ELSEVIER

Acta Psychologica 102 (1999) 319–343

acta psychologica

Object identification is isolated from scene semantic constraint: evidence from object type and token discrimination

Andrew Hollingworth *, John M. Henderson [1]

Department of Psychology, 129 Psychology Research Building, Michigan State University, East Lansing, MI 48824-1117, USA

Received 13 May 1998; received in revised form 31 August 1998; accepted 6 October 1998

Abstract

Two models of the interaction between scene meaning and object identification were tested: the *description enhancement model* and the *criterion modulation model*. The former proposes that the early activation of a scene schema facilitates the initial perceptual analysis of schema-consistent objects, the latter that schema activation modulates the amount of information necessary to indicate the presence of an object of a particular perceptual type. In Experiment 1, we employed a forced-choice, type-discrimination paradigm. Participants were asked to determine which of two semantically consistent objects or which of two semantically inconsistent objects had appeared in a briefly presented scene. Contrary to the prediction derived from both of these models, discrimination performance was better for semantically inconsistent versus consistent objects. In Experiments 2 and 3 we introduced a forced-choice, token-discrimination paradigm to further test the description enhancement model. Contrary to the prediction of that model, discrimination performance was no better for semantically consistent versus inconsistent tokens. These results suggest that both the initial perceptual analysis of an object and the matching of an object's constructed visual description to stored descriptions are isolated from stored knowledge about real-world contingencies between scenes and objects.

* Corresponding author. E-mail: andrew@eyelab.msu.edu
[1] E-mail: john@eyelab.msu.edu

0001-6918/99/$ – see front matter
PII: S0001-6918(98)00053-5

PsycINFO classification: 2323

Keywords: Object identification; Scene perception; Context effects

1. Introduction

When humans view a natural scene, to what extent is the identification of individual objects influenced by stored knowledge about that scene type? For example, is the identification of a swing-set influenced by whether it appears in a playground (consistent with our knowledge of playgrounds) or in a bedroom (inconsistent with our knowledge of bedrooms)? This question has received considerable attention within scene perception research (see Henderson & Hollingworth, 1998), because it addresses a central issue in visual perception and cognition: To what degree does our knowledge of the world influence perception of the visual environment? In scene perception research, three main positions have emerged. First, contextual constraint may interact with the initial perceptual analysis of objects. Second, contextual constraint may modulate the threshold amount of information necessary to indicate that an object of a particular perceptual type is present. Third, object identification processes may be isolated from semantic information stored in memory about real-world contingencies between objects and scenes. [2]

The first of these positions we will term the *description enhancement model* (Biederman, 1981; Biederman, Mezzanotte & Rabinowitz, 1982). [3] According to this view, the rapid identification of a scene as a particular type activates a memory representation (a *schema*) that contains information about the objects and spatial relations among objects that form that type. The activation of a scene schema facilitates the initial perceptual analysis (i.e., the extraction of features and/or construction of a visual description) of schema-consistent objects, leading to facilitated identification of objects that are consistent versus inconsistent with scene context. The second position we will term the *criterion modulation model* (Bar & Ullman,

[2] We assume that object identification consists of at least three component processes. First, the current pattern of retinal stimulation is translated into perceptual primitives. Second, visual descriptions of the object tokens in the scene are constructed from these primitives. Third, constructed descriptions are matched to stored long-term memory descriptions of object types. When a match is found, identification has occurred, and semantic information stored in memory about that object type becomes available.

[3] In the past (Henderson & Hollingworth, 1999; Hollingworth & Henderson, 1998), we have referred to the two positions described in this paragraph as the *perceptual schema model* and the *priming model*. However, because the modulation of a perceptual matching criterion could also be considered a version of a perceptual schema model, we now use the more precise term *description enhancement model* to describe the position that schema activation influences the initial perceptual analysis of objects in a scene. In addition, we have replaced the term *priming model* with *criterion modulation model* to describe the position that schema activation influences activation thresholds during the matching stage of object identification.

1996; Friedman, 1979; Friedman & Liebelt, 1981; Kosslyn, 1994; Palmer, 1975; Ullman, 1996). This view proposes that the contextual constraint generated by activation of a scene schema influences the stage when the constructed visual description of an object token is matched against stored descriptions of object types. The activation of a scene schema serves to lower the threshold amount of information needed to indicate a match to the stored descriptions of schema-consistent objects. As a result, relatively less perceptual information will need to be encoded to select the stored description of an object consistent with scene context compared to that of an object inconsistent with scene context, facilitating the identification of the former compared to the latter.

The description enhancement model and the criterion modulation model both hypothesize that identification is facilitated for objects that are consistent with the scene in which they appear. The description enhancement model gives rise to the further hypothesis that the constructed visual description of a consistent object should be more detailed (and/or complete) than that of an inconsistent object, because scene context serves to facilitate the initial encoding of perceptual information for consistent objects. The criterion modulation model, however, does not lead as directly to a hypothesis regarding the visual description of consistent versus inconsistent objects. Friedman (1979) has proposed that more perceptual information will be encoded for inconsistent objects under free viewing conditions, because those objects will require more perceptual analysis to reach an identification threshold. However, this proposal requires adding an assumption to the criterion modulation model that perceptual encoding ceases as soon as an object is identified.

The third position, the *functional isolation model*, proposes that object identification is not influenced by the scene context in which an object appears, because object identification processes are isolated from knowledge about the real-world contingencies between scenes and objects (Hollingworth & Henderson, 1998). Thus, the functional isolation model predicts that experiments examining the identification of objects in real-world scenes should find no effect of the relation between object and scene. However, context effects on object processing measures may arise in experiments that are sensitive to later, post-identification influences of scene constraint.

Support for the criterion modulation and description enhancement models has come from two principal paradigms: an eye movement paradigm and an object detection paradigm. In eye movement studies, the length of time an object is fixated during the free viewing of a scene tends to be shorter when the object is semantically consistent with the scene (i.e., likely to appear within the scene) compared to when it is semantically inconsistent (i.e., unlikely to appear within the scene) (De Graef, Christiaens & d'Ydewalle, 1990; Friedman, 1979; Henderson, Weeks & Hollingworth, 1999). This result has been interpreted as support for the general claim that consistent scene context facilitates object identification, and for the criterion modulation model in particular (Friedman, 1979). However, the conclusion that differences in fixation duration uniquely reflect differences in ease of identification is not well supported. It is clear that fixation durations are influenced not only by ease of identification, but by other, post-identification factors, such as conceptual

integration and memory encoding (Henderson, 1992; Henderson & Hollingworth, 1999; Rayner & Pollatsek, 1992).

The strongest evidence supporting the hypothesis that consistent scene context facilitates object identification comes from the object detection paradigm introduced by Biederman and colleagues (Biederman, 1981; Biederman, Mezzanotte & Rabinowitz, 1982). Biederman, Mezzanotte and Rabinowitz (1982) asked participants to decide whether a target object had appeared within a briefly presented scene at a cued location. During each trial, a label naming a target object was presented until the participant was ready to continue, followed by a line drawing of a scene for 150 ms, followed by a pattern mask with an embedded location cue. Participants indicated whether the object described by the target label had appeared in the scene at the cued location. The key manipulation in this paradigm was the consistency between the object appearing at the cued location and the scene. The cued object could be either consistent with the scene or violate scene constraint along one or more dimensions, including probability (semantic consistency), position, size, support, and interposition (whether the object occluded objects behind it or was transparent). Detection sensitivity (d') was best when the cued object did not violate the constraints imposed by scene meaning.

