Context in Object and Scene Perception

by

Jodi L. Davenport

B.S., Cognitive Science
University of California, Los Angeles, 1998

Submitted to the Department of Brain and Cognitive Sciences in Partial
Fulfillment of the Requirements for the Degree of
Doctor of Philosophy In Cognitive Science

at the
Massachusetts Institute of Technology
June 2005

Signature of Author:____

Department of Brain and Cognitive Sciences
June 2, 2005

Certified by:____

Mary C. Potter
Professor of Psychology
Thesis Supervisor

Accepted by:____

Earl K. Miller
Picower Professor of Neuroscience
Chairman, Department Graduate Committee

Context in Object and Scene Perception

by

Jodi L. Davenport

Submitted to the Department of Brain and Cognitive Sciences in
Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy in Cognitive Science

Abstract

In the world, objects and settings tend to co-occur. Cars usually appear on streets with other vehicles, not in kitchens next to refrigerators. The present studies provide evidence that the semantic consistency of an object and its setting is available in a glimpse and affects perception. Objects are perceived more accurately in typical rather than atypical settings and when they appear with related objects regardless of the setting. Backgrounds are perceived more accurately when they contain plausible rather than unlikely foreground objects. Objects and scenes are processed interactively, not in isolation.

Table of Contents

Acknowledgements

First and foremost, I would like to thank my advisor, Molly Potter. I feel so fortunate to have had the opportunity to work with such a wonderful mentor. Molly is a role model of academic excellence and I'll always be in awe of her ability to rapidly understand the core of any topic, talk or experimental design. She has challenged me to conduct rigorous research and to constantly question my assumptions. I sincerely appreciate her advice, support and encouragement over the past five years, and I look forward to being her colleague in the future.

I would also like to thank my thesis committee, Nancy Kanwisher, Aude Oliva and Marvin Chun for their helpful comments and suggestions and their continued involvement in my research.

Finally, I would like to thank my friends and family who made graduate school such a positive chapter in my life.

Chapter 1: Introduction

The gist, or meaning, of a scene can be understood with a presentation as short as 100 ms (Biederman, 1972; Biederman, Rabinowitz, Glass & Stacy, 1974; Potter, 1975, 1976). How is such rapid understanding accomplished? Does it depend on perception of one or more objects in the scene, on global perception of background properties, or on some combination of these?

The current series of experiments investigate the influence of context on perception. In the world, objects and backgrounds tend to co-occur. Cars usually appear in streets, not in kitchens. Does knowledge of these co-occurrences influence the initial perception of a scene? If so, scenes containing an object likely to appear in a setting (e.g., a pig on a farm) should be perceived more readily than scenes containing an implausible object (e.g., an Eskimo on a farm). Prior research investigating eye movements, detection performance and two alternative forced choice (2AFC) recognition has focused on how a rapidly extracted scene schema, or information about the setting, influences the perception of objects in a picture. The current research extends the study of consistency effects to how an object influences the perception of its setting, and how objects affect the perception of other objects in a scene.

Early studies used eye movement data to investigate the effects of scene schemata on object perception (Friedman, 1979; Loftus & Mackworth, 1978; Henderson, Weeks & Hollingworth, 1999). The underlying assumption of this paradigm was that the duration and order of fixations on objects would reflect the processing demands of perceiving objects. Specifically, longer fixation times would imply more difficulty in processing, and earlier fixations would be made to regions of "semantic informativeness" (e.g., regions likely to provide meaningful information about a picture). Participants were asked to prepare for a long term memory task as they inspected line drawings of scenes (e.g., a farm). The drawings either contained objects likely to appear in the setting (e.g., a tractor) or objects that were not plausible (e.g., an octopus). Fixations to the unlikely objects were longer than fixations to plausible objects in all three studies, suggesting that more time was required to process objects that were inconsistent with the context of their settings.

While all studies found longer fixation durations for implausible objects, the question of when the consistency effect emerged was inconclusive. Loftus and Mackworth (1978) found that observers fixated on low probability objects earlier, more often, and with longer durations. They concluded that the rapid extraction of gist and partial processing of objects in the periphery directed initial eye movements towards the low probability objects. However, Henderson, Weeks and Hollingworth (1999) failed to find that inconsistent objects were fixated earlier, though they were fixated more often and for longer.

A potential problem with the task of preparing for a long term memory test is that longer fixations may have reflected the additional time needed to integrate an implausible object with the scene schema in memory as opposed to time for initial perception. De Graef, Christiaens and d'Ydewalle (1990) had participants perform a non-object search task to determine whether inconsistent objects would require longer fixations even when no memory encoding was required. In the non-object search task, zero to three "non-objects" (meaningless object-like figures) were placed in a scene (e.g., a playground) in addition to real objects that were consistent (e.g., a slide) or inconsistent (e.g., a tea cup) with the setting. Participants were instructed to look around the scene and count occurrences of non-objects. Again, the fixations for consistent objects were shorter than for inconsistent objects, but this difference only occurred after several fixations were made in the scene. Attention was not immediately drawn to inconsistent objects, and De Graef et al. raise the possibility that the context effects may have been driven by inter-object priming

Even when objects need not be remembered, it is unclear what types of processing are reflected in longer fixation durations for the inconsistent objects. Since fixations generally last about 300 ms and objects may be identified in less time than that (Thorpe, Fize & Marlot, 1996), it is unlikely that fixation duration directly reflects the amount of time required to identify an object. Longer fixations may either reflect processing to identify an object, time to integrate it with the scene, or interest in that region. Overall, due to long scene exposures and task specific effects, it is difficult to use the results of eye movement data to determine whether the initial perception of a scene is influenced by the consistency between objects and their settings.

The influence of semantic consistency of a background on object perception has also been studied using brief presentations of scenes. Many researchers have used the object detection paradigm first introduced by Biederman, Mezzanotte and Rabinowitz (1982). In the object detection paradigm, participants are shown the name of an object followed by a masked 150 ms presentation of a scene. Their task is to determine whether the named object was present in the scene. Objects in consistent settings were detected more accurately than objects in inconsistent settings, leading Biederman et al. to conclude that scene information must have been available early enough to influence object processing.

Boyce and her colleagues found further evidence that object perception is affected by the semantic consistency of a scene (Boyce, Pollatsek & Rayner, 1989; Boyce & Pollatsek, 1992). Using the object detection paradigm, Boyce et al. investigated whether object identification was mediated by a consistent scene, or simply by the presence of consistent objects within that scene. By dissociating inter-object consistency and scene consistency she found that objects were detected more easily in consistent backgrounds, regardless of whether the other objects were consistent or inconsistent with the background. In a naming study, Boyce and Pollatsek (1992) found that the naming latency for an object that "wiggled" (moved slightly to provide a visual transient) was shorter for a consistent than for an inconsistent object.

Not all research involving brief presentations of a scene has found consistency effects. Hollingworth and Henderson (1998; 1999) propose that the reported effects of semantic consistency on perception using the object detection paradigm were due to response bias. In the object detection studies, the task was to respond yes or no whether a target label (e.g., a sofa) matched an object in a briefly presented scene (e.g., a sofa in a living room). Detection sensitivity, *d'*, was higher when an object appeared in a consistent rather than inconsistent setting. However, the *d'* measure did not take into account response bias from the target label on catch trials in which the target object was not present. In the original study (Biederman et al., 1982) false alarm rates were higher when the catch trial target label (e.g., sofa) and pictured scene were consistent (e.g., a living room) rather than inconsistent (e.g., a street), yet the false alarm rate for all catch trials was averaged together. This led to an overestimation of sensitivity in consistent conditions and an underestimation of sensitivity in inconsistent conditions. Hollingworth and Henderson (1998, Exp. 1) replicated this effect when they calculated *d'* in this way, but not when they calculated *d'* with separated false alarm rates for the two catch trial conditions.

In further studies, Hollingworth and Henderson (1998, 1999) failed to find consistency effects using a two-alternative forced choice (2AFC) test. No object label was presented beforehand. Participants reported which object had appeared by choosing between two names or two pictures after viewing a briefly presented picture. To ensure the choice was not made purely on the basis of the category of the setting, the two test options were either both consistent or both inconsistent with the setting. As they found no consistency effects were present in any of their manipulations, Hollingworth and Henderson proposed a functional isolation account of object processing; suggesting that objects and their setting are processed independently from one another.

Although Hollingworth and Henderson did not find consistency effects using the 2AFC task, some aspects of their experiments limit the generality of their findings. Inconsistent objects had an asymmetrical advantage in their 1998 and 1999 studies: Whenever an inconsistent object appeared in a scene, it was always the item that would be tested. Consistent objects had no such distinction. Additionally, critical objects and scenes were repeated multiple times, in both consistent and inconsistent conditions, which might have attenuated the consistency effect.

