THE SCENE SUPERIORITY EFFECT: OBJECT RECOGNITION IN THE CONTEXT OF NATURAL SCENES

BY

RICHARD YAO

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Arts in Psychology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2011

Adviser:

Professor Daniel J. Simons

ABSTRACT

Four experiments investigate the effect of background scene semantics on object recognition. Although past research has found that semantically consistent scene backgrounds can facilitate recognition of a target object, these claims have been challenged as the result of post-perceptual response bias rather than the perceptual processes of object recognition itself. The current study takes advantage of a paradigm from linguistic processing known as the Word Superiority Effect. Humans can better discriminate letters (e.g., D vs. K) in the context of a word (WORD vs. WORK) than in a non-word context (e.g., WROD vs. WROK) even when the context is non-predictive of the target identity. We apply this paradigm to objects in natural scenes, having subjects discriminate between objects in the context of scenes. Because the target objects were equally semantically consistent with any given scene and could appear in either semantically consistent or inconsistent contexts with equal probability, response bias could not lead to an apparent improvement in object recognition. The current study found a benefit to object recognition from semantically consistent backgrounds, and the effect appeared to be modulated by awareness of background scene semantics.

TABLE OF CONTENTS

INTRODUCTION

Imagine that you are picking up your friend John from the airport. Although you are good work friends, the two of you never really spent time together outside the office. In your mind, John and the office are strongly associated with one another. As you scan the crowds pouring out of the "arrivals" terminal, you look among the tired travelers' faces for John's, perhaps even modulating your attention to selectively attend only to the men in the crowd. As you dutifully search the rush of human traffic for your friend, you suddenly feel a tap on your shoulder. "Hey! Hope I haven't kept you waiting." Instead of responding, you stare at this man blankly. A long second (or two) goes by before recognition finally washes over your face and you realize, in a surprise twist, John has found you!

Such an experience of delayed recognition is common. Although a number of factors could come into play when explaining the phenomenon (e.g., devoting attention to a demanding task, the task-switching cost of moving attention from said task), our interest here lies in the effect of context: how, and to what degree, does visual context modulate our ability to correctly recognize an object? (For the sake of our story, let us just assume, for now, that recognizing John is like recognizing an exemplar of any object category, conveniently ignoring a massive literature on human face processing.) We can imagine conditions under which semantic consistency between an object and a scene can influence recognition. In our example above, trying to recognize John in an unusual environment where you normally do not see him may have hindered recognition. To put it another way, an object's typical context may prime semantic networks relevant to the recognition process. On the other hand, placing an object outside its typical context may make it stick out, thereby attracting attention and speeding up

recognition in cases where attention has not already been brought to the object (as it was in our example with John).

Classic Research on Object Recognition in Context

The investigation of object recognition and context goes back decades, though the first papers of notoriety today began to appear in the 1970s. Experimenters would present subjects with line drawings of objects (e.g., a rooster) in the context of either a semantically consistent scene (in this case, a farm) or a semantically inconsistent scene (a city street, for example) and record subsequent behavior (e.g., Palmer, 1975; Loftus and Mackworth, 1978; Biederman, Mezzanote, and Rabinowitz, 1982). One such experiment tracked subjects' eye gaze while they viewed such scenes under instructions to scan the images as if their memories for them would be tested later. Compared to the semantically consistent objects, subjects fixated semantically inconsistent objects sooner, more frequently, and for longer (Loftus and Mackworth, 1978). Objects that do not "belong" in a scene attract our gaze. Whether or not that actually modulates recognition, however, is left unanswered.

The first experiment to directly address object recognition in scene context primed subjects with a scene for 2 seconds and then asked them to identify a rapidly presented object in isolation. Relative to a no-scene baseline condition, semantically consistent scenes increased reporting accuracy whereas inconsistent scenes decreased it (Palmer, 1975). Although it would seem that a semantically consistent context facilitates object recognition, this particular experiment does not speak to the nature of that facilitation. For instance, a scene may have simply biased subjects to report a semantically consistent object when probed. Depending on the target object and the likelihood of subjects guessing it spontaneously, a semantically consistent

scene may have simply improved guessing by constraining the range of objects a subject guessed. Furthermore, because the scenes and objects never appeared together, the scenes could have only acted as a conceptual prime, increasing the likelihood that the nature of the facilitation was in postperceptual processes (e.g., response bias) rather than object recognition processes proper.