Using a similar paradigm, Boyce, Pollatsek and Rayner (1989) manipulated the consistency of the cued object with both the global scene and with other cohort objects appearing in the scene. Detection sensitivity was facilitated when the cued object was semantically consistent with the global scene in which it appeared compared with when it was semantically inconsistent with the global scene. In contrast, there was no effect of the consistency of the cued object with the cohort objects in the scene. Boyce et al. concluded that the global meaning of the scene, rather than the specific objects present in the scene, is functional in facilitating object identification.

The above results provide the strongest evidence that consistent scene context facilitates object identification. However, a number of methodological problems have been identified regarding these paradigms (De Graef, 1992; De Graef, Christiaens & d'Ydewalle, 1990; Henderson, 1992; Hollingworth & Henderson, 1998). First, the signal detection methodology of previous object detection studies may not have adequately eliminated response bias from sensitivity measures. These studies (Biederman, Mezzanotte & Rabinowitz, 1982; Boyce, Pollatsek & Rayner, 1989) did not compute sensitivity using the correct detection of a particular signal when it was present and the false detection of the same signal when it was absent, as required by signal detection theory. Catch trials presented the same scene (and cued object) as in target-present trials but merely changed the label appearing before the scene. In addition, in the Biederman et al. studies (Biederman, Mezzanotte & Rabinowitz, 1982; Biederman, Teitelbaum, & Mezzanotte, 1983) false alarms were computed in both consistent and inconsistent cued object conditions by averaging across catch trials on which the target label was semantically consistent and semantically inconsistent with the scene. Because the false alarm rate was lower when the target label was inconsistent with the scene, the averaged false alarm rate likely resulted in an overestimation of sensitivity in the consistent cued object condition and an

underestimation of sensitivity in the violation conditions, as shown by Hollingworth and Henderson (1998).

The second concern with these experiments (Biederman, Mezzanotte & Rabinowitz, 1982; Boyce, Pollatsek & Rayner, 1989) is that participants may have searched areas of the scene where the target object was likely to be found. If the spatial positions of semantically consistent objects were more predictable than those of inconsistent objects, detection of the former would have been facilitated compared to the latter, even if there were no differences in the perceptibility of each type of object (Hollingworth & Henderson, 1998). Supporting this idea, Henderson, Weeks and Hollingworth (1999) found that participants could more quickly locate semantically consistent versus inconsistent objects in a free-viewing, visual search task.

Hollingworth and Henderson (1998) explored both of these issues. To investigate concerns about signal detection methodology, we replicated the Biederman, Mezzanotte and Rabinowitz (1982) study first using the original signal detection design and then using a corrected design in which participants attempted to detect the same object on corresponding target-present and catch trials. The experiment using the original design replicated the consistent object detection advantage found by Biederman, Mezzanotte and Rabinowitz and Boyce, Pollatsek and Rayner (1989). However, the experiment using the corrected design showed no advantage for the detection of semantically consistent versus semantically inconsistent objects. In addition, Hollingworth and Henderson tested whether differences in search efficiency influence performance in the object detection paradigm. Instead of presenting the target label before the scene, we presented it after the scene, so that participants could not use the positional constraints of the scene to facilitate consistent object search. Contrary to earlier studies, we found a reliable advantage for the detection of semantically *inconsistent* objects. These results suggest that the consistent object advantage in previous object detection experiments likely arose from the inadequate control of response bias and from differences in search efficiency as a function of object consistency, not from the faciliatory influence of scene context on consistent object identification.

To investigate the identification of objects in scenes independently of response bias, Hollingworth and Henderson (1998) introduced a, forced-choice object type-discrimination paradigm, similar to that developed by Reicher (1969). A scene was presented briefly (250 ms) and could contain either one of two semantically consistent target objects or one of two semantically inconsistent target objects. The scene was followed by a pattern mask for 30 ms, and the mask was followed immediately by a forced-choice response screen displaying two labels corresponding either to the two consistent targets or to the two inconsistent targets. Under these conditions, response bias should be eliminated because the two object alternatives are of equivalent semantic consistency. Contrary to the prediction of the description enhancement and criterion modulation models, Hollingworth and Henderson found no advantage for the discrimination of consistent versus inconsistent objects: The non-reliable trend was in the direction of better inconsistent object discrimination. These results indicate that object identification may be isolated from stored knowledge about the type of scene in which an object appears.

1.1. The present study

One goal of the present study was to test further whether consistent scene context facilitates object identification by extending our forced-choice, type-discrimination paradigm (Hollingworth & Henderson, 1998, Experiment 4). One potential concern with this paradigm is that the initial scene was presented for 250 ms, 100 ms longer than other studies of object identification in scenes (Biederman, Mezzanotte & Rabinowitz, 1982; Boyce, Pollatsek & Rayner, 1989). It is possible that consistent scene context facilitates object identification very early within the visual processing of a scene. Later in viewing, however, attention may be preferentially directed to inconsistent objects, perhaps because they are more difficult to integrate with other conceptual information in the scene and require further analysis. If this is correct, then our paradigm may not have been sensitive to early, faciliatory effects of consistent context, but instead may have primarily reflected later allocation of attention to inconsistent objects. To address this concern, in Experiment 1 we replicated the type-discrimination paradigm but presented the scene for 150 rather than 250 ms.

A second goal of this study was to provide an additional test of the description enhancement model. The description enhancement model proposes that the processes leading to the generation of a visual description of an object are facilitated when that object is consistent versus inconsistent with scene context. As a result, the constructed visual description of a consistent object should be more detailed (and/or complete) than that of an inconsistent object. In Experiments 2 and 3 we introduced a forced-choice, token-discrimination paradigm to test specifically the relative perceptual detail of the representations constructed for objects that are semantically consistent versus inconsistent with the scene in which they appear. Because two tokens of the same object type (e.g., a sedan and a sports car) will have similar conceptual-level representations, discrimination performance will depend primarily on the encoding of perceptual detail from the object token presented in the scene.

2. Experiment 1

Experiment 1 replicated the forced-choice, type-discrimination paradigm developed by Hollingworth and Henderson (1998). To assure that discrimination performance reflected the initial influence of scene context on object identification, the presentation of the scene was limited to 150 ms. On each trial, the presented scene could contain either one of two semantically consistent target objects or one of two semantically inconsistent target objects. Semantically consistent target objects were chosen as likely to appear in the scene. Semantically inconsistent target objects were chosen as unlikely to appear in the scene. Fig. 1 shows an example of a stimulus scene and the semantic consistency manipulation. The scene was followed by a pattern mask for 30 ms, and the mask was followed by a forced-choice response screen containing two labels. One object label named the target object presented in the scene, and the second label named the other target object of the equivalent semantic consistency. For example, when a consistent target object (a cello or a harp)