The current series of experiments address the influence of expectations about which objects and settings tend to co-occur on the initial perception of briefly presented scenes. The next three chapters describe studies using new materials, presenting a new task, and asking new questions to shed light on these issues. Most previous work investigating consistency effects has been limited to black and white line drawings of wide angle views of scenes. In our studies, we used naturalistic color photographs because color has been shown to improve both object and scene recognition (Oliva & Schyns, 2000; Wurm, Legge, Isenberg, & Luebker, 1993). The vantage point of images also differed from past studies. Rather than presenting wide angle views with many objects, our scenes contained one or two prominent foreground objects pasted into a setting using Adobe Photoshop.

In order to understand how consistency information influences initial perception of a scene, we used a new immediate report paradigm and did not repeat any objects or backgrounds. In the task, participants viewed an 80 ms presentation of a picture followed by a mask. After the presentation, participants typed their responses in a dialog box. This immediate recall task is likely to be a more sensitive measure of initial perception than either the object detection task or the 2AFC task because no suggested identity is provided ,either before or after presentation. Additionally, since each picture was shown a single time, participants could not be biased by previous exposure to the stimuli.

While other studies have mainly focused on how a pictured setting influences object perception, our studies ask new questions. In the first series of experiments, Chapter 2, stimuli consisted of 28 pictures containing a single foreground object that was either consistent or inconsistent with its setting. The following questions are addressed: Does the background context influence perception of a salient foreground object? Does a salient foreground object influence how the setting is perceived? Does the consistency effect remain when both the object and the background must be attended and reported?

The second series of experiments, Chapter 3, investigates the influence of consistency in pictures containing two foreground objects. Does an additional object from the same setting reduce the influence from the background on object perception? Does the presence of an additional object influence background perception? Finally, since objects tend to co-occur with each other, will objects be reported more accurately in the presence of another object from the same setting compared to another object from a different setting?

The experiments in Chapter 4 address the temporal constraints of the consistency effect. Will the consistency effect occur when the object and background are presented sequentially rather than simultaneously?

Finally, in the concluding Chapter 5, the results are summarized and related to ongoing work in neuroimaging.

References

Biederman, I. (1972). Perceiving real-world scenes. *Science, 177(4043),* 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.

Biederman, I., Rabinowitz, J. C., Glass, A. L., & Stacy, E. W. Jr. (1974). On the information extracted from a glance at a scene. *Journal of Experimental Psychology, 103,* 596-600.

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.

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., 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(4),* 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.

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., & Leubker, A. (1993). Color improves object recognition in normal and low vision. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 899-911.

Chapter 2: Scene Consistency in Object and Background Perception

Published as: Davenport, J.L., & Potter, M.C. (2004). Scene Consistency in Object and Background Perception. *Psychological Science, 15,* 559-564

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 post-identification 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.

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.

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 – Racecar	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.

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 x 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 x 768 pixels with a refresh rate of 75 Hz. As displayed, pictures were 17.64 x 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.

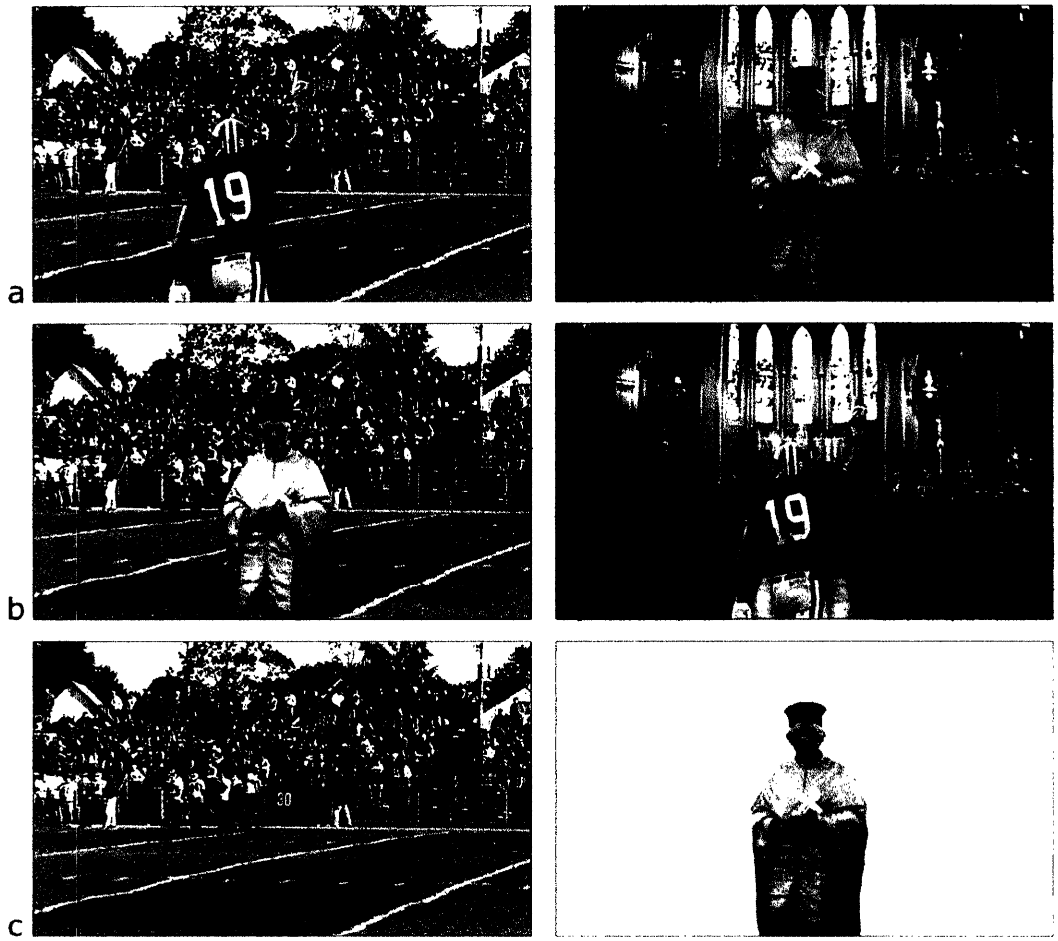

Fig. 1. Examples of consistent scenes (a), inconsistent scenes (b), and isolated objects and backgrounds (c).

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]

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.

---

[1] Intrusions were rare. Out of 168 inconsistent trial sin each experiment, 1 object intrusion occurred in Experiment1, 6 background intrusions occurred in Experiment 2, and 1 object and 12 background intrusions occurred in Experiment 4.

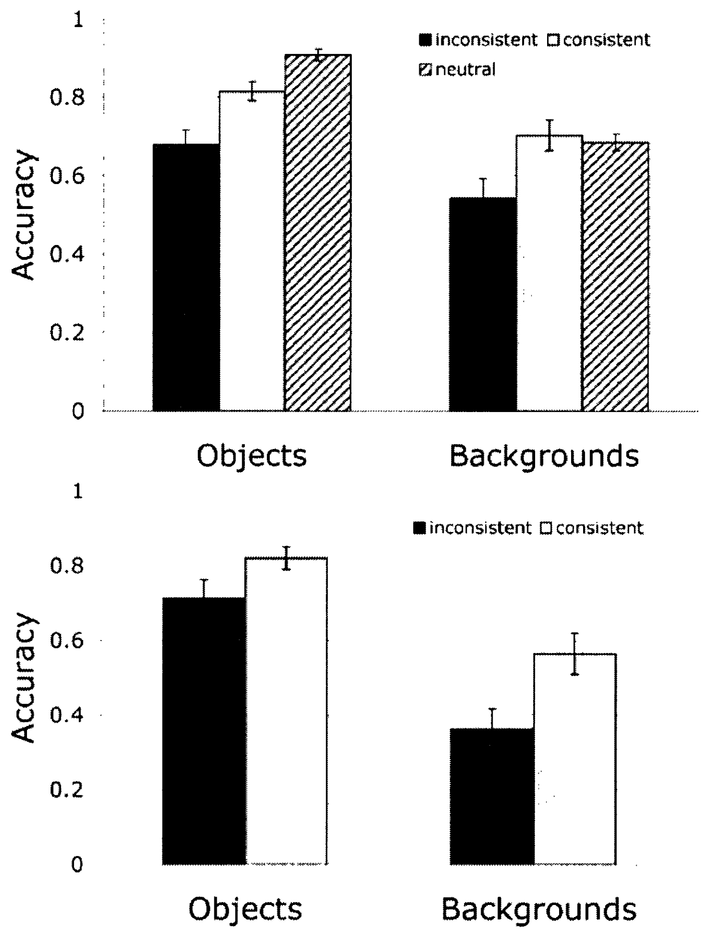

Fig. 2. Accuracy in reporting foreground objects and backgrounds. The top panel shows results when only objects or only backgrounds were reported (Experiments 1-3), and the bottom panel shows results when both objects and backgrounds were reported (Experiment 4). In Experiments 1, 2, and 4, the foreground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

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.