A more extensive study on object recognition sought to remedy some of these issues and investigate the effects of scene context on the perceptual processes of object recognition (Biederman, Mezzanote, and Rabinowitz, 1982). The experimenters used a detection paradigm in which subjects first saw the name of an object that could appear on the display in the coming trial. A scene appeared for 150 ms, followed by a mask and a cue. Subjects then had to report whether or not they saw the object named at the beginning of the trial at the cued location. The design allowed experimenters to use signal detection theory (see: MacMillan and Creelman, 2005) to estimate subjects' sensitivity to perceiving objects (d') under different conditions, independent of response bias. Objects could violate a variety of properties normally found in a well-formed scene: occlusion, physical support, typical size, typical position, and—most importantly for the current study—semantic consistency. Compared to the baseline condition, which consisted only of well-formed scenes containing semantically consistent objects, sensitivity decreased additively with the introduction of more physical violations. In short, object recognition suffers in the context of a semantically inconsistent scene.

A Challenge to the Effect of Context

The literature up to this point supports the hypothesis that context does in fact influence object recognition in some way; however, our hypothesis may not actually hold true. The

previously described study (Biederman et al., 1982) has been directly challenged on the grounds of numerous experimental design issues that could overstate the effect of context (Hollingworth and Henderson, 1998). First, the data were not computed in line with the background theoretical assumptions of signal detection theory. Sensitivity (when measured by d') is a function of a subject's hit rate and false alarm rate to a stimulus (the "signal") under conditions that differ crucially in the presence or absence of a single target stimulus. For example, a "hit" trial might begin with the label "chicken," then display and test the subject on an image that actually contains a chicken. The corresponding "false alarm" or "catch" trial would also begin with the label "chicken" and then use an image without a chicken.

In the previously discussed study, however, the "catch trials" from which the experimenters calculated false alarm rate were not an appropriately matched, uniform baseline. Instead, 70% of catch trials began with a semantically consistent label (e.g., "horse" in a farm scene) and 30% began with a semantically inconsistent label (e.g., "television" in a farm scene). Without identical labels in the hit and false alarm conditions, we face two problems. First, the different object labels conflate two different types of detection; the "hit" trials described above measured subjects' ability to detect a chicken, whereas the "catch" trials described above measure subjects' ability to detect horses and televisions. The design therefore violates a core assumption of signal detection theory: that d' is calculated from the hit and false alarm rates for a single detection task.

Second, the different object label trials conflated in the false alarm rate could have introduced biases that inflated the consistency effect. People might reasonably assume a semantically consistent object appeared in a scene rather than a semantically inconsistent one. Seeing a horse in a farm pasture seems much more likely than seeing a television. That intuition

increases false alarm rates for semantically consistent labels and decreases false alarm rates for semantically inconsistent labels. Unfortunately, the data indicate people have exactly that bias. Consequently, the false alarm rate is lower in the consistent background condition than it would have been if the label were always semantically consistent; likewise, the false alarm rate is higher in the inconsistent background condition than it would have been if the label were always inconsistent. In turn, d' appears higher than it should be in the consistent condition and lower than it should be in the inconsistent condition. Indeed, replicating the original experiment yields an effect of context on object detection sensitivity, but remedying the issues with the experimental design makes that effect disappear (Hollingworth and Henderson, 1998). Therefore, the effect of scene context may not act on processes as fundamental as perception or recognition of the object, but rather reflects a postperceptual response bias.