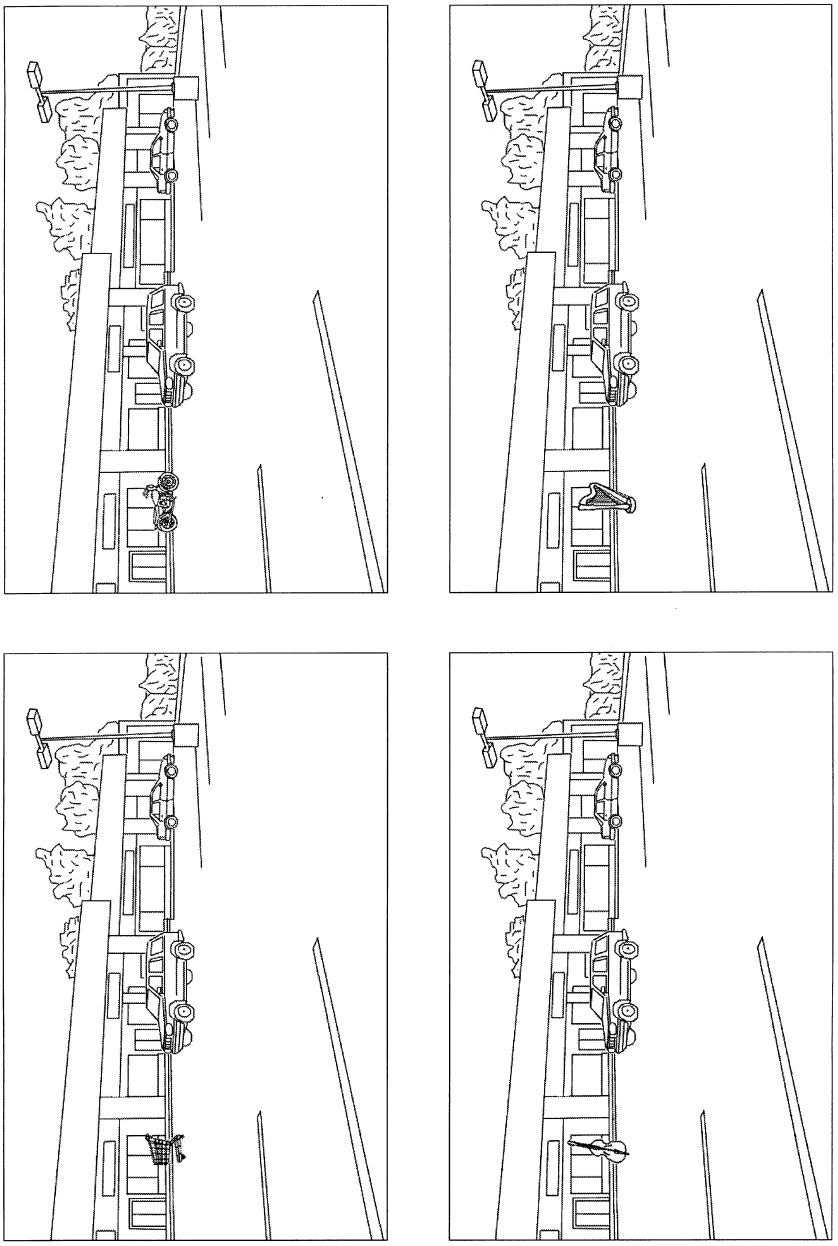

Fig. 1. An example of the type of scene used and the semantic consistency manipulation. The top two scenes contain semantically consistent target objects (shopping cart and motorcycle), whereas the bottom two scenes contain semantically inconsistent target objects (cello and harp). This parking lot scene was paired with a concert hall scene in which the cello and harp were consistent and the shopping cart and motorcycle were inconsistent.

was presented in the concert hall scene, the forced-choice screen presented the labels "cello" and "harp". When an inconsistent target object (a motorcycle or a shopping cart) was presented in the concert hall scene, the forced-choice screen presented the labels "motorcycle" and "shopping cart". The participants' task was to indicate which of the two labels named an object that had been presented in the scene. Fig. 2 depicts the sequence of events in an experimental trial.

Both the criterion modulation model and the description enhancement model predict that percent correct discrimination performance should be better when the target object is consistent versus inconsistent with the scene in which it appears, because they propose that consistent scene context facilitates the identification of objects. The criterion modulation model proposes facilitated identification through the lowering of identification thresholds for consistent objects, whereas the

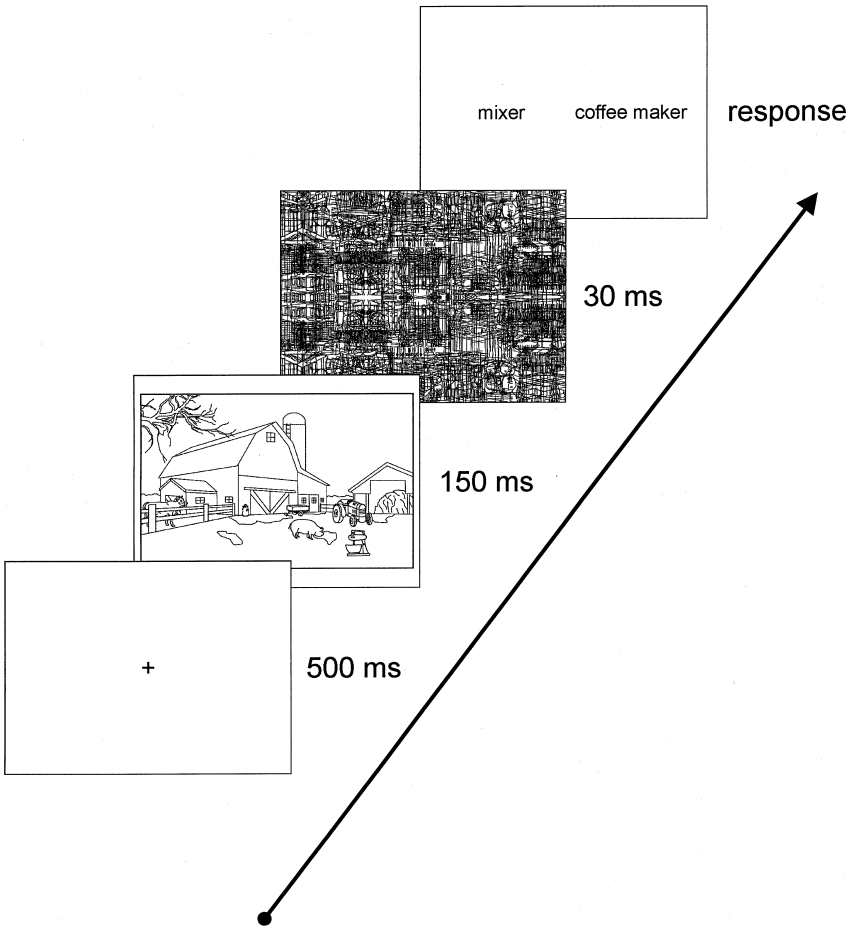

Fig. 2. Schematic illustration of a trial in Experiment 1.

description enhancement model proposes facilitated identification through the enhanced encoding of perceptual information for consistent objects. In contrast, the functional isolation model predicts no advantage for the discrimination of consistent versus inconsistent objects, because it proposes that object identification processes are isolated from semantic knowledge about the real-world contingencies between objects and scenes.

2.1. Method

Participants. Twenty-six Michigan State University undergraduate students participated in the experiment for course credit. All participants had normal or corrected-to-normal vision. The participants were naive with respect to the hypotheses under investigation.

