Research Article

Scene Consistency in Object and Background Perception

Jodi L. Davenport and Mary C. Potter

Massachusetts Institute of Technology

ABSTRACT—*Does knowledge about which objects and settings tend to co-occur affect how people interpret an image? The effects of consistency on perception were investigated using manipulated photographs containing a foreground object that was either semantically consistent or inconsistent with its setting. In four experiments, participants reported the foreground object, the setting, or both after seeing each picture for 80 ms followed by a mask. In Experiment 1, objects were identified more accurately in a consistent than an inconsistent setting. In Experiment 2, backgrounds were identified more accurately when they contained a consistent rather than an inconsistent foreground object. In Experiment 3, objects were presented without backgrounds and backgrounds without objects; comparison with the other experiments indicated that objects were identified better in isolation than when presented with a background, but there was no difference in accuracy for backgrounds whether they appeared with a foreground object or not. Finally, in Experiment 4, consistency effects remained when both objects and backgrounds were reported. Semantic consistency information is available when a scene is glimpsed briefly and affects both object and background perception. Objects and their settings are processed interactively and not in isolation.*

Objects and settings tend to co-occur in the world; cars usually appear in streets, not in kitchens. Is initial perception of a scene influenced by this knowledge? Are objects more easily perceived in a typical than in an unusual setting? Is perception of a background setting facilitated when the scene includes a probable rather than an improbable object? Despite some well-known studies demonstrating that an appropriate setting facilitates object processing, other evidence has suggested that objects are processed in isolation from the scene in which they occur. No prior studies have addressed the question of whether a conspicuous object can influence perception of the background setting.

Using line drawings, eye-tracking studies have shown that objects inconsistent with a scene are fixated longer than consistent objects (De Graef, Christiaens, & d'Ydewalle, 1990; Friedman, 1979; Henderson, Weeks, & Hollingworth, 1999; Loftus & Mackworth, 1978). Fixation durations may reflect both object identification processing and postidentification processing, such as integrating object identity with scene context. As fixation times and patterns vary greatly across tasks, eye-tracking measures currently do not provide clear evidence that consistency with the scene speeds the initial identification of an object.

Mixed results have been obtained in behavioral studies investigating the effects of scene context on object perception. Some find a consistency advantage. Objects are identified more accurately when primed by a semantically consistent scene than when primed by an inconsistent scene (Palmer, 1975). Additionally, objects are detected more accurately and named more quickly when they appear in a semantically consistent setting (Biederman, Mezzanotte, & Rabinowitz, 1982; Boyce & Pollatsek, 1992; Boyce, Pollatsek, & Rayner, 1989; Cheng & Simons, 2001). However, Hollingworth and Henderson (1998, 1999; Henderson & Hollingworth, 1999) attributed some of the prior results to response bias. When they controlled for false alarm rates and a location-cue advantage in the Biederman et al. study, Hollingworth and Henderson failed to find a consistency advantage. They concluded that object identification is functionally isolated from information about the scene context.

The current experiments shed new light on the question of consistency in object and scene perception by using new materials and a new task and by asking new questions. Most previous work investigating consistency effects has been limited to black-and-white line drawings. In our studies, we used naturalistic color photographs as color has been shown to improve both object and scene recognition (Oliva & Schyns, 2000; Wurm, Legge, Isenberg, & Luebker, 1993). Each picture consisted of a background scene and a single foreground object that was either semantically consistent or semantically inconsistent with the scene. Participants saw each picture only once, for 80 ms followed by a mask, and were asked to type the name of the foreground object, the background, or both. Immediate report of identity may be a more sensitive gauge of processing than a yes/no object detection or two-alternative forced-choice task, as no specific information is provided before viewing or in the testing phase. In addition, because each picture was shown a single time, participants could not be biased by previous exposures to the stimuli or influenced by having seen a given object in multiple settings or a given scene with different objects. Twelve native English speakers with normal or corrected-to-normal vision volunteered for each experiment and were paid for their participation. None participated in more than one of the reported studies.