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: Report of both background and object.

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)528.59, 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, $F(1, 22)56.832, p < .05$, 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 co-occur 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.

Chapter 3: Consistency effects between objects in scenes

Abstract

How are objects recognized in context? Objects appear in a given setting with surrounding objects. However, most studies investigating object recognition have presented objects in isolation. Do objects in scenes exert contextual influences on each other? Do these influences interact with background consistency? Three experiments investigated the role of object-to-object context on object and scene perception. Objects (Experiments 1 and 3) and backgrounds (Experiment 2) were reported more accurately when objects and their settings were consistent rather than inconsistent regardless of the number of foreground objects. In Experiment 3, related objects (from the same setting) were reported more accurately than unrelated objects (from different settings) independent of consistency with the background. Object-to-object context and scene context each independently affect object perception.

While a colleague may be immediately recognized at a meeting, it may take longer to recognize this same person in a grocery store. How does context influence our perception of objects and scenes? Objects rarely appear without some context: They are always located spatially within a setting, and usually appear with other objects. However, most prior work on object recognition has focused on objects in isolation. Here, three experiments investigated how objects in scenes interact with each other and their background. The studies bridge two lines of research on semantic consistency effects; the influence of background context on objects in scenes and the influence of object relatedness outside of scenes.

Prior studies of the influence of a background setting on object perception have given mixed results. Some found that objects in typical settings (e.g., a chicken on a farm) were detected more accurately than objects in implausible settings (e.g., a chicken in a living room) (Biederman, Mezzanotte, & Rabinowitz, 1982; Boyce & Pollatsek, 1992; Boyce, Pollatsek, & Rayner, 1989). Others observed no influence of scene context (Hollingworth & Henderson, 1998, 1999). More recently, Davenport and Potter (2004) found robust semantic consistency effects on both object and background perception.

Related objects have also been shown to facilitate each other's perception when presented outside of scenes. Henderson, Pollatsek, and Rayner (1987), found that priming from a previously fixated object shortened the naming latency of a related object. More recently, Auckland, Cave and Donnelly (submitted) used a six-alternative forced-choice paradigm that showed that target objects (e.g., a hand of cards) were identified more accurately when surrounded by semantically related objects (e.g., pictures of dice, dominos and poker chips) than by semantically unrelated objects (e.g., different types of fruit).

While prior research suggests that background settings influence object perception in scenes and that related objects influence object perception in arrays, the interaction of objects in scenes with each other and with their background has rarely been explored. In the single prior study addressing how objects interact in scenes, Boyce et al. (1989) manipulated whether a target object was presented with related or unrelated cohort objects and found no influence of object to object context. However, the sparse line drawings used as stimuli may have made the items difficult to interpret outside of their scene context.

In the present study, Experiment 1 tested whether the presence of related foreground objects eliminates consistency effects with the background. Experiment 2 investigated the effects of related foreground objects on background perception. In Experiment 3, the relatedness between foreground objects was manipulated to determine whether relatedness and background consistency make independent contributions to object perception.

Experiments 1 and 2

Using brief, masked presentations of color photographs of backgrounds with a single foreground object, Davenport and Potter (2004) found that both objects and backgrounds were reported more accurately when they were semantically consistent rather than inconsistent. As scenes usually contain more than one object, and related objects may influence each other, Experiments 1 and 2 test whether the object-background consistency effects are modulated by the number of foreground objects. If objects exert contextual influences on each other, the presence of two related foreground objects (i.e., related because both may be found in the same scene) might eliminate or reduce contextual influence from the setting. However, if the contextual influence of the background is sufficiently strong, the effect of object-background consistency should be remain when two related foreground objects are present. Additionally, two related objects consistent with the background might make accurate perception of the background more likely than when a single object is present, whereas two inconsistent objects might have the opposite effect.

In Experiment 1 participants were instructed to report the prominent foreground object or objects in a briefly presented scene. In Experiment 2 participants reported the background setting of each scene. Scenes had either one or two foreground objects: if two objects were present, they were always episodically related to each other. The background setting was either semantically consistent or inconsistent with the object(s).

Method

Participants. In each experiment, 16 fluent English speakers with normal or corrected to normal vision from the Massachusetts Institute of Technology community volunteered and were paid for their participation. No participant was in more than one experiment.

Materials and apparatus. Stimuli consisted of 40 color photographs of diverse settings. For each background image, two objects that would be likely to appear in that setting were selected from different source photographs. Although each object was consistent with the scene, the objects themselves were not necessarily strong associates. See Table 1 for a list of objects and backgrounds. In this paper objects are termed "related" if they fit into the same consistent scene, whether they were presented to a given participant with that scene or a different, inconsistent scene. The backgrounds and objects were taken from commercially available Photo CDs and the web. Objects were people, animals, furniture, vehicles and the like.

Table 1. Scenes and Objects

Farm (Pig, Tractor)	Arctic (Eskimo, Igloo)
Classroom (Teacher, Overhead projector)	Yard (Lawnmower, Wheelbarrow)
Warehouse (Crates, Forklift)	Forest (Bear, Moose)
Intersection (Ambulance, Traffic cone)	Desert (Rock, Cactus)
Bed (Cat, Teddy Bear)	Beach (Sandcastle, Beach ball)
Park (Stroller, Park bench)	Stage (Piano, Cello)
Pool (Lifeguard, Pool chair)	Track (Runner, Hurdle)
Hospital (Doctor, Nurse)	Church (Priest, Nun)
Sky (Biplane, Helicopter)	Outer Space (Astronaut, Satellite)
Fireplace (Broom, Logs)	Undersea (Turtle, Fish)
Ice rink (Hockey player, Goal)	Ocean (Buoy, Sailboat)
Snowy mountain (Sledder, Skier)	Lake (Jet ski, Motorboat)
Pond (Duck, Frog)	Stove (Kettle, Frying pan)
Serengeti (Zebra, Photographer)	Library (Armchair, Book cart)
Parking lot (Motorcycle, Car)	Arena (Matador, Bull)
Apartment (Sofa, Lamp)	Trail (Cyclist, Jogger)
Patio (Grill, Patio table)	Soccer field (Soccer player, Soccer ball)
Football Field (Football player, Referee)	Hotel Lobby (Bellhop, Luggage)
Restaurant (Chef, Waiter)	Basketball court (Hoop, Basketball player)
Tree (Squirrel, Bird)	Bar (Wine, Wineglass)

Note. Items on the same line were paired

All image manipulation was performed using Adobe Photoshop 7.0. To create the consistent stimuli, the two selected objects were pasted into the "consistent" background. To create the inconsistent stimuli, the background photos were paired, and the objects were exchanged between scenes. For example, an Eskimo and an igloo were consistent in an arctic setting, but inconsistent in a farm setting. The foreground objects appeared in the same locations in each picture and were pasted so that size and support relations were not violated. See Figure 1 for example stimuli. To create the one-object condition, one foreground object was removed. Whether a given setting appeared with one or two objects, which object was presented in the one-object condition and whether these objects were consistent or inconsistent with the setting were fully counterbalanced between subjects. A set of masks was generated by cutting 6 other pictures into a 20 x 20 grid of rectangles and rearranging them randomly.

Figure 1. *Example stimuli. Top: Consistent, Middle: Inconsistent; Bottom (Experiment 3 only): Unrelated. In Experiments 1 & 2 one object was removed in half of the trials.*

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 inch monitor was set to a resolution of 1024 x 768 pixels with a refresh rate of 75 Hz. Pictures as displayed were 17.64 x 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 and the room was normally illuminated.

Design and Procedure. Each of the 40 trials consisted of a single picture followed by a mask. Each participant saw half of the pictures in the consistent condition and half in the inconsistent condition. Of these, half the pictures contained a single object, and half contained two objects from the same setting. The pictures were randomly intermixed and participants never saw an object or background more than once.

Each trial consisted of a fixation "+" for 300 ms, a blank of 200 ms, the test picture for 80 ms and a mask for 200 ms. A dialog box appeared immediately after the mask. In Experiment 1, the dialog box had a single entry area on one-object trials and two entry areas on two-object trials: Participants did not know in advance whether the picture in a given trial would contain one or two objects.

Participants were informed that each picture would contain one or two objects that might or might not belong with the background. In Experiment 1, their task was to type the name of the foreground object(s) into the response box. In Experiment 2, their task was to type the name of the background setting or type of place. Participants were to type

"?" if they did not see the object(s) or background. At the end of the experiment participants were shown each picture again for 500 ms without a mask and named the object(s) and the background setting.