Stimuli. The stimuli were the same as in Hollingworth and Henderson (1998, Experiment 4). Line drawings of 20 scenes were generated from photographs of natural scenes. Fourteen scenes were generated from those used by van Diepen and De Graef (1994), and the other six scenes were generated from photographs taken in the East Lansing, Michigan area. The images generated from the two sources were not distinguishable. The main contours of the scenes were traced using commercial software to create gray-scale line drawings. Two semantically consistent target objects for each scene were also created by digitally tracing scanned images. The 20 scenes were paired, and the semantically inconsistent target conditions were created by swapping objects across scenes. For example, a fire hydrant and a parking meter were the semantically consistent target objects in a street scene, whereas a chair and a television (on a stand) were the semantically consistent target objects in a living room scene. These targets were swapped across scenes so that the fire hydrant and parking meter were the semantically inconsistent targets in the living room scene, and the chair and television were the inconsistent targets in the street scene. All target objects appeared in the same position in each scene, which did not coincide with the experimenter-determined initial fixation position. This position was chosen as a place within the scene where the consistent target objects might reasonably appear. This paired-scene design was employed so that each scene served as a control for its partner, reducing the influence of such factors as object size, eccentricity, and lateral masking. Appendix A lists the scenes and target objects in Experiment 1.

All scene and object manipulations were conducted using commercially available software. The scenes subtended a visual angle of 23° (width) by 15° (height) at a viewing distance of 64 cm. Target objects subtended about 2.73° on average (range = 1.30–5.23°), measured along the longest axis. All images were displayed as gray-scale contours on a white background at a resolution of 800 by 600 pixels by 16 levels of gray. The pattern mask presented after the scene consisted of overlapping line segments, curves, and angles, and was slightly larger than the scene stimuli. The scenes were completely obliterated when presented simultaneously with the pattern mask. Target labels were created using lower-case, 24-point, antialiased Arial font. To create the forced-choice response screen, the centers of the

two labels were presented an equal distance to the left and right of the center of the screen.

Apparatus. The stimuli were displayed on a NEC MultiSync XE15 SVGA monitor with a 100 Hz refresh rate. Responses were collected with a button box connected to a dedicated input–output (I–O) board. Depression of a button stopped a millisecond clock on the I–O board. The display and I–O systems were interfaced with a 486–based microcomputer that controlled the experiment.

Procedure. Participants were tested individually. The experimenter first explained that the task on each trial was to view a briefly displayed scene and to determine which of two labels described an object that had appeared in the scene. Participants were instructed that one of the labels always described an object that had appeared in the scene and the other label always described an object that had not appeared in the scene.

The participant was then seated in front of a computer monitor, with one hand resting on the left-hand button and the other on the right-hand button of the button box. Viewing distance was maintained by a forehead rest. Participants saw a fixation screen (containing a central fixation point at which participants were to direct their gaze) for 500 ms, followed by presentation of the scene for 150 ms, followed by the pattern mask for 30 ms, followed by the forced-choice response screen. There was no delay (i.e., the inter-stimulus interval was zero) between each display. The forced-choice response screen remained in view until the participant pressed the left button to indicate that the object named by the left-hand label had appeared in the scene or the right button to indicate that the object named by the right-hand label had appeared in the scene. After the response, there was a 4 s delay while the stimuli for the next trial were loaded into video memory, and then the prompt for the next trial appeared.

Participants took part in a practice block of 16 trials. The two scenes used in the practice block were not used in the experimental trials. After the practice trials, the experimenter answered any questions the participant had about the procedure, and the participant proceeded to the experimental trials. Each participant then saw 160 experimental trials that were produced by a within-participant factorial combination of 2 target object consistency conditions $\times$ 2 target objects per scene $\times$ 2 label positions in the forced-choice response screen $\times$ 20 scenes. Because the latter two factors were not of theoretical interest, the two levels of those factors were combined in the statistical analysis. Trial order was randomized independently for each participant. The entire session lasted approximately 40 min.

2.2. Results

Percent correct analysis. The influence of object consistency on percent correct discrimination performance was analyzed via a simple effects test. There was a reliable effect of the consistency of the target object, $F(1,25) = 5.61$, $\mathrm{MSE} = 0.0075$, $p < 0.05$. Participants responded correctly 64.2% of the time when the target object was consistent with the scene and 67.0% of the time when the target object was inconsistent with the scene, an inconsistent object advantage of 2.8%.

2.3. Discussion

In Experiment 1, we replicated the forced-choice, type-discrimination paradigm developed by Hollingworth and Henderson (1998). To assure that discrimination performance reflected the initial influence of scene context on object identification, we limited the presentation of the scene to 150 ms. Contrary to the prediction of the criterion modulation and description enhancement models, discrimination performance was higher when the target object was semantically inconsistent rather than consistent with the scene in which it appeared. (We will discuss potential explanations for this inconsistent object advantage in Section 5.) This result supports earlier findings that when response bias is eliminated from object detection paradigms and when differences in search efficiency are controlled, no advantage is obtained for the identification of consistent objects (Hollingworth & Henderson, 1998). Together, these studies suggest that results from earlier experiments that found better detection of consistent versus inconsistent objects (Biederman, Mezzanotte & Rabinowitz, 1982; Boyce, Pollatsek & Rayner, 1989) may not have reflected the influence of scene context on object identification. Instead, the consistent object advantage in those experiments appears to have been caused by the inadequate control of response bias and by differences in search efficiency for consistent versus inconsistent objects.

3. Experiment 2

The central claim of the description enhancement model is that the activation of a scene schema interacts with the initial perceptual analysis of objects in the scene, facilitating the perceptual analysis of objects consistent with the scene. As a result, the constructed visual description of a consistent object will be more detailed compared to that of an inconsistent object, leading to facilitated identification. Previous experiments that have failed to find an advantage for the identification of consistent versus inconsistent objects (Experiment 1; Hollingworth & Henderson, 1998) have used paradigms requiring participants to detect a particular object type or to discriminate between two object types. In addition, these experiments have identified target objects using labels, abstracted from the visual form of the objects presented in the scenes. Thus, the specific hypothesis that the constructed visual descriptions of consistent objects will be more detailed than those of inconsistent objects has not been tested directly.

To test this hypothesis, we employed a forced-choice, token-discrimination paradigm. The basic paradigm was the same as Experiment 1 (see also, Masson, 1991), but instead of discriminating between two object types (e.g., whether a chair or a television appeared in the scene), participants were asked to discriminate between two object tokens of the same type (e.g., which of two different chairs appeared in the scene). Given that discriminating between two tokens should be more difficult than between two object types, we presented the scene for 250 ms rather than 150. Fig. 3 illustrates the paradigm. A scene containing one of two object tokens was presented