Address correspondence to Jodi L. Davenport, Department of Brain and Cognitive Sciences, MIT, NE20-439, 77 Massachusetts Ave., Cambridge, MA 02139; e-mail: jodi@mit.edu.

Our stimuli and methods—color photographs, brief masked presentations, immediate report of identity, and no repetition of materials—allowed us to address new questions about consistency in scene perception. Does consistency between an object and its setting affect the initial perception of the object? Does a foreground object affect the initial perception of its background? Finally, are consistency effects modulated when both objects and their backgrounds must be attended and reported? The following experiments provide evidence that semantic consistency affects the perception of both a foreground object and its background in the first glimpse of an unfamiliar picture.

EXPERIMENT 1: THE EFFECT OF CONSISTENCY ON OBJECT PERCEPTION

In the first experiment, we asked whether perception of a foreground object is influenced by its consistency with its setting when a scene is presented for a very short duration. Participants viewed each of 28 scenes a single time and were instructed to report the foreground object of each scene.

Method

Materials and Apparatus

Stimuli were 28 color photographs of diverse settings, each with one foreground object edited into the picture. The backgrounds and objects were taken from commercially available CDs of photographs, the Web, and other sources. The objects in the stimuli were animals, vehicles, people, articles of furniture, and the like. To ensure that the objects and settings had agreed-upon common names, we asked 8 raters in a pilot study to name each background without a foreground object and each object on a plain background, after viewing each for 500 ms without a mask. Only consistently named backgrounds and objects were selected for the experiment. The names (including paraphrases and synonyms) generated by raters were used to score the responses. A list of the backgrounds and objects is given in Table 1.

For each background image, a semantically consistent object was selected from a different source photograph. The object was chosen to be likely to appear in the matched background setting (e.g., a zebra in an African plains setting, a sofa in a living room) and was pasted into the background using Adobe Photoshop 7.0. Semantically inconsistent scenes were created by pairing scenes and exchanging their consistent objects (e.g., putting the zebra in the living room and the sofa in the African plains). Although there were often other objects in the background, such as a fireplace in the living room and trees on the plain, the critical object was the only one clearly in the foreground. The foreground object was pasted so size and support relations were not violated. See Figure 1 for examples of the stimuli.

A set of masks was generated by cutting six other pictures into a 20×20 grid of rectangles and rearranging them randomly.

All pictures and masks consisted of jpeg files 500 pixels in width by 300 pixels in height. They were presented on an Apple PowerMac G3 computer with a 400-MHz processor. The 17-in. monitor was set to a resolution of 1024×768 pixels with a refresh rate of 75 Hz. As displayed, pictures were 17.64×10.53 cm, subtending approximately 22° of visual angle horizontally and 13° of visual angle vertically when viewed from a normal viewing distance of 45 cm. The experiments were written in Matlab, using the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997). The pictures appeared on a black background that was present throughout the experiment. The room was illuminated normally.

TABLE 1
Backgrounds and Objects Used as Stimuli

Stage–Ballerina	Road–Cyclist
Living room–Sofa	African plains–Zebra
Football field–Football player	Church–Priest
Beach–Sand castle	Mud–Pig
Racetrack–Race car	Farm–Tractor
Park–Jogger	Arena–Bull
Bowling alley–Bowler	Range–Buffalo
Ice rink–Figure skater	Horse track–Racehorse
Intersection–Ambulance	Desert–Camel
Lake–Duck	Snowy mountain–Sledder
Earth–Space shuttle	Underwater–Sea turtle
Mountain valley–Woman on donkey	Parking lot–Car
Parade–Trumpeter	Forest–Deer
War–Soldier	Library–Student

Note. Stimuli that appear on the same line were paired when objects were swapped to form the inconsistent pictures.

Design and Procedure

On each of the 28 trials, a single picture was presented and followed by a mask. Each participant saw half of the background pairs with consistent objects and half with inconsistent objects; consistent and inconsistent trials were randomly intermixed. Each object and background appeared only once.