Recent Work and Photographic Stimuli

In spite of the challenges to a semantic consistency effect on object recognition, more recent evidence continues to point to its existence. In one experiment (Green and Hummel, 2007), subjects completed an object detection paradigm similar to those described above, but rather than scene backgrounds, the targets appeared in proximity to another object upon which they could act (e.g., a screwdriver and a screw, a pitcher and a glass). Sensitivity increased when the objects appeared in a spatial relationship that made action between them possible relative to a spatial relationship that made action impossible. Context effects on object recognition might therefore operate at a level even subtler than background scene gist, in this case depending upon familiar spatial relations and potential for action.

Modern approaches investigating the effects of scene context on object recognition have shifted towards the use of photographic stimuli. The use of natural scenes is of theoretical importance to the degree that the perception of scene gist (and, in turn, the extraction of semantic information from a scene) depends upon global, low-level image statistics (see Oliva and Torralba, 2006, for a discussion of the possible role of global scene statistics on the perception of gist). Line drawings do not provide the same low-level visual information that natural scenes do, and the facilitation of object recognition may very well depend on that information. Neurological models of object recognition have proposed that low spatial frequency information serves to prime representations of scene category, which activates a set of candidate object representations (e.g., Bar, 2004). Such a model predicts facilitation when an object and context are semantically consistent, and impediment when they are inconsistent, which is exactly what the behavioral data appears to show. (However, see Loschky et al., 2007 for a discussion of the limitations of global scene statistic accounts of scene perception.)

Looking at the relationship between context and object recognition in the other direction, the presence of a salient, semantically consistent object in a scene has been shown to facilitate rapid scene categorization (Joubert, Rousselet, Fize, and Fabre-Thorpe, 2007). Although the authors originally proposed interacting, ultra-rapid object and scene recognition processes working in parallel, an alternative explanation depends entirely upon the bottom-up processing of low-level image statistics (Mack and Palmeri, 2010). Computational models of scene categorization suggest that the visual characteristics of a salient object greatly influence global scene statistics, making them resemble the scene category to which the object belongs. When the object and scene are semantically consistent, categorization is fast and accurate because the global scene statistics are highly typical of the object and scene's category. When they are

inconsistent, however, the scene statistics fall outside the typical distribution for either the object or the scene's categories, hampering recognition.

Recent research using photographic stimuli has addressed the specific question of object recognition in context more directly (Davenport and Potter, 2004; Davenport, 2007). Experimenters had subjects rapidly (80 ms stimulus duration) identify a salient foreground object that appeared on one of three possible backgrounds: a semantically consistent scene, a semantically inconsistent scene, or a white screen. All the stimuli came from a finite set of semantically associated object/scene pairs that were intermixed to create stimuli for the inconsistent condition. Subjects performed best without a scene background (possibly because the white background increased the object's salience and obviated the need for object segmentation), but still had greater reporting accuracy when objects appeared on a semantically consistent background than an inconsistent one. Once again, objects had a reciprocal effect on the scenes, increasing accuracy for reporting the background in the consistent condition, as well. Unfortunately, studies that rely on free reporting accuracy of objects in scenes do not adequately account for the issue of response bias. Especially in an experiment where the stimuli all come from a finite set of object/scene pairs, reporting accuracy can improve for completely non-perceptual reasons.

For example, say a subject sees a football player either in an inconsistent or blank background earlier on in the experiment. The subject may pick up on the fact that the stimuli come from a finite pool of pairs, or may simply be primed for football players in general. When that same subject sees a football field later on, regardless of the foreground object, that subject will be biased towards reporting "football player." As a result, this bias will increase accuracy in the consistent condition and decrease accuracy in the inconsistent condition, regardless of

whether the subject actually recognized the foreground object on the trial. If we are to conclude that semantic consistency truly modulates object recognition at the level of object perception, we must effectively account and control for response bias. An ideal paradigm would remove any strategic value for response bias altogether. Luckily, such a paradigm exists in the language literature.