Scoring. All results were scored blind to condition. Responses were scored as correct if they were the same names provided by participants in the post-experimental naming session, or synonyms at an equal level of descriptiveness (e.g. "runner" and "jogger"). Responses were scored as incorrect if they were a different or more general name (e.g. "animal" instead of "pig").

If a subject guessed the object or background that would have been consistent when the inconsistent scene was presented (e.g., responded "igloo" or "farm" when a pig was presented in the arctic scene) the response was considered an intrusion. To correct for such pure guesses based on the background or foreground objects, for each intrusion made by a given participant, one correct consistent response was subtracted. Such intrusions were extremely rare. Intrusions occurred in 0.6% of the responses in Experiment 1, 4% of responses in Experiment 2, and 0.4% of the responses in Experiment 3. All analyses were carried out on the corrected data.

Results and Discussion

In Experiment 1, an analysis of variance (ANOVA) was carried out to determine if accuracy in reporting an object varied as a function of background consistency (consistent vs. inconsistent) and number of objects (1 vs. 2). A highly significant main effect of object-background consistency was found, with objects in consistent settings ($M = .74$) reported more accurately than objects in inconsistent settings ($M = .59$), $F(1, 15) = 49.96$, $p < .001$. No main effect of number of objects was found, $p = .13$, and the interaction of consistency and number of objects did not approach significance, $F(1, 15) < 1$. Results are shown in Figure 2a.

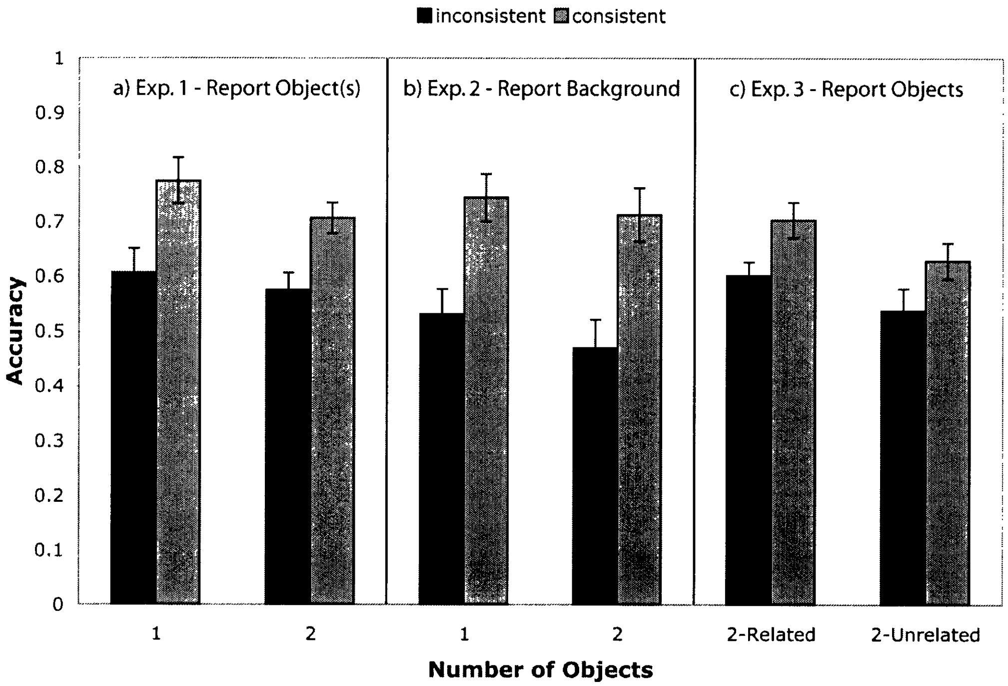

Figure 2. Results of Experiments 1 – 3.

In Experiment 2, in which participants reported the background, a strong main effect of object-background consistency was found with backgrounds reported more accurately when they contained consistent foreground object(s) ($M = .72$) than inconsistent foreground object(s) ($M = .50$), $F(1, 15) = 49.09, p < .001$. Neither the main effect of number of foreground objects, $F < 1$, nor the interaction of consistency and number of objects, $F < 1$, approached significance. Results are shown in Figure 2b.

Thus, robust effects of object-background consistency were present in both experiments, replicating Davenport and Potter's (2004) findings for single objects. Although the background could have been ignored in Experiment 1, the presence of two related objects did not diminish the influence of the background on object perception. In Experiment 2, background report was influenced by the consistency of the foreground object(s), but not by the number of foreground objects. Object-background effects are robust and are not modulated by the presence of an additional object.

Experiment 3

Do objects influence each other's perception? If object-to-object consistency plays a role in perception, an object should be identified more accurately in the presence of a related object than an unrelated object. In Experiment 3, participants named the two foreground objects in each scene. On half the trials the objects were related, so that both or neither were consistent with the background. On the other half of trials, the objects were unrelated; one object was consistent with the background, the other object was inconsistent.

Method

The method of Experiment 3 was identical to that of Experiments 1 and 2 except as noted.

Design and procedure. The same backgrounds and objects as those in Experiments 1 and 2 were used in Experiment 3. Two objects appeared in each picture; on related trials, the objects were both consistent or both inconsistent with the background, as in Experiment 1. On unrelated trials, one object was consistent with the background, one was inconsistent. For example, the farm and arctic scenes were paired to create unrelated trials in which the farm scene contained a tractor and an igloo and the arctic scene contained an Eskimo and a pig. Whether a given setting appeared with two consistent objects, two inconsistent objects or one consistent and one inconsistent object was fully counterbalanced between subjects. Participants were instructed to report the two objects in each trial.

Results and Discussion

An ANOVA was carried out on responses to objects, separately for each object, with object-background consistency and object relatedness as variables. There were significant main effects of object-background consistency and object relatedness. Objects in a consistent setting ($M = .67$) were reported more accurately than objects in an inconsistent setting ($M = .57$), $F(1, 15) = 14.31, p < .01$. Related objects were reported more accurately ($M = .65$) than unrelated objects ($M = .58$), $F(1, 15) = 5.12, p < .05$. Object-background consistency and object relatedness did not interact, $F < 1$. The results are shown in Figure 2c.

The results suggest that object perception is influenced independently by the background setting and by the presence of a related object. Note that object relatedness is defined in this study as potential co-appearance in a given setting; related objects were not selected with the criteria of being strong associates. Thus, it is perhaps surprising that the two objects (e.g., pig and tractor or ambulance and traffic cone) were mutually supportive even when the background was inconsistent (e.g., the arctic or the desert scene).

General Discussion

Three experiments tested the role of semantic consistency in object and background perception. In Experiment 1, objects in consistent scenes were reported more accurately than objects in inconsistent scenes, regardless of the number of foreground objects. In Experiment 2, backgrounds with consistent foreground objects were reported more accurately regardless of the number of foreground objects. In Experiment 3, in addition to a consistency effect, objects appearing with a related object were reported more accurately than objects appearing with an unrelated object, regardless of background. The two effects were independent. Together the findings suggest that objects and scenes are processed interactively, but show separate context effects for object-to-object relatedness and object-background consistency.

The strong object-background consistency effects are in line with prior findings (Biederman et al., 1982; Boyce & Pollatsek, 1992; Boyce et al., 1989; Davenport & Potter, 2004). Attention to multiple foreground objects did not eliminate the influence of the background, even when the background could have been ignored. The object-background consistency effects are robust and the ability to report objects is influenced by their settings.

Experiment 3 demonstrates for the first time that related objects in scenes may influence each other's perception. These results conflict with the findings of Boyce et al. (1989), who failed to find an effect of episodic relatedness between objects in scenes. However, stimuli in the Boyce et al. study were sparse line drawings; objects in isolation may have been difficult to interpret. Also, because the task was to detect the presence of an object named in advance, the task might not have been sufficiently sensitive to detect effects of object relatedness.

The object relatedness effects found in the current study extend prior semantic relatedness effects between objects (Auckland et al., submitted) to objects in scenes. A notable difference between the object relatedness effects in the present experiments and the previous findings is the choice of related objects. Henderson et al. (1987) and

Auckland et al.'s studies used related objects that were strong semantic associates of each other (e.g., "hand and foot" or "hammer, screwdriver, nails and pliers"). However, in the present experiments, related objects were not deliberately chosen to be strong associates of each other. Related objects that are plausible in a given setting (e.g., a park bench and a stroller in a park) were sufficient to facilitate object perception.

The results are inconsistent with a strict version of the functional isolation model of scene processing proposed by Hollingworth and Henderson (1998, 1999) after they failed to find consistency effects in several experiments. That model of scene processing proposed that context does not influence the processing of objects in scenes. The discrepancies between the current studies and the results of Hollingworth and Henderson may have been due to the task, the design or the stimuli. In their studies, the task was either object detection, in which participants indicated yes or no whether a named object was present in a scene, or two alternative forced choice (2AFC) in which subjects indicated which of two named or pictured objects had been presented. Their experiments were designed so that objects and backgrounds were repeated multiple times and the stimuli were black and white line drawings.