Each trial began with the phrase "press any key to continue." After the key press, a fixation cross "+" appeared for 300 ms, followed by a blank screen for 200 ms, the test picture for 80 ms, and a mask for 200 ms. A dialogue box appeared immediately after the mask, and participants were to type their response in the box.

Participants were instructed to report the foreground object in each picture and were informed that the object might or might not fit the background. If participants missed the picture, they were able to leave the response box blank. There were six practice trials with backgrounds and objects not used in the main experiment.

Scoring

All results were scored blind to condition. Names provided by the raters in the norming study and synonyms at an equal level of descriptiveness were marked as accurate (e.g., "runner" and "jogger"). Names that were at a more general level of description than those given by raters were marked as incorrect (e.g., "animal" instead of "zebra").

In the inconsistent condition, a response was scored as an intrusion if participants guessed the object that would have been consistent with the background (e.g., if they said "sofa" when a zebra was presented in the living room scene). To correct for such pure guesses, for each intrusion in the inconsistent condition we subtracted one correct response from that participant's score for consistent trials. A similar correction was made in the item analysis.[1]

[1]Intrusions were rare. Out of 168 inconsistent trials in each experiment, 1 object intrusion occurred in Experiment 1, 6 background intrusions occurred in Experiment 2, and 1 object and 12 background intrusions occurred in Experiment 4.

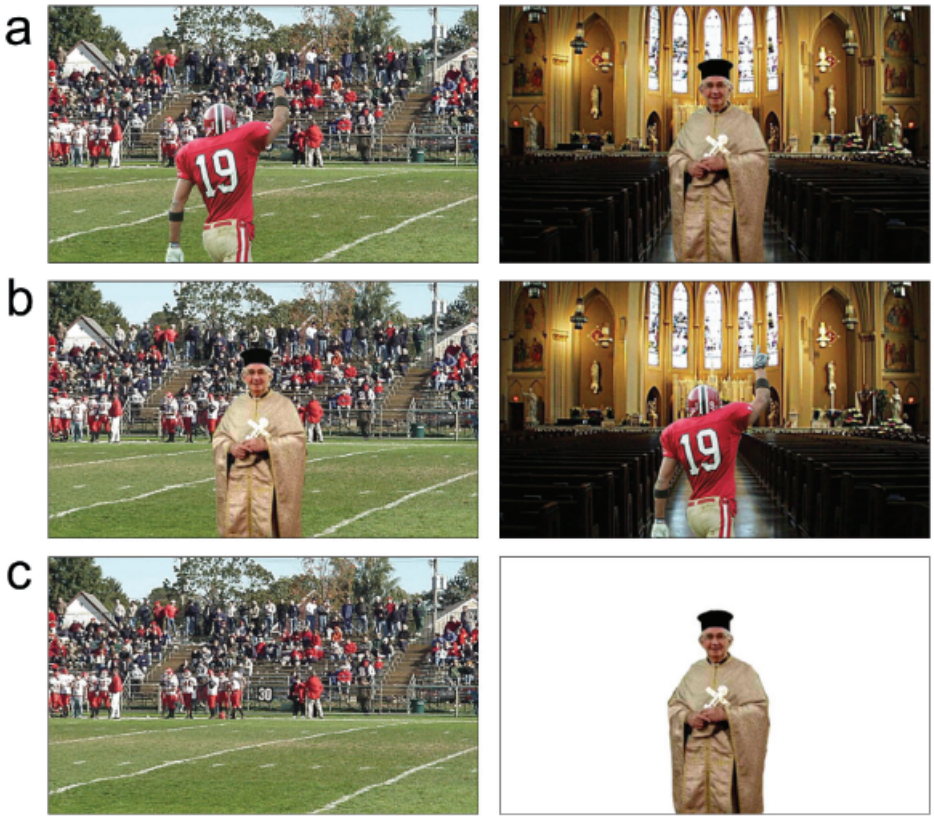

Fig. 1. Examples of consistent scenes (a), inconsistent scenes (b), and isolated objects and backgrounds (c).

