from fixation. Although the saccadic barrier can readily be breached for the extraction of the general setting or gist of a scene, the reliable detection of individual targets requires direct fixation.

ACKNOWLEDGMENTS

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, MH07891.

REFERENCES

Alpern, M. Effector mechanisms in vision. In J. A. Kling and L. A. Riggs (Eds.) *Woodworth and Schlosberg's experimental psychology* (3rd ed.). New York: Holt, Rinehart and Winston, 1971.

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 and J. R. Pomerantz (Eds.) *Perceptual organization*. Hillsdale, NJ: Lawrence Erlbaum, 1981.

Biederman, I., Mezzanotte, R. J., and Rabinowitz, J. C. Scene Perception: Detecting and Judging Objects Undergoing Relational Violations. *Cognitive Psychology*, in press.

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

Di Lollo, V. Temporal characteristics of iconic memory. *Nature*, 1977, *267*, 241-243.

Klatsky, G. J., Teitelbaum, R. C., Mezzanotte, R. J., and Biederman, I. Evidence for mandatory processing of contextual information in real-world scenes. Paper presented at the meetings of the Eastern Psychological Association, Hartford, CT, 1980.

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

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

Sperling, G., Budiansky, J., Spivak, J., and Johnson, M. C. Extremely rapid visual search: The maximum rate of scanning letters for the presence of a numeral. *Science*, 1971, *174*, 307-311.

Teitelbaum, R. C. Perceiving real-world scenes: Does one glance facilitate the processing from another glance even when the views differ? Unpublished doctoral dissertation, Department of Psychology, State University of New York at Buffalo, 1979.

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

Yarbus, A. L. *Eye movements and vision*. New York: Plenum, 1967.