The current experiments tested whether the semantic consistency between objects and scenes affects conscious perception the very first time a naturalistic picture is seen. The naming task provides an indication of what information the participant was able to extract in a brief presentation of 80 ms with no further information given in that trial. In contrast, both the object detection and the 2AFC tasks used by Hollingworth and Henderson (1998, 1999) involved a more extended viewing time (150 or 250 ms) and provided information in the trial other than the picture (e.g., the target label in the object detection task or the names in the 2AFC task). This additional information may have enabled post-perceptual reasoning. For example, although the participant may not have extracted enough information to name an object (e.g., a chicken), given two choices (e.g., a chicken or a pig), the participant may have reflected back on the picture and determined that one choice was more likely based on partially processed image. The 2AFC task cannot distinguish whether the viewer saw some sort of bird-like object and so chose "chicken" over "pig," or actually saw a chicken. In the current experiments, the correct name of an object had to be provided at the basic level (e.g., “Eskimo” not “person” and “pig” rather than “animal”). Further, Hollingworth and Henderson may have failed to find consistency effects because the same objects appeared in different scenes and if an inconsistent object was present it was always the item later tested whereas consistent objects had no such distinction. Finally, the present experiments used natural color photographs with salient foreground objects, which may convey consistency information more readily than black and white line drawings with small objects (Cheng & Simons, 2001).

Are the current results due to response bias? As no information other than the picture was presented during a trial, the naming task used here was not susceptible to the type of response bias present in earlier object detection studies (e.g., Biederman et al. 1982; Boyce et al. 1989, as pointed out in Hollingworth & Henderson, 1998, 1999). In those studies, the task was to respond yes or no whether a target label (e.g., sofa) matched an object in a briefly presented scene (e.g., a sofa in a living room). Detection sensitivity, d', was higher when an object appeared in a consistent rather than inconsistent setting, but the d' measure did not take into account response bias from the target label in catch trials.

In the Biederman et al. (1982) study and the replication by Hollingworth & Henderson (1998, Experiment 1), false alarm rates were higher when the target label (e.g., sofa) and pictured scene were consistent (e.g., a living room) rather than inconsistent (e.g., a street); the use of the combined false alarm rate in calculating *d′* led to an overestimation of sensitivity in consistent conditions and an underestimation of sensitivity in inconsistent conditions. However, this type of response bias was not possible in the current studies.

Another possibility is that participants only processed partial information and were using consistency to guess which objects or backgrounds were presented. This type of pure guessing was conservatively corrected for before the data analysis. For example, if the task was to report the objects, and participants saw only the background and simply guessed a related object, their responses would have included intrusions of consistent objects when an inconsistent object was presented (e.g., reporting "lawn mower" when an overhead projector was presented in a yard). Intrusions of this type were very rare, and were corrected for by subtracting one correct consistent response for each such intrusion. Thus, if a participant used a pure guessing strategy, all intrusions would be subtracted from correct consistent responses and the net score would be 0 in both conditions.

So how might the consistency effect be occurring? The current studies address how context influences conscious perception of briefly presented pictures. It seems likely that, rather than penetrating and influencing early vision, conceptual knowledge about the world affects later stages of perceptual processing, biasing perceptual interpretation in a Bayesian fashion (e.g., Knill & Pouget, 2004). When viewing time is very short, as in the present experiments, there will be some degree of uncertainty about the background, the objects, or both. The prior likelihood that an igloo is on a farm is low, so more evidence is required before the perceptual system concludes that there actually is an igloo. For a pig, however, the prior likelihood is much higher, so less evidence is required. This kind of bias has been shown to be common in vision, for example, in identifying letters in words (e.g., Massaro & Cohen, 1991). In everyday life this may allow us to more easily recognize people or objects far away (or in other degraded conditions), with less information that would be needed to recognize an unexpected item.

Can this inherent perceptual bias toward the more probable interpretation – what Helmholtz (1925) called "unconscious inference" – be distinguished from another, more conscious kind of bias often called "sophisticated guessing"? For example, in the latter case the viewer sees a farm scene with something that looks like an animal, reasons that it is more likely to be a pig than a wolf or hyena, and so simply guesses that it is a pig. Both the sophisticated guessing hypothesis and the Bayesian model assume a bias toward the probable, but the Bayesian claim is that the bias is built into perception, rather than being a conscious guessing strategy subsequent to perception. Because participants in the present experiments were asked to report what they saw and to enter a question mark if they did not know, and because they rarely came up with a guess (in the inconsistent condition) that matched the corresponding item in the consistent condition, it seems likely that the effects of consistency and relatedness were unconscious, affecting what participants perceived rather than what they simply guessed. However, further work will be required to show conclusively that these effects occur before conscious perception.

The current results suggest an interactive model of object perception in scenes in which object-background and object-object effects are independent. The gist or meaning of a scene may be extracted from a brief glimpse of 100 ms (Potter, 1975) leading prior models of semantic consistency in scene perception to focus on how background information influences object perception (Biederman et al., 1982; Boyce & Pollatsek, 1992; Friedman, 1979). The present study also looks at how an object influences the perception of another object, in the presence of a consistent or inconsistent background. An interactive model of scene processing suggests that scenes are processed holistically, with mutually constraining object and background processing occurring in parallel. Scenes containing related objects and consistent settings may require less perceptual information for identification of the individual elements. In summary, the present study shows that backgrounds influence how objects are perceived, and that objects influence perception of other objects and their backgrounds.

References

Auckland, M., Cave, K. R., & Donnelly, N. (submitted). Response bias can only partially explain contextual facilitation in object recognition.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. Cognitive Psychology, 14(2), 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(3), 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(3), 556-566.

Brainard, D. H. (1997). The psychophysics toolbox. Spatial Vision, 10(4), 433-436.

Cheng, E. K., & Simons, D. J. (2001). Perceiving the internal consistency of scenes. Paper presented at the 42nd Annual Meeting of the Psychonomic Society, Orlando, FL.

Davenport, J. L., & Potter, M. C. (2004). Semantic consistency in object and background perception. Psychological Science, 15(8), 559-564.

Friedman, A. (1979). Framing pictures: Role of knowledge in automatized encoding and memory for gist. Journal of Experimental Psychology-General, 108(3), 316-355.

Helmholtz, H. (1925). Physiological Optics, Vol. III: The perceptions of Vision (J. P. Southall, Trans.). Rochester, NY: Optical Society of America.

Henderson, J. M., Pollatsek, A., & Rayner, K. (1987). Effects of Foveal Priming and Extrafoveal Preview on Object Identification. Journal of Experimental Psychology-Human Perception and Performance, 13(3), 449-463.

Hollingworth, A., & Henderson, J. M. (1998). Does consistent scene context facilitate object perception? Journal of Experimental Psychology-General, 127(4), 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(2-3), 319-343.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. Trends in Neuroscience, 27(12), 712-719.

Massaro, D. W., & Cohen, M. M. (1991). Integration Versus Interactive Activation - the Joint Influence of Stimulus and Context in Perception. Cognitive Psychology, 23(4), 558-614.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. Spatial Vision, 10(4), 437-442.

Potter, M. C. (1975). Meaning in visual search. Science, 187(4180), 965-966.

Chapter 4: Consistency effects in scenes: Guessing or unconscious inference?

Abstract

In prior studies (Davenport, under review; Davenport & Potter, 2004), knowledge about which objects and backgrounds tend to co-occur influenced perception. Objects that were semantically consistent with their settings (e.g., a sofa in a living room) were identified more accurately than objects that were inconsistent with their settings (e.g., a sofa in a street). Was this consistency effect the result of unconscious inference during perception, or was it the result of a post-perceptual process of sophisticated guessing? The consistency effect did not occur when objects and backgrounds were presented sequentially (Experiment 1), but did occur when objects and backgrounds appeared simultaneously in the same picture (Experiment 2). The results suggest that the consistency effect is temporally limited and is not due to sophisticated guessing.

How does knowledge about the world influence perception? Objects are not distributed throughout space at random, but are more likely to appear in some settings than others. For instance, sofas usually appear with lamps and other furniture in living rooms, but rarely on the beach. In recent studies (Davenport, under review; Davenport & Potter, 2004) strong consistency effects were demonstrated in masked, briefly presented scenes when the task was to report either the object or setting of the picture. Both objects and backgrounds were reported more accurately when presented in consistent scenes (e.g., an igloo in the arctic) compared to inconsistent scenes (e.g., an igloo on a farm). Additionally, objects were reported more accurately when presented with an object from the same setting (e.g., an igloo and an Eskimo), compared with an object from a different setting (e.g., an igloo and a pig). Consistency information was available in an 80 ms glimpse of a picture, suggesting that perception makes use of unconscious inferences about co-occurrence based on Bayesian principles.