Results and Discussion

As shown in the top panel of Figure 2, participants reported objects more accurately when they appeared with a consistent background (.82) than when they appeared with an inconsistent background (.68). An analysis of variance (ANOVA) with consistency as a within-subjects variable found a highly significant main effect of consistency, $F(1, 11) = 14.73$, $p < .01$, $\eta^2 = .59$. In an ANOVA with items as random variables, the consistency effect remained significant, $F(1, 27) = 5.55$, $p < .05$.

The results suggest that when a scene is glimpsed briefly, the consistency of an object with its background affects its perception, even when the background can be ignored.

EXPERIMENT 2: THE EFFECT OF A FOREGROUND OBJECT ON BACKGROUND PERCEPTION

Experiment 1 provided evidence that object processing could be affected by an object's consistency with its background. Because a salient foreground object may also evoke a schema or context, in Experiment 2 we examined whether a foreground object could affect the processing of its background. Experiment 2 was identical to the first experiment, with the exception that participants were asked to report the background or type of place rather than the foreground object.

Method

The method was identical to that of Experiment 1, with the exception that instructions in Experiment 2 asked participants to report just the background setting of each picture.

Results and Discussion

As shown in the top panel of Figure 2, participants reported backgrounds appearing with a consistent foreground object (.70) more accurately than backgrounds with an inconsistent foreground object (.54). A within-subjects ANOVA revealed a significant main effect of consistency, $F(1, 11) = 7.40$, $p < .05$, $\eta^2 = .4$. In an ANOVA with items as a random variable, the main effect of consistency was again significant, $F(1, 27) = 9.45$, $p < .01$.

The results of Experiment 2 provide new evidence that perception of a background may be modulated by a to-be-ignored foreground object. Settings or backgrounds do not seem to be processed independently of the objects they contain. Instead, perception of an object and perception of its background appear to occur concurrently and interactively even when only one needs to be attended.

EXPERIMENT 3: OBJECTS WITHOUT BACKGROUNDS AND BACKGROUNDS WITHOUT OBJECTS

To establish a baseline for ability to report the objects without backgrounds and the backgrounds without objects, we had subjects report backgrounds alone and objects alone.