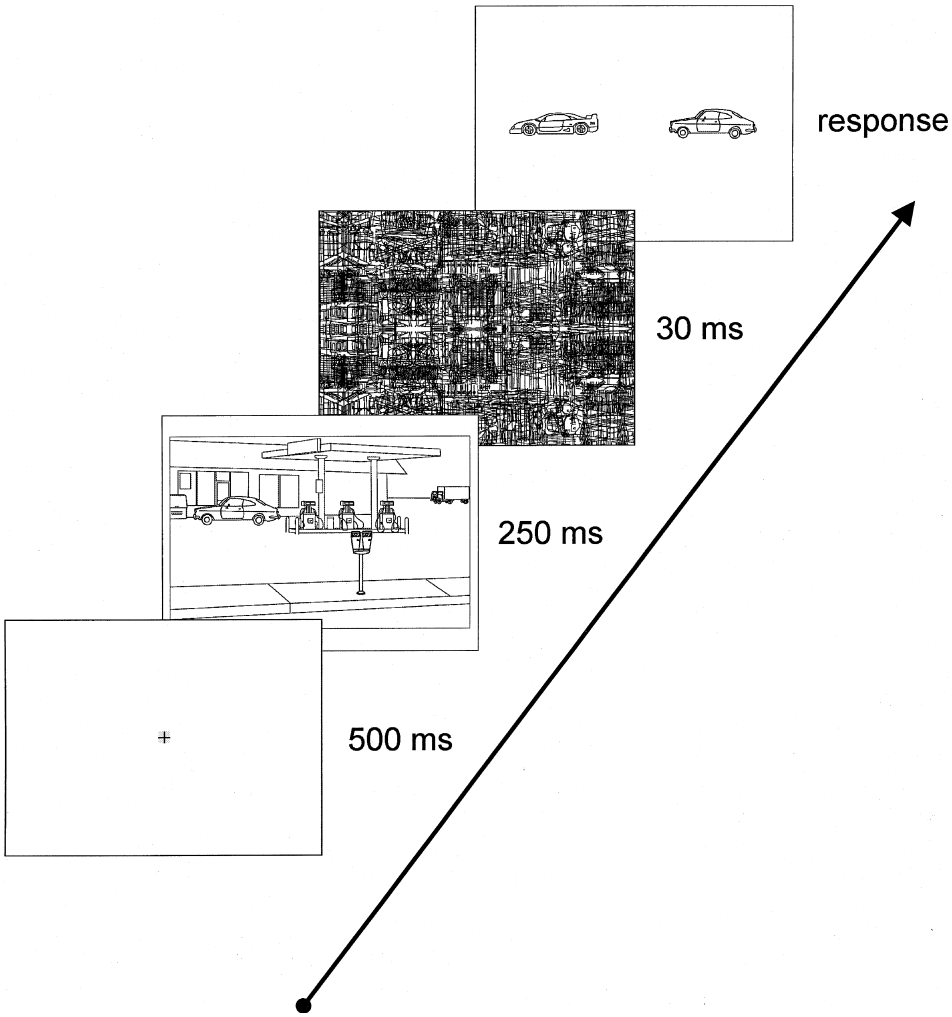

Fig. 3. Schematic illustration of a trial in Experiments 2 and 3.

for 250 ms, followed by a pattern mask for 30 ms, followed by a screen displaying the two token alternatives. One token had been presented in the scene and the other had not. The participants' task was to indicate which of the two tokens had been presented in the scene. The semantic consistency between the scene and the token presented in the scene was manipulated. Each token was equally likely to appear in each consistency condition.

In this experiment, we did not distinguish between token differences at the level of subordinate categorization (e.g., a sedan versus a sports car) and token differences between visually distinct exemplars of the same basic or subordinate

category (e.g., two visually different chairs). In either case, conceptual-level differences in representation should be minimized, and discrimination performance should be based primarily on the amount of perceptual information encoded from the object token presented in the scene. Fig. 4 presents an example of the token manipulation. Because the description enhancement model proposes that the initial perceptual analysis for consistent objects is facilitated compared to inconsistent objects, this model predicts that discrimination performance will be higher when the presented token is consistent versus inconsistent with the scene in which it appears. In contrast, the functional isolation model predicts no advantage for consistent tokens.

3.1. Method

Participants. Twenty-four Michigan State University undergraduate students participated in the experiment for course credit. All participants had normal or corrected-to-normal vision. The participants were naive with respect to the hypotheses under investigation. None had participated in Experiment 1.

Stimuli. The stimuli were the same as in Experiment 1 with the following modifications. For each scene, one consistent target object was chosen and a second token of the same object type was created. In three scenes, two new tokens were created due to difficulty finding an appropriate second token for an existing target object. Object tokens subtended about 3.18° on average (range = 1.25–7.2°), measured along the longest axis. As in Experiment 1, the semantically inconsistent condition was created by pairing scenes and swapping object tokens between them. In the forced-choice response screen, the pictures of the two token alternatives (presented in exactly the same form as they appeared in the scenes) were centered vertically and positioned to the left and right of fixation. Appendix B lists the scenes and target objects in Experiment 2.

Apparatus and procedure. The apparatus was the same as in Experiment 1. Participants saw a scene for 250 ms, followed by a pattern mask for 30 ms, followed by a forced-choice response screen containing two object token alternatives. There was no delay (i.e., the inter-stimulus interval was zero) between each display. The forced-choice response screen remained in view until the participant pressed either the left button to indicate that the left-hand token had appeared or the right button to indicate that the right-hand token had appeared in the scene.

Each participant took part in a practice block of 16 trials. The two scenes used in the practice block were not used in the experimental trials. Each participant then saw 160 experimental trials produced by a within-participant factorial combination of 2 token consistency conditions × 2 tokens × 2 positions in forced-choice response screen × 20 scenes. To assess practice effects, the trials were distributed into two blocks. The position of the tokens in the forced-choice response screen display was partially counterbalanced within a block and fully counterbalanced between blocks. Block order was counterbalanced between participant groups. Because the token factor and the position in the forced-choice response screen factor were not of theoretical interest, the two levels of each factor were combined in the statistical

Fig. 4. An example of the token manipulation.

analyses. Trial order within block was randomized independently for each participant. The entire session lasted approximately 40 min.

3.2. Results

Percent correct analysis. First, there was a main effect of block, with better performance in the second block (68.2%) than in the first block (63.8%), $F(1,23) = 6.65$, $\mathrm{MSE} = 0.0142$, $p < 0.05$. Because block did not interact with semantic consistency, $F < 1$, the subsequent analysis collapsed across the blocking factor. There was no effect of the consistency of the target object, $F < 1$. Participants responded correctly 65.3% of the time when the target object was consistent with the scene and 66.8% of the time when the target object was inconsistent with the scene. The 95% confidence interval around these means was ±2.82%. Thus, the experiment had enough power to detect a 3.99% effect (see Loftus & Masson, 1994).

3.3. Discussion

Experiment 2 employed a forced-choice, token-discrimination paradigm to test the prediction of the description enhancement model that the encoding of perceptual information for an object stimulus will be facilitated when that object is consistent with scene context, leading to better token discrimination in the consistent versus inconsistent condition. Contrary to that prediction, but in accord with the prediction of the functional isolation model, no advantage was found for the discrimination of consistent object tokens. The non-reliable trend was in the direction of better token discrimination when the presented token was semantically inconsistent with the scene. One potential concern with this experiment, however, is that performance was not particularly high. Participants responded correctly 66% of the time in a paradigm in which chance is 50%. Thus, in Experiment 3, we modified the Experiment 2 stimuli to improve overall performance.

4. Experiment 3

In Experiment 3, the position in which the object tokens were placed was moved closer to fixation in a number of scenes. In addition, the object tokens were made slightly larger in a number of other scenes. These modifications were intended to improve participants' ability to discriminate between the two token alternatives. As in Experiment 2, the description enhancement model predicts a consistent token discrimination advantage, whereas the functional isolation model predicts that no such advantage should be obtained.

4.1. Method

Participants. Forty Michigan State University undergraduate students participated in the experiment for course credit. All participants had normal or

corrected-to-normal vision. The participants were naive with respect to the hypotheses under investigation. None had participated in Experiments 1 or 2.