An open question from prior work is whether the consistency results reflect the initial interpretation of an image or a sophisticated guessing process on the part of participants. Pure guessing ("I saw a farm, so I'll guess a pig") was controlled for in these studies, by subtracting the number of such false guesses on inconsistent trials from correct reports on consistent trials. However conscious sophisticated guessing based on partial information ("I saw an artic scene so I'll guess that the white blob was an igloo") was not ruled out. That is, the consistency benefit might not have occurred during perception of the scene, but only by a reasoning process after the image disappeared.

If participants are reasoning about what they have seen after the presentation, consistency effects should be present and perhaps amplified if the object and background information are presented serially rather than simultaneously. However, if the consistency of an object and background influences how the scene is initially perceived, the consistency effect should be reduced or eliminated by presenting the object and background separately.

Experiment 1

Experiment 1 tested whether consistency effects would occur when object and background information were presented serially instead of simultaneously. If the previously reported consistency effects were the result of deliberate sophisticated guessing rather than unconscious inference, presenting the object and background apart from one another should enable consistency effects to occur and perhaps be magnified. However, if scene information constrains how an object is interpreted during the presentation, displaying the object and background separately should show no such consistency effects.

Participants saw 80 ms presentations of a masked object in noise and a masked background scene, separated by a 500 ms blank, black screen. Their task was to report only the object, and they were informed that it would sometimes appear first and sometimes appear second.

Method

Participants. Sixteen fluent English speakers with normal or corrected to normal vision from the Massachusetts Institute of Technology community volunteered and were paid for their participation. No participant had been in any prior scene consistency experiment.

Materials and apparatus. Stimuli consisted of 80 color photographs of objects and 40 color photographs of diverse settings, which were modified from Davenport (under review). For each background image, two objects that would be likely to appear in that setting were selected from different source photographs. The objects and backgrounds were taken from commercially available Photo CDs and the web. Objects were people, animals, furniture, vehicles and the like.

All image manipulation was performed using Adobe Photoshop CS. Objects were cropped out of their original pictures and pasted onto a mottled colored background, to make them approximately as easy to recognize as if they had been presented in a natural scene (see Figure 1). Objects were always pasted near the center of the image, in a location that would have been plausible had they appeared in the associated scene. Background photographs contained no large foreground objects and were not further modified.

A set of masks was generated by cutting 6 pictures taken from similar sources into a 20 x 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 inch monitor was set to a resolution of 1024 x 768 pixels with a refresh rate of 75 Hz. Pictures as displayed were 17.64 x 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 and the room was normally illuminated.

Design and Procedure. Each of the 40 trials consisted of two masked presentations, an object and a setting, separated by a 500 ms blank. Trials were either consistent (the background was appropriate for that object) or inconsistent (the background was inappropriate for that object). To create inconsistent trials, objects from each consistent combination were exchanged (e.g., the igloo appeared with the farm scene and the pig with the arctic scene). As there were two potential objects consistent with each background setting, a given participant saw only half the objects (counterbalanced between subjects) Whether the object and setting were consistent and whether the object appeared first or second were counterbalanced within and between subjects. No object or background was presented more than once.

Participants pressed the space bar when they were ready to begin a trial. Each trial consisted of a fixation "+" for 300 ms, a blank of 200 ms, either the object or background for 80 ms and a mask for 200 ms. The first mask was followed by a blank screen for 500 ms, the background or the object for 80ms and a second mask for 200ms. A dialog box appeared immediately after the mask with a single entry area and the prompt "object?". Participants did not know in advance whether the object or background would appear first.

Participants were informed that on each trial there would be a picture of an object embedded in colorful noise and a background scene. Their task was to type the name of the object into the response box. Participants were to type "?" if they did not see the object.

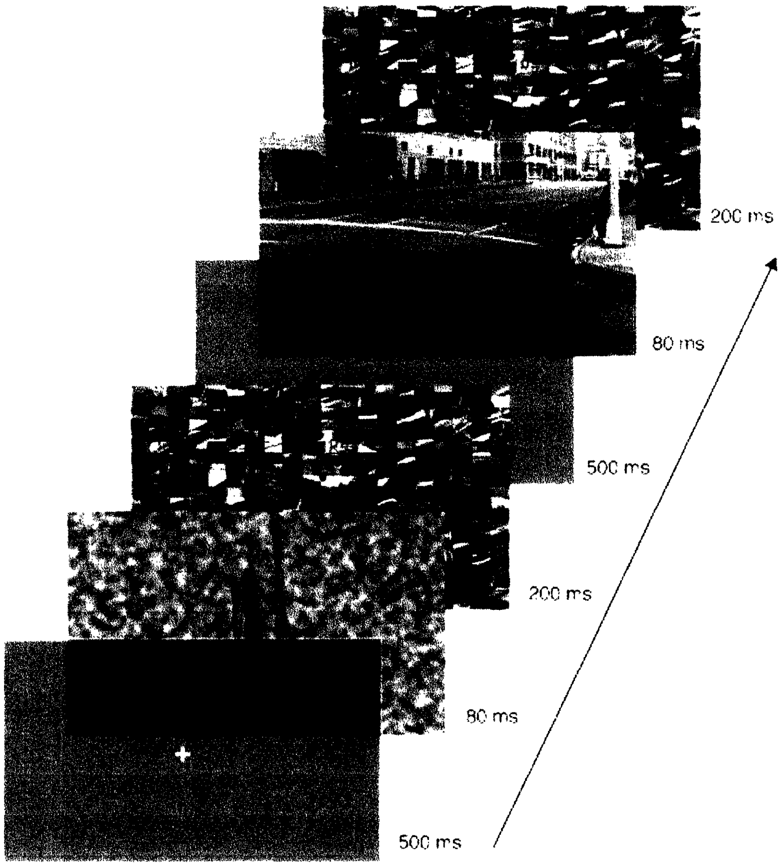

Figure 1. Sample trial from Experiment 1.

Scoring. All results were scored blind to condition. Responses were scored as correct if they were the same names provided by participants in the post-experimental naming session from the Davenport (under review) series of studies, or synonyms at an equal level of descriptiveness (e.g. "runner" and "jogger"). Responses were scored as incorrect if they were a different or more general name (e.g. "animal" instead of "pig").

Results and Discussion

An analysis of variance (ANOVA) was carried out to determine whether accuracy in reporting the object varied as a function of the consistency of the background picture presented on the trial (consistent vs. inconsistent) or as a function of whether the object appeared first or second on the trial (see Figure 2). No significant main effects or interactions were found. Objects on trials with consistent backgrounds ($M = .64$) were not reported significantly more accurately than objects on trials with inconsistent backgrounds ($M = .63$, $F < 1$). No significant main effect of order ($p = .21$) or interaction between order and consistency was found ($F < 1$).

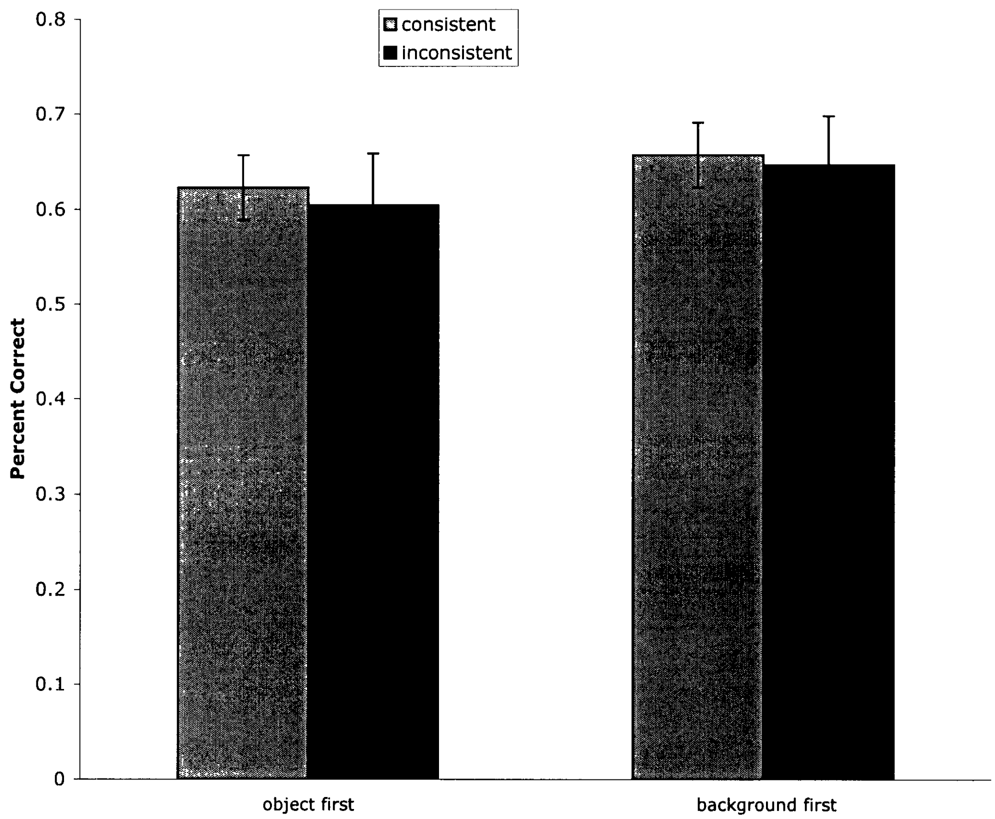

Figure 2. Results of Experiment 1.