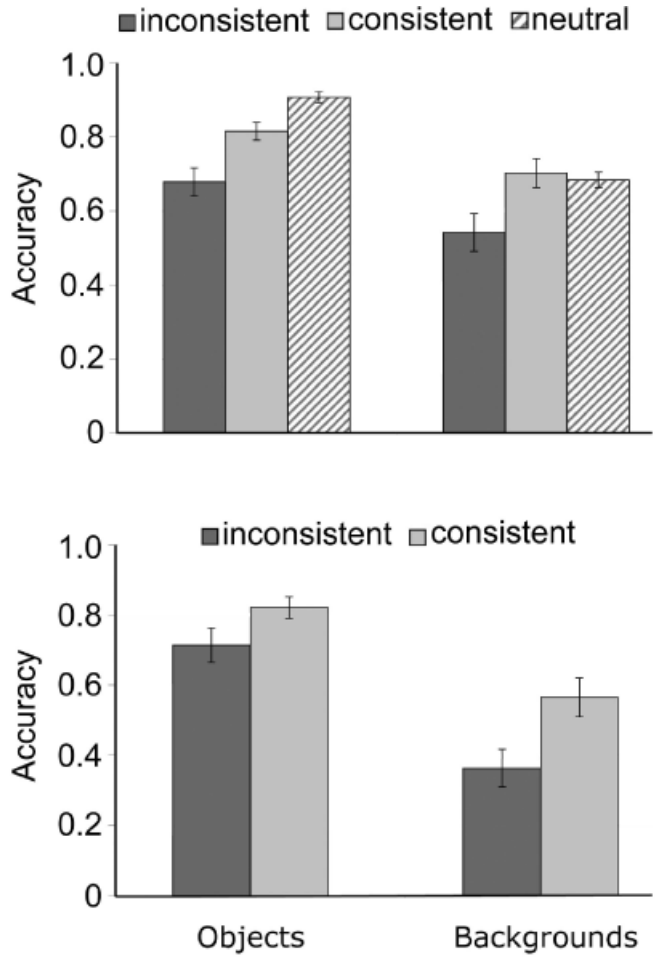

Fig. 2. Accuracy in reporting foreground objects and backgrounds. The top panel shows results when only objects or only backgrounds were reported (Experiments 1–3), and the bottom panel shows results when both objects and backgrounds were reported (Experiment 4). In Experiments 1, 2, and 4, the foreground object was consistent with its background in half the scenes and inconsistent with its background in the other half of the scenes. In Experiment 3 (''neutral''), objects were presented without backgrounds, and backgrounds were presented without foreground objects.

Method

The method was like that of Experiment 1, except as specified. Stimuli were the pictures of backgrounds with no object pasted in and the objects alone on a white background (see Fig. 1c). The trials were blocked and counterbalanced such that half the participants saw and named objects in the first block and backgrounds in the second block, whereas the other half saw and named backgrounds in the first block and objects in the second block. The order of trials within each block was randomized for each subject.

Results and Discussion

An ANOVA was carried out with task (report objects vs. report backgrounds) as a within-subjects variable. Objects (.91) were reported more accurately than backgrounds (.68), $F(1, 10) = 111.166$, $p < .001$. In a series of planned comparisons, accuracy in Experiment 3 was compared with accuracy in Experiments 1 (report objects) and 2 (report backgrounds; see Fig. 2, top panel). Objects presented without a background were reported more accurately than objects with either a consistent background, $F(1, 22) = 9.297$, $p < .01$, or an inconsistent background, $F(1, 22) = 33.56$, $p < .001$. Backgrounds presented with no foreground object were reported more accurately than backgrounds appearing with an inconsistent object, $F(1, 22) = 6.74$, $p < .05$, but at the same level of accuracy as backgrounds with a consistent object, $F(1, 22) < 1.0$.

The results for reporting the backgrounds suggest that an inconsistent foreground object interferes with the processing of the setting and that briefly presented images are more difficult to recognize when an inconsistency is present than when there is no inconsistency. The results for reporting the objects are more difficult to interpret. Object identification was best when objects were presented in isolation, whereas background identification was equally good with and without a foreground object. The difference may be that an isolated object benefits from a clear contour, whereas the outer contours of a background do not change regardless of whether a foreground object appears with the background.

EXPERIMENT 4: CONSISTENCY EFFECTS WHEN BOTH BACKGROUND AND OBJECT MUST BE REPORTED

Experiments 1 and 2 demonstrated that the consistency of the background with a foreground object influences perception of the object and that the consistency of a foreground object with its background influences perception of the background. However, in both experiments, participants were instructed to attend to only the object or the background. In Experiment 4, we asked whether the same pattern of results would be found when the task was to report both the object and the background. In the prior experiments, did selective attention facilitate participants' ability to report just the object or just the background? Would perception of one, the other, or both suffer when both had to be reported?

Method

The method was identical to that of Experiment 1 with the exception that the instructions in Experiment 4 asked participants to report both the foreground object and the background setting, in either order. The dialogue box had two lines for responses.

Results and Discussion

Separate ANOVAs were carried out to determine the effects of consistency and item type on accuracy in the current experiment (see Fig. 2, bottom panel) and to compare the effects of item type on accuracy across experiments.

In an ANOVA with consistency and item types as variables, the consistency effect was again highly significant, $F(1, 11) = 31.51$, $p < .001$, $\eta^2 = .74$, with items (both objects and backgrounds) in consistent scenes (.69) reported more accurately than items in inconsistent scenes (.54). The main effect of item type was also significant, $F(1, 11) = 44.78$, $p < .001$, $\eta^2 = .8$, with objects (.77) reported more accurately than backgrounds (.46). The interaction between consistency and item type was not significant, $F(1, 11) = 1.98$, $p = .19$. Item analyses were conducted separately for backgrounds and objects. The consistency effect remained highly significant for

backgrounds across items, $F(1, 27) = 28.59$, $p < .001$, and approached significance for objects, $F(1, 27) = 2.94$, $p = .09$.