Stimuli. The stimuli were the same as in Experiment 2 with the following modifications. To raise performance, object tokens were moved to a position closer to fixation in a number of scenes. In other scenes, the object tokens were enlarged slightly (though not so much that the size of the resulting object would appear at all incongruous). In addition, four more scene stimuli were added to the original 20. Object tokens subtended about 3.09° on average (range = 1.40–7.20°), measured along the longest axis. [4] Appendix C lists the four new scenes employed in Experiment 3.

Apparatus and procedure. The apparatus was the same as in Experiments 1 and 2. Each participant took part in a practice block of 16 trials. The two scenes used in the practice block were not used in the experimental trials. Each participant then saw 192 experimental trials produced by a within-participant factorial combination of 2 token consistency conditions × 2 tokens × 2 positions in forced-choice response screen × 24 scenes. The stimuli were distributed into 8 blocks in order to assess practice effects. Each block contained all 24 scenes, and within each block 3 scenes were presented in each of the 8 conditions. Block order was counterbalanced between participant groups. Because the token factor and the position in the forced-choice response screen factor were not of theoretical interest, the two levels of each factor were combined in the statistical analyses. Trial order within block was randomized independently for each participant. The entire session lasted approximately 45 min.

4.2. Results

Percent correct analysis. First, there was a main effect of block, with better performance in later blocks than in earlier blocks, $F(7,39) = 10.24$, $\mathrm{MSE} = 0.0625$, $p < 0.001$. Second, there was a non-reliable trend in the direction of better performance when the presented object was inconsistent versus consistent with the scene. Participants responded correctly 72.8% of the time when the presented object was inconsistent with the scene and 71.5% of the time when the presented object was consistent with the scene, $F(1,39) = 2.60$, $\mathrm{MSE} = 0.0433$, $p = 0.11$. The 95% confidence interval for the means in this contrast was ±1.17%. Thus, the experiment had enough power to detect a 1.65% effect.

Although block did not interact with semantic consistency, $F < 1$, examination of these data suggested that an advantage for inconsistent token discrimination may have been present in the early blocks. Because the division of the experiment into 8 blocks resulted in a relatively large amount of variability in each block, the initial test

[4] The average size of objects in Experiment 3 was actually smaller than in Experiment 2. This was due to the fact that the target objects in the four new scenes in Experiment 3 were smaller than average. These objects, however, were placed relatively close to fixation, consistent with our goal of improving performance in Experiment 3.

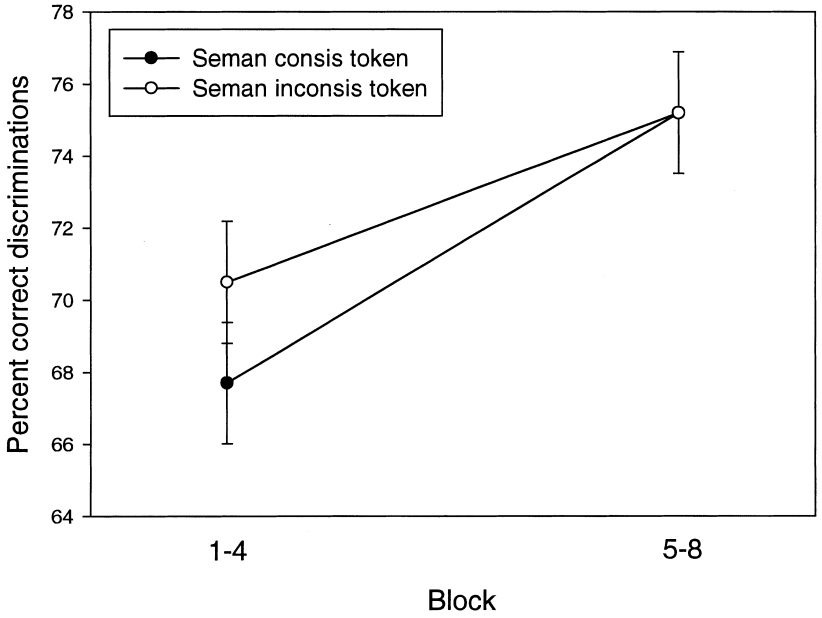

Fig. 5. Percent correct token discrimination performance as a function of semantic consistency and block in Experiment 3. Error bars are 95% confidence intervals based on the error term from the interaction between semantic consistency and first versus second half of the trials.

of the interaction between block and semantic consistency may not have had enough power. Thus, we conducted a post hoc test of discrimination performance in the first half of the trials (first four blocks) versus the second half of the trials (second four blocks) as a function of semantic consistency. These data are displayed in Fig. 5. There was a marginally reliable interaction between half of the trials and semantic consistency, $F(1,39) = 3.03$, $\mathrm{MSE} = 0.0467$, $p = 0.09$, with a trend toward a larger inconsistent discrimination advantage in the first half of the experiment than in the second.

4.3. Discussion

In Experiment 3, the stimuli were modified to improve discrimination performance. Performance did improve: Discrimination performance was 6.1% higher in Experiment 3 than in Experiment 2. Contrary to the prediction of the description enhancement model, however, there was no evidence of a consistent token discrimination advantage. As in Experiment 2, the non-reliable trend was in the direction of better inconsistent token discrimination. These results in conjunction with those from Experiment 1 provide strong evidence against the description enhancement model. Instead, these data support the conclusion that object identification is functionally isolated from knowledge about the real-world contingencies between objects and scenes.

One potential explanation of these results can be drawn from Friedman's (Friedman, 1979) criterion modulation model (see also Masson, 1991). According to that model, top-down influence from an activated scene schema mediates the amount of perceptual information necessary to reach identification threshold. Objects that are consistent with a scene can be identified fairly "automatically" through the encoding of relatively few (perhaps global) object features. The identification of inconsistent objects, however, requires a resource-intensive analysis of a relatively greater number of object features. Assuming that the encoding of perceptual information ceases upon reaching an identification threshold, the visual description of an inconsistent object will be more detailed than that of a consistent object. Although this model appears to be able to account for the trends observed in Experiments 2 and 3 toward better inconsistent token discrimination, the hypothesis regarding the visual descriptions of consistent versus inconsistent objects cannot be separated from the hypothesis regarding differences in the ease of identification. The model proposes that the perceptual representation of inconsistent objects will be more detailed as a direct result of the fact that consistent objects are identified more easily at the level of object type. In other words, it should be more difficult to determine which specific chair appeared in a scene in just that case when it is easier to determine that a chair in fact appeared (i.e., when the chair is in a consistent context). However, Experiment 1 (and Hollingworth & Henderson, 1998) found no evidence for the facilitated identification of semantically consistent objects at the level of object type. Thus, the Friedman model does not appear able to account for the full pattern of data found in this study.

How can we be confident that the absence of a consistent object advantage in Experiments 2 and 3 did not arise from lack of power? First, the semantic consistency manipulation in these experiments was quite strong. In previous experiments using similar stimuli (Hollingworth & Henderson, 1998), participants demonstrated robust response biases as a function of the semantic relationship between presented object and scene. Thus, the semantic information needed to constrain object identification processing as a function of semantic consistency should have been available. Second, Experiments 2 and 3 had a good deal of statistical power. In Experiment 3, we could have detected a 1.65% effect of semantic consistency. Not only was such an effect absent, but the trends in the consistency effect were in the direction of better inconsistent token discrimination. Thus, we feel safe concluding that consistent scene context does not facilitate object identification, and in particular, that consistent scene context does not facilitate the initial perceptual analysis of objects.