In Experiment 1 the consistency effect was not present when the object and background were presented serially. If the consistency effect in previous studies was a result of sophisticated guessing, that is, conscious reasoning about the object based on the background scene, a similar or magnified consistency effect should have been present in the current experiment. Whether the background information was presented before or after the target object did not influence accuracy.

Experiment 2

In Experiment 1, no consistency effect was found when the task was to report an object presented in noise. Experiment 2 was run as a comparison experiment with the same materials presented simultaneously in a single scene to replicate the consistency effect observed in previous studies. In Experiment 2, participants were told that each trial would consist of two events, a masked picture of mottled colors and a masked picture of a scene. Their task was to report the foreground object of the scene, and they were instructed that the scene would appear either before or after the masked picture of mottled colors.

Method

The method of Experiment 2 was the same as Experiment 1 except as noted.

Participants. Sixteen fluent English speakers with normal or corrected to normal vision from the Massachusetts Institute of Technology community volunteered and were paid for their participation. No participant had been in any prior scene consistency experiment.

Materials and apparatus. Stimuli were modified from those used in Experiment 1 and consisted of 40 color photographs of diverse settings. To create the consistent stimuli, each of the two selected objects was pasted separately into the center of the "consistent" background. To create the inconsistent stimuli, the object in an inconsistent pairing was pasted into its scene. Foreground objects appeared in the same locations in each picture and were pasted so that size and support relations were not violated. The consistency of a given trial and whether the object appeared first or second were fully counterbalanced between subjects, as in Experiment 1. Participants never saw an object or background more than once. The mottled noise image and the masks were created by the procedures described in Experiment 1.

Design and Procedure. Each of the 40 trials consisted of two masked presentations separated by a 500 ms blank. The scene was either consistent or inconsistent. One scene and one noise image was presented on each trial; order was counterbalanced.

Participants pressed the space bar when they were ready to begin a trial. Each trial consisted of a fixation "+" for 300 ms, a blank of 200 ms, either the scene or mottled noise for 80 ms and a mask for 200 ms. The first mask was followed by a blank screen for 500 ms, the noise or scene for 80ms and a second mask for 200ms. A dialog box appeared immediately after the mask with a single entry area and the prompt “object?”. Participants did not know in advance whether the scene or noise image would appear first.

Participants were informed that on each trial there would be a picture of a scene with a foreground object. Their task was to type the name of the object into the response box or to type "?" if they did not see the object.

Scoring. Results were scored as in Experiment 1. Additionally, if a subject guessed the object that would have been consistent when the inconsistent scene was presented (e.g., responded "igloo" when a pig was presented in the arctic scene) the response was considered an intrusion. To correct for such pure guesses based on the background, for each intrusion made by a given participant, one correct consistent response was subtracted. Such intrusions were extremely rare and only occurred twice out of 320 inconsistent trials. All analyses were carried out on the corrected data.

Results and Discussion

An ANOVA was carried out to determine whether consistency (consistent vs. inconsistent) or order (scene 1st or 2nd) varied with the accuracy in reporting the foreground object (See Figure 3). A significant main effect of consistency was found. Objects appearing with consistent backgrounds ($M = .65$) were reported more accurately than objects appearing with inconsistent backgrounds ($M = .57$), $p < .05$. There was also a significant interaction of consistency and order, $p < .05$. A planned comparison revealed that the consistency effect was significant when the scene appeared first ($p < .01$) but not when the scene appeared second ($F < 1$). Finally, an ANOVA comparing the object-first trials of Experiment 1 with the comparable scene-first trials of Experiment 2 found a significant interaction of experiment and consistency, $p < .05$.

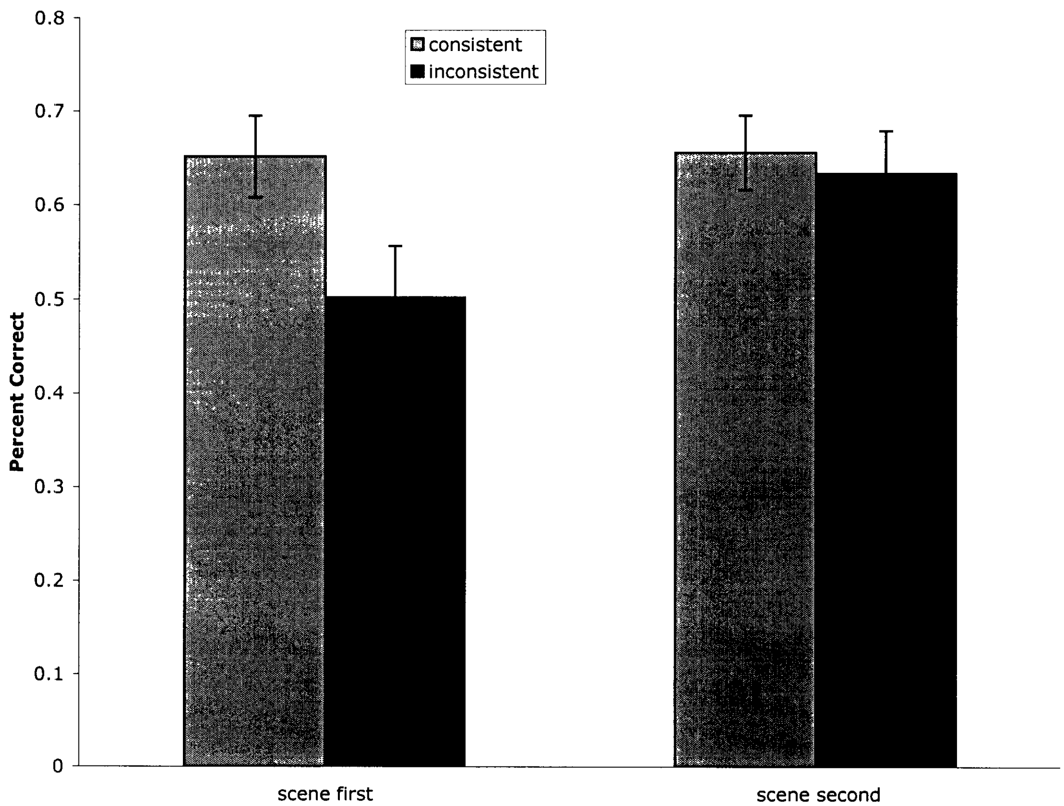

Figure 3. Results of Experiment 2.

Replicating prior research, a consistency effect was found when the scene appeared as the first item in the trial with objects reported more accurately in consistent rather than inconsistent settings. Unlike Experiment 1, in which the background on the trial did not influence object perception, the current results suggest that the object and background must appear simultaneously in order for a consistency effect to occur. However, even when the object and background appeared together, when the scene was in the second position, no consistency effect was found.

General Discussion

Two experiments investigated which conditions led the background setting to influence object processing. In Experiment 1, objects and backgrounds were presented sequentially and no consistency effects were found. In Experiment 2, objects and backgrounds were presented together in a single picture, and consistency effects were found when the picture appeared at the beginning of the trial. Taken together, these studies suggest that the consistency effect is not a product of conscious reasoning or sophisticated guessing, but instead may reflect an unconscious inferential process of cognition.

When in processing do the consistency effects of briefly presented scenes occur? In prior studies (Davenport, under review; Davenport & Potter, 2004), we suggested that knowledge about the world (e.g., what objects and background tend to co-occur) biases perception in a Bayesian fashion through a process of unconscious inference. With a short glimpse of a picture, there may be uncertainty about the objects or setting of a scene. The prior likelihood that a cactus is in the middle of an intersection is low, so more information may be needed before the perceptual system concludes there is actually a cactus present. The likelihood of an ambulance on a street is much higher, so less information may be required. In everyday life context may facilitate recognition of objects that are occluded, far away or in other degraded condition, with less information than would be needed to perceive unexpected items.