Comparisons with Experiments 1 (objects) and 2 (backgrounds) were conducted independently, to determine whether attending to both objects and backgrounds had an effect on processing. In an ANOVA comparing accuracy for objects in Experiment 1 and Experiment 4, there was a main effect of consistency, $F(1, 22) = 19.78$, $p < .001$. However, there was no main effect of experiment, $F(1, 22) = 0.23$, and no significant interaction of consistency and experiment, $F(1, 22) = 0.30$.

Comparisons of accuracy in reporting backgrounds in Experiment 2 and Experiment 4 revealed a highly significant main effect of consistency, $F(1, 22) = 23.70$, $p < .001$, and a main effect of experiment, $F(1, 22) = 6.832$, $p < .05$, with backgrounds reported more accurately in Experiment 2 (.62), when only backgrounds were reported, than in Experiment 4 (.46), which required both backgrounds and objects to be reported.

Experiment 4 provides further evidence that consistency information is available when an image is glimpsed briefly and affects the processing of objects and backgrounds. Reporting both objects and backgrounds had a selective cost for backgrounds. As backgrounds were reported first 53% of the time and objects were reported first 47% of the time, the reduced accuracy in background perception cannot be attributed to memory decay in reporting the second item. The processing of backgrounds may require greater attention than the processing of objects, or objects may have had an advantage because they usually appeared closer to fixation than backgrounds did.

GENERAL DISCUSSION

The current experiments provide evidence that knowledge about the world affects observers' perception. Information about semantic consistency is available when an image is presented for a very brief duration of 80 ms and affects how objects and their settings are perceived. In the experiments, objects and backgrounds were reported more accurately when they were semantically consistent with each other than when they were inconsistent. Experiment 1 demonstrated that objects in a consistent background setting were reported more accurately than objects in an inconsistent setting. Experiment 2 provided the first clear evidence that backgrounds may also be influenced by their consistency with objects appearing in their foreground. Experiment 4 showed that the consistency effect remained even when both objects and their settings were attended, and that perception of backgrounds was selectively impaired in that condition. In each condition, objects were perceived more accurately than backgrounds. This asymmetry may be intrinsic to perception of objects and backgrounds, but this result should be interpreted with caution.

The current findings are in line with previous studies that found an effect of semantic consistency on object perception (Biederman, 1972; Biederman et al., 1982; Boyce & Pollatsek, 1992; Boyce et al., 1989; Cheng & Simons, 2001; Friedman, 1979). However, the results conflict with those of Hollingworth and Henderson (1998, 1999; Henderson & Hollingworth, 1999), who failed to find an effect of consistency in a series of two-alternative forced-choice object detection tasks using line drawing as stimuli. They proposed that object perception is functionally isolated from stored semantic knowledge about scenes.

The discrepancies between our results and the findings of Hollingworth and Henderson (1998, 1999) are likely due to differences in tasks and materials. The identification task used in the present experiments may have been a more sensitive measure of perception than the two-alternative forced-choice task, as subjects had no information about the scene prior to viewing and were not forced to guess if they were unable to see the picture. Also, the format of Hollingworth and Henderson's studies may have given inconsistent objects an asymmetrical advantage. On inconsistent trials, the object that did not fit with the scene was always the object later tested; however, on consistent trials, any one of many objects could be selected for testing. Our materials differed as well. Compared with gray-scale photographs or line drawings, full-color photographs may improve object and scene recognition (Oliva & Schyns, 2000; Wurm et al., 1993) and enhance the ability to detect semantic inconsistencies (Cheng & Simons, 2001). In addition, our pictures always included a prominent foreground object, which may have been more salient than smaller objects in line drawings.