5. General discussion

This study investigated whether stored knowledge about scene types and the objects that form those types influences the identification of objects appearing in scenes. Specifically, we tested two views of the potential interaction between scene

context and object identification. First, the description enhancement model proposes that the early activation of a scene schema facilitates the initial perceptual analysis of objects consistent with the scene, leading to better identification of consistent versus inconsistent objects. Second, the criterion modulation model proposes that scene knowledge interacts with the matching of a constructed visual description to stored descriptions, so that less perceptual information is required for consistent object representations to reach identification threshold activation, again leading to facilitated identification of objects consistent with the scene. These models receive support from earlier object detection studies that found facilitated detection of semantically consistent versus inconsistent objects appearing in briefly presented scenes (Biederman, Mezzanotte & Rabinowitz, 1982; Boyce, Pollatsek & Rayner, 1989). However, Hollingworth and Henderson (1998) have demonstrated that when methodological problems with these paradigms are corrected, no consistent object advantage is obtained. The consistent object advantage in previous object detection experiments (Biederman, Mezzanotte & Rabinowitz, 1982; Boyce, Pollatsek & Rayner, 1989) appears to have been caused by inadequate control of response bias and by differences in search efficiency as a function of object consistency.

The goal of the present study was to extend the experiments in Hollingworth and Henderson (1998) and to test further the criterion modulation and description enhancement models of object identification in scenes. In Experiment 1, we replicated the forced-choice, type-discrimination paradigm of Hollingworth and Henderson (in press-a, Experiment 4), but reduced the duration of the scene presentation of assure that discrimination performance reflected the initial influence of scene context on object perception. Contrary to the prediction of the criterion modulation and description enhancement models, discrimination performance was better for semantically inconsistent compared to consistent objects.

In Experiments 2 and 3, we sought to test the hypothesis derived from the description enhancement model that the constructed visual description of a semantically consistent object will be more detailed than that of an inconsistent object. We employed a forced-choice, token-discrimination paradigm: Participants saw a briefly presented scene, followed by a pattern mask, followed by a forced-choice response screen displaying two tokens of an object type, only one of which had been presented in the scene. Contrary to the prediction of the description enhancement model, token discrimination performance was no better for consistent versus inconsistent object tokens. In both experiments the non-reliable trend was in the direction of better inconsistent token discrimination.

These results do not provide support for the hypothesis that object identification is influenced by the constraints imposed by scene meaning. Instead, the results suggest that both the initial perceptual analysis of objects and the matching of constructed visual descriptions to stored descriptions are functionally isolated from stored knowledge about the real-world contingencies between objects and the scenes in which they appear. Such isolation may derive from structural properties of the functional architecture of the visual system (Hollingworth & Henderson, 1998; see Fodor, 1983; Pylyshyn, 1980, Pylyshyn, in press). Specifically, visual architecture

may isolate object identification from knowledge of scene–object contingencies, because without such isolation, the potentially relevant information to any scene perception task would be so large as to make the discrimination between relevant and irrelevant information resource intensive and time consuming. Therefore, the adaptive value of quickly recognizing objects in the environment may have produced an object identification system that consults only a very limited set of information (Fodor, 1983).

This point raises the question of exactly what types of information are consulted during object identification. We propose that both the construction of a visual description of an object and the matching of constructed visual descriptions to stored descriptions occurs presemantically. Thus, an object is identified as a perceptual type within a system devoted to the analysis of object form, independently of object meaning and other associative information such as an object's semantic relation to the scene in which it appears. This *perceptual identification system* would represent information about visual features and the routines necessary to construct a visual object description from these features. In addition, it would store visual descriptions of real-world object types, against which constructed descriptions are matched, leading to entry-level recognition. This view is consistent with recent theories of object recognition that propose no role for the influence of contextual constraint on the identification of objects (Biederman, 1987; Bülthoff, Edelman, & Tarr, 1995; see also Marr & Nishihara, 1978). [5] In fact, much of the current literature on object identification has employed novel objects to discriminate between competing theories (e.g., Biederman & Gerhardstein, 1993; Tarr, Bülthoff, Zabinski, & Blanz, 1997) upon the assumption that object form is solely functional in the identification of an object as a particular type. In addition, this view is consistent with evidence from implicit memory research suggesting separate memory systems for the representation of object form on the one hand and the representation of semantic and associative information about objects on the other (e.g., Schacter, Cooper, & Delaney, 1990). Finally, this view is consistent with neuropsychological evidence suggesting that there are dissociable systems responsible for perceptual classification versus semantic classification (e.g., Riddoch & Humphreys, 1987).

One final issue raised by this study concerns the reliable advantage for the discrimination of inconsistent objects found in Experiment 1. We have now obtained an inconsistent object advantage in three different paradigms designed to investigate object perception in scenes: a type-discrimination paradigm (Experiment 1), an object detection paradigm similar to that of Biederman, Mezzanotte & Rabinowitz (1982) (Hollingworth & Henderson, 1998, Experiment 3), and a change detection

[5] In contrast, interactive theories, such as the description enhancement and criterion modulation models, were developed when it was generally believed that information encoded from the retina was insufficient to support object identification (see Bruner, 1973; Neisser, 1967). Thus, the application of real-world knowledge to perceptual tasks was deemed necessary to resolve ambiguity inherent in the visual input.

paradigm in which a target object is changed between two subsequent presentations of a scene (Hollingworth & Henderson, in press). Hollingworth and Henderson (1998) identified two potential hypotheses to explain this effect, both of which are compatible with the view that object identification is functionally isolated from stored scene knowledge. First, according to a *memory schema hypothesis*, perceptual analysis of semantically consistent and inconsistent objects proceeds equivalently, but information about semantically inconsistent objects is preferentially remembered, perhaps as part of a list noting deviations from the default values in the schema (Friedman, 1979). This hypothesis is supported by scene memory studies that have shown better long-term memory for semantically inconsistent versus consistent objects (e.g., Friedman, 1979). Second, an *attention hypothesis* proposes that the perception of semantically consistent and inconsistent objects is not influenced directly by scene knowledge, but attention is preferentially allocated to already-identified objects that violate the constraints imposed by scene meaning. The additional attentional resources devoted to an inconsistent object would then produce a more complete visual description of that object, leading to better detection performance. Regardless of which (if either) explanation is correct, the results from this study do not support the general hypothesis that consistent scene context facilitates object identification, but instead support the view that object identification is isolated from stored knowledge about real-world contingencies between objects and scenes.

Acknowledgements

This research was supported by a National Science Foundation Graduate Fellowship to Andrew Hollingworth and grants from the U.S. Army Research Office (DAAH04-94-G-0404) and the National Science Foundation (SBR 96-17274) to John M. Henderson. The contents of this article are those of the authors and should not be construed as an official Department of the Army position, policy, or decision. We would like to thank Johan Wagemans, Peter De Graef, and Sandy Pollatsek for their helpful comments on an earlier version of the manuscript and Jennifer Johnson for help in running the experimental sessions.

Appendix A

Semantically consistent and inconsistent target objects for each scene in Experiment 1. The labels appearing in the forced-choice screen were presented as they appear below.