The goal of the current experiments was to separate unconscious inference from a more deliberate process of sophisticated guessing. The sophisticated guessing hypothesis suggests that participants are partially processing the object information, and then consciously reasoning in order to guess a likely object. For example, if the farm scene is shown, subjects may see some type of animal, and then reason that a pig would be more likely than a wolf to appear in such a setting. If the background serves to provide a clue about the type of object that is later used to reason with, presenting the background apart from the object should enable the same sophisticated guessing strategy as presenting the object and background together in one picture. The fact that no consistency effect occurred when the objects and backgrounds were presented serially suggests that this type of sophisticated guessing was unlikely. Additionally, subjects were able to report "?" if they did not see the object and extremely few responses in the inconsistent condition matched the correct response in the consistent condition.

The results of Experiment 1 differ from those in a prior study by Palmer (1975) that found that a scene primed object perception. In Palmer's experiment, all stimuli were line drawings and the scene appeared for 2 seconds, without a mask, prior to the briefly presented object. Here, both objects and backgrounds were presented briefly and masked. Because the presentation times and design vary, the two studies may be tapping into different processes.

The present experiments also revealed that while the consistency effect appeared when a briefly presented picture must be interpreted, it may not occur in other circumstances. In particular, in Experiment 2, when the scene was preceded by a masked noise image, the consistency effect did not occur. The results of Experiment 1 do not seem to have been influenced in this manner, as no consistency effect was present regardless of whether the background appeared in the first or second position. In further studies, in which both the object and background were to be reported, a significant consistency effect did occur for pictures in the second position.

In conclusion, object perception may only be influenced by its background when the object and background are presented simultaneously, suggesting that the consistency effect is due to unconscious inference rather than sophisticated guessing. Future studies will investigate what conditions are necessary for the effect to occur.

References

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433-436.

Davenport, J. L. (under review). Contextual effects in object and scene processing.

Davenport, J. L., & Potter, M. C. (2004). Semantic consistency in object and background perception. *Psychological Science*.

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

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*(4), 437-442.

Chapter 5: Conclusions

Does knowledge about what objects and backgrounds tend to co-occur influence object and background perception? Three series of studies provide evidence that the semantic consistency between objects and their settings is available in a glimpse of 80 ms and influences perception.

The first series of experiments investigated the influence of semantic consistency in naturalistic photographs containing a single foreground object in a setting. In Experiment 1, objects in consistent settings were reported more accurately that objects in inconsistent settings. In Experiment 2, backgrounds were reported more accurately when they contained a consistent rather than inconsistent foreground object. In Experiment 3, objects and backgrounds were presented alone and results were compared with the first two experiments. Objects in isolation were reported more accurately than objects in any background. Backgrounds with no foreground object were reported as accurately as backgrounds containing a consistent object, and more accurately than those containing an inconsistent object. Finally, in Experiment 4, the object-background consistency effect was robust and remained when both objects and backgrounds were to be reported.

Just as objects and backgrounds tend to co-occur, objects usually appear with a subset of other objects. The second series of experiments investigated consistency effects between objects in settings. Objects (Experiments 1 and 3) and backgrounds (Experiment 2) were reported more accurately when objects and their settings were consistent rather than inconsistent regardless of the number of foreground objects. In Experiment 3, related objects (from the same setting) were reported more accurately than unrelated objects (from different settings) independent of consistency with the background. Object-to-object context and scene context each independently affect object perception.

Finally, the last series of studies explored the temporal limitations of the consistency effects. In Experiment 1, no consistency effects on object perception were found when the object and background were presented serially with a stimulus onset asynchrony (SOA) of 780 ms., providing no support for the suggestion that the consistency effect is driven by post-perceptual reasoning processes. In Experiment 2, objects in consistent settings were reported more accurately than objects in unusual settings. However, the consistency effect was not present if the picture was presented after masked noise.

How might this consistency effect be occurring? During the initial perception of a scene, objects and backgrounds may be mutually constraining, biasing perception in a Bayesian fashion. When viewing time is limited, there will be some uncertainty about the background, the objects or both. For example, the prior likelihood of a buffalo appearing in a bowling alley is extremely low, so more perceptual information may be required for identification than when a buffalo appears on the plains. In everyday life, this may allow us to recognize people or objects from a distance or in other degraded conditions with less perceptual information than would be required for an unexpected item.

Bar (2004) has proposed a model of contextual facilitation that speculates how individual object perception may be influenced by context. In this model, two types of information, a context frame and candidate objects, are processed in parallel and used to constrain the interpretation of an object. The context frame, based on global low spatial frequency scene information, provides prototypical information about what objects usually occur and where. Candidate objects are selected by using shape information from local low spatial frequency image areas. Bar proposes that the context frame is generated in parahippocampal cortex, whereas candidate objects are generated in prefrontal cortex. These two sources of information are combined in the inferior temporal cortex (ITC), and their intersection produces reliable selection of the object identity.

While Bar's model accounts for object facilitation, a modification would be necessary to account for the influence of a foreground object on background perception. As suggested by neuroimaging research, summarized below, objects and backgrounds may be preferentially processed in separate areas of the cortex. While ITC may represent object identity, other areas may represent backgrounds or settings.

Related neuroimaging results

As the present experiments use only behavioral measures, the results cannot speak to the time course or neurological loci of the consistency effects. It seems likely that, rather than penetrating and influencing early vision, conceptual knowledge influences later stages of processing. In an event-related potential (ERP) study investigating the time course of scene consistency on object perception, Ganis and Kutas (2003) compared waveforms for objects in consistent versus inconsistent scenes. They found an N390 congruity effect and found no congruity effects before 300 ms or after 500 ms. As the N390 waveform is similar to the N400 found for semantic inconsistency in sentence comprehension, Ganis and Kutas concluded that scenes only influence object perception after semantic knowledge has been activated.

While ERP studies provide good temporal resolution, they provide limited spatial resolution. Functional magnetic resonance imaging (fMRI), may be better suited to determining where the consistency effect may be occurring. Prior research has suggested that different areas of the brain may be selective for scene versus object processing.

Epstein and Kanwisher (1998) found a brain region selective to the processing of scenes rather than objects in the parahippocampal cortex, dubbed the parahippocampal place area (PPA). On the other hand, a brain region in the lateral occipital complex (LOC) produces selective activation to objects compared to scenes or faces (Grill-Specter, Kourtzi & Kanwisher, 2001).

One possible locus of the consistency effect may be regions in the parahippocampal cortex. Goh, Siong, Park, Gutchess, Hebrank and Chee (2004) found areas responsible for binding objects and scenes in bilateral parahippocampal and right hippocampal regions distinct from scene processing areas. Using an adaptation paradigm, they showed participants four pictures in a series under four conditions; in each case the object and setting were semantically consistent. In the OO condition (Old object, Old scene), the same object and scene was repeated four times, in the ON condition (Old object, New scene) the same object was repeated in four different scenes, in the NO condition (New object, Old scene) four new objects appeared in a repeated background scene, and finally in the NN (New object, New scene) condition, four different object and scene pairs were presented. Goh et al. found LOC adaptation when the object remained the same but the background changed and PPA adaptation when the background remained the same but the object changed. The binding regions only showed adaptation when the unique object and background pairing was repeated.

Bar and Aminoff (2003) suggested that regions in the parahippocampal cortex may mediate contextual processing. They found that "strong-contextual objects" (objects strongly associated with a type of place, e.g., a beach chair) produced greater activation in these areas than "weak-contextual objects" (objects that could appear in many settings, e.g. a fly or a camera). However, as neither study used semantically inconsistent stimuli, further research is required to determine whether these regions would respond with differential activation to consistent versus inconsistent object-background pairs.

Conclusions

Knowledge about the world, particularly what objects and backgrounds tend to co-occur, influences the perception of briefly presented scenes. Robust behavioral data suggests that background perception is influenced by the presence of salient foreground objects and that object perception is influenced both by their backgrounds and other objects. Objects and their settings are processed interactively, not in isolation.

Further, the results raise interesting questions about how, where and when these consistency effects occur. At some point before conscious perception, perceptual information makes contact with stored conceptual representations. Does this contact happen in all circumstances or only in situations where perceptual information is limited due to short viewing time or long distances? How early in processing does conceptual information influence what is perceived? Future neuroimaging studies may be able to resolve these open questions and give a more detailed account of how what we know influences what we see.

References

Bar, M., & Aminoff, E. (2003). Cortical analysis of visual context. *Neuron, 38,* 347-358.

Bar, M. (2004). Visual objects in context. *Nature Neuroscience Reviews, 5.* 617-629.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature, 392,* 598-601.

Ganis, G., & Kutas, M. (2003). Electrophysiological study of scene effects on object identification. *Cognitive Brain Research, 16,* 123-144.

Goh., J.O.S., Siong, S.C., Park, D., Gutchess, A., Hebrank, A, & Chee, M.W.L. (2004). Cortical areas involved in object, background, and object-background processing revealed with functional magnetic resonance adaptation. *Journal of Neuroscience, 24,* 10223-10228.

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research, 41,* 1409-1422.