In addition to testing the effects of background consistency on object perception, our study addressed new questions. First, we investigated the effects of consistency on initial perception of stimuli with very brief presentation times. Prior work has studied long-term memory, used eye fixation data and long scene exposures, or repeated scene and object stimuli numerous times. Participants in our experiments were exposed to each object and background a single time, for immediate report, whereas subjects in prior studies saw each object and background several times. If consistency information is most critical the first time a scene is processed, the repeated viewing of objects and scenes in various combinations would reduce the consistency effect. Second, we asked for the first time whether a foreground object affects background perception, as well as the reverse.

Our findings suggest that objects and scenes are processed interactively, and that knowledge of which objects and settings tend to cooccur influences perception. Objects and backgrounds may be mutually constraining; less perceptual information may be required for identification when scenes are semantically consistent than when they are inconsistent. A qualitative overview of our participants' reports indicated that the types of errors made reflected a lack of detailed perceptual information. Many errors consisted of omitted items or vague responses at a more general level of description than the correct response (e.g., reporting "indoors" instead of "living room"). Other incorrect responses were names of perceptually similar but conceptually dissimilar settings or objects (e.g., reporting "ice rink" when the actual stimulus was a car racetrack with a passing resemblance to an ice rink).

Experiment 4 suggests that a foreground object may have a special status in processing, as objects were reported as accurately when both objects and backgrounds were reported as when only objects were reported. Backgrounds, in contrast, demonstrated a cost when both objects and backgrounds were reported, though this did not reduce the consistency effect. Foreground objects may automatically attract attention, and may contribute substantially to the gist of a picture that is extracted early in processing (Biederman, 1972; Potter, 1975, 1976). The current study investigated consistency effects only when there was a single foreground object. Further work is planned to investigate interactions among two or more foreground objects and a background.

In conclusion, the present research provides strong evidence that information about the semantic relationship between objects and their

background is available when a scene is presented briefly and affects perception of the scene. A foreground object and the background of a scene seem to be processed interactively, reflecting knowledge about which objects and settings co-occur in the world.

Acknowledgments—This research was supported by National Institute of Mental Health Grants MH47432 and MH20007. We thank Casey Dugan and Chris Meyer for help with creating the stimuli and programming.

REFERENCES

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

Biederman, I., Mezzanotte, R.J., & Rabinowitz, J.C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology*, *14*, 143–177.

Boyce, S.J., & Pollatsek, A. (1992). Identification of objects in scenes: The role of scene background in object naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 531–543.

Boyce, S.J., Pollatsek, A., & Rayner, K. (1989). Effect of background information on object identification. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 556–566.

Brainard, D.H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436.

Cheng, E.K., & Simons, D.J. (2001, November). *Perceiving the internal consistency of scenes*. Poster presented at the annual meeting of the Psychonomic Society, Orlando, FL.

De Graef, P., Christiaens, D., & d'Ydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychological Research*, *52*, 317–329.

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

Henderson, J.M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology*, *50*, 243–271.

Henderson, J.M., Weeks, P.A., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 210–228.

Hollingworth, A., & Henderson, J.M. (1998). Does consistent scene context facilitate object perception? *Journal of Experimental Psychology: General*, *127*, 398–415.

Hollingworth, A., & Henderson, J.M. (1999). Object identification is isolated from scene semantic constraint: Evidence from object type and token discrimination. *Acta Psychologica*, *102*, 319–343.

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

Oliva, A., & Schyns, P.G. (2000). Diagnostic colors mediate scene recognition. *Cognitive Psychology*, *41*, 176–210.

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

Pelli, D.G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.

Potter, M.C. (1975). Meaning in visual search. *Science*, *187*, 965–966.

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

Wurm, L.H., Legge, G.E., Isenberg, L.M., & Luebker, A. (1993). Color improves object recognition in normal and low vision. *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 899–911.

(RECEIVED 6/9/03; REVISION ACCEPTED 9/2/03)