Scene	Consistent target objects	Inconsistent target objects
Bar	Wine bottle, cocktail	Boots, teddy bear
Bedroom	Boots, teddy bear	Wine bottle, cocktail
Beach	Snorkeling mask, flippers	Iron, hanger

Launderette	Iron, hanger	Snorkeling mask, flippers
Classroom	Globe, backpack	Wheelbarrow, lawnmower
Front yard	Wheelbarrow, lawnmower	Globe, backpack
Dining room	Lamp, candle	Tricycle, skateboard
Playground	Tricycle, skateboard	Lamp, candle
Farmyard	Chicken, pig	Mixer, coffee maker
Kitchen	Mixer, coffee maker	Chicken, pig
Gas station	Tow truck, bicycle	Boat, dock
Pond	Boat, dock	Tow truck, bicycle
Living room	Chair, television	Fire hydrant, parking meter
Street	Fire hydrant, parking meter	Chair, television
Locker room	Barbell, tennis shoes	Flower vase, water pitcher
Restaurant	Flower vase, water pitcher	Barbell, tennis shoes
Parking lot	Shopping cart, motorcycle	Cello, harp
Concert hall	Cello, harp	Shopping cart, motorcycle
Patio	Barbecue grill, dog	Coat rack, computer
Library	Coat rack, computer	Barbecue grill, dog

Appendix B

Semantically consistent and inconsistent target objects for each scene in Experiment 2. Objects of equivalent semantic consistency were two different tokens of an object type. Object tokens that were visually different but did not clearly differ at a subcategory level are listed as A and B. Objects marked with an asterisk were used as targets in Experiment 1

Scene	Consistent target objects	Inconsistent target objects
Bar	Martini, highball*	Men's shoes, women's pumps
Bedroom	Men's shoes, women's pumps	Martini, highball*
Beach	Snorkeling mask A*, B	Iron A*, B
Launderette	Iron A*, B	Snorkeling mask A*, B
Classroom	Globe A*, B	Wheelbarrow A*, B
Front yard	Wheelbarrow A*, B	Globe A*, B
Dining room	Lamp A*, B	Tricycle A*, B
Playground	Tricycle A*, B	Lamp A*, B
Farmyard	Rooster*, hen	Standing mixer*, hand mixer
Kitchen	Standing mixer*, hand mixer	Rooster*, hen
Gas station	Sports car, sedan*	Row boat*, speed boat
Dock	Row boat*, speed boat	Sports car, sedan*
Living room	Chair A*, B	Fire hydrant A*, B
Street	Fire hydrant A*, B	Chair A*, B

Locker room	Tennis racket, badminton racket	Water pitcher A*, B
Restaurant	Water pitcher A*, B	Tennis racket, badminton racket
Parking lot	Motorcycle A*, B	Grand piano, upright piano
Concert hall	Grand piano, upright piano	Motorcycle A*, B
Patio	Charcoal grill*, gas grill	Coat rack A*. B
Library	Coat rack A*. B	Charcoal grill*, gas grill

Appendix C

New scenes added for Experiment 3. Objects of equivalent semantic consistency were two different tokens of an object type. Object tokens that did not clearly differ at a subcategory level are listed as A and B. Objects marked with an asterisk were used as targets in Experiments 1 and 2.

Scene	Consistent target objects	Inconsistent target objects
Cemetery (paired with beach)	Flower bouquet A, B	Snorkeling mask A*, B
Laboratory (paired with launderette)	Microscope A, B	Iron A*, B
Office (paired with bathroom)	Rotary phone, cordless phone	Hair dryer A*, B
Bathroom (paired with office)	Hair dryer A, B	Rotary phone, cordless phone

References

Bar, M., & Ullman, S. (1996). Spatial context in recognition. *Perception*, *25*, 343–352.

Biederman, I. (1981). On the semantics of a glance at a scene. In M. Kubovy, & J. R. Pomerantz, *Perceptual organization* (pp. 213–253). Hillsdale, NJ: Erlbaum.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*, 115–147.

Biederman, I., & Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: Evidence and conditions for 3-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 1162–1182.

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

Biederman, I., Teitelbaum, R. C., & Mezzanotte, R. J. (1983). Scene perception: A failure to find a benefit from prior expectancy of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*, 411–429.

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.

Bruner, J. S. (1973). *Beyond the information given*. New York: W. W. Norton.

Bülthoff, H. H., Edelman, S. Y., & Tarr, M. J. (1995). How are three-dimensional objects represented in the brain? *Cerebral Cortex*, *3*, 247–260.

De Graef, P. (1992). Scene-context effects and models of real-world perception. In K. Rayner, *Eye Movements and Visual Cognition: Scene Perception and Reading* (pp. 243–259). New York: Springer.

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

Fodor, J. A. (1983). *Modularity of Mind*. Cambridge, MA: MIT Press.

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

Friedman, A., & Liebelt, L. S. (1981). On the time course of viewing pictures with a view towards remembering. In D. F. Fisher, R. A. Monty, & J. W. Senders, *Eye Movements: Cognition and Visual Perception* (pp. 137–155). Hillsdale, NJ: Erlbaum.

Henderson, J. M. (1992). Object identification in context: The visual processing of natural scenes. *Canadian Journal of Psychology, 46* (*Special Issue*), 319–341.

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

Henderson, J. M., Weeks, P. A., Jr., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex 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. (in press). Semantic informativeness mediates the detection of changes in natural scenes. *Visual Cognition: Special Issue on Change Detection and Visual Memory*.

Kosslyn, S.M. (1994). *Image and brain.* Cambridge, MA: MIT Press.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subjects designs. *Psychonomic Bulletin & Review*, *1*, 476–490.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London B*, *200*, 269–294.

Masson, M. E. J. (1991). Constraints on the interaction between context and stimulus information. In K. J. Hammond & D. Gentner, *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society* (pp. 540–545). Hillsdale, NJ: Erlbaum.

Neisser, U. (1967). *Cognitive Psychology,* Englewood Cliffs, NJ: Prentice-Hall.

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

Pylyshyn, Z. (1980). Computation and cognition: Issues in the foundations of cognitive science. *Behavioral and Brain Sciences*, *3*, 111–132.

Pylyshyn, Z. (in press). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioural and Brain Sciences*.

Rayner, K., & Pollatsek, A. (1992). Eye movements and scene perception. *Canadian Journal of Psychology, 46 (Special Issue),* 342–376.

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, *81*, 275–280.

Riddoch, M. J., & Humphreys, F. W. (1987). Visual object processing in optic aphasia: A case of kinantic aphasia. *Cognitive Neuropsychology*, *4*, 131–185.

Schacter, D. L., Cooper, L. A., & Delaney, S. M. (1990). Implicit memory for unfamiliar objects depends on access to structural descriptions. *Journal of Experimental Psychology: General*, *119*, 5–24.

Tarr, M. J., Bülthoff, H. H., Zabinski, M., & Blanz, V. (1997). To what extent do unique parts influence recognition across changes in viewpoint? *Psychological Science*, *8*, 282–289.

Ullman, S. (1996). *High-level vision: Object recognition and visual cognition.* Cambridge, MA: MIT Press.

van Diepen, P. M. J., & De Graef, P. (1994). *Line-drawing library and software toolbox*. (Psych. Rep. No. 165). Leuven, Belgium: University of Leuven, Laboratory of Experimental Psychology.