Eliminating Bias: The Word Superiority Effect Paradigm

The word superiority effect has origins that reach as far back as the nineteenth century. Subjects viewing tachistoscopically presented letter strings could report more letters when they formed a word than when they did not (Cattell, 1886). Modern studies use a paradigm developed by Reicher (1969) and Wheeler (1970). Subjects perform a discrimination task between two rapidly presented letters, such as D and K. The letters can appear in a variety of contexts, such as letters that form words (e.g., WORD and WORK), non-word letter strings (e.g., RWOD and RWOK), or non-letter character strings (e.g., &&&D and &&&K). Importantly, the context is in no way informative or predictive of what letter appeared on any given trial. Either letter fits equally well into the context, yet discrimination accuracy is highest when the context forms a word with the target letter.

The Scene Superiority Effect

Some authors have used interactive activation models to account for the word superiority effect (e.g., McClelland and Rumelhart, 1981; Chase and Tallal, 1990). In such models, word recognition includes three levels: a feature level, a letter level, and a word level, each of which combine elements to form the level after it (i.e., features form letters, letters form words).

Activation from the stimulus feeds upward to the word level, but then candidate words feed activation back to the letter level, resulting in enhanced recognition for individual letters. An alternative account suggests that the word superiority effect results from word contexts constraining the set of possible letters to be recognized at the target location (Paap, Newsome, McDonald, and Schvaneveldt, 1982). Although resolving this linguistic debate goes beyond the scope of the current study, both accounts bear a passing resemblance to the processes of object recognition in scenes. Scene category information may facilitate object recognition by constraining the set of possible objects to recognize. Likewise, activation at the "scene level" could feed activation back to representations of any objects that do appear in the scene, making recognition faster and/or more accurate.

Thus, the current experiments seek to explore the degree to which the logic and methodology of the word superiority effect apply to object recognition in scenes. Instead of discriminating between two letters that fit equally well into the same word context, subjects will discriminate between two objects that can fit equally well inside the same scene context. "Fit," in the case of objects, will refer to the degree of semantic consistency with the context scene. If subjects show better (i.e., faster or more accurate) object recognition amidst semantically consistent scenes compared to semantically inconsistent ones, it would suggest that scene context truly influences the perceptual processes associated with object recognition rather than post-recognition processes like response bias. Although numerous studies have shown a positive effect of semantic consistency on object recognition, no study to date has yet to adequately separate the impact of context on object recognition from its influence on response bias (Henderson and Hollingworth, 1999).

GENERAL PARADIGM

All experiments were carried out on Apple eMac computers at a resolution of 1024 x 768 and refresh rate of 89 Hz. Subjects viewed the screen from a distance of approximately 56 cm through the aperture of a proprietary viewing hood that blocked the screen from alternative viewpoints.

The experimenter rapidly presented objects at central fixation surrounded by one of three image contexts: a semantically consistent scene, a semantically inconsistent scene, or a semantically meaningless scene (either an inverted or phase-scrambled scene). The scenes occupied only the central portion of the screen, leaving a medium gray border around them. Subjects had 3 seconds to respond in a forced-choice task paradigm with the object they believe appeared on that trial.

Parameters specific to each experiment are described in their respective sections.

EXPERIMENT 1

The first experiment served as an exploratory study to see if our stimuli and paradigm could elicit a semantic consistency effect. It does not follow the word superiority effect paradigm as described previously, but instead uses an 8-alternative forced choice recognition task.

Subjects

Nineteen undergraduate students from the University of Illinois participated in the study for course credit as part of the introductory psychology subject pool. An experimenter obtained informed consent from each subject and explained the instructions of the task as a test of visual perception. After they completed the task, the experimenter debriefed the subjects on the hypotheses and nature of the study. Debriefing included surveying subjects for any observations they made during the task that might be relevant to the researchers trying to interpret their data.

Stimuli

The context scenes consisted of 6 stock photographs of beaches and 6 stock photographs of offices. All scene images had a resolution of 800 x 600 pixels (subtending approximately 25.5˚ of visual angle horizontally and 19.125˚ vertically). The scenes could appear upright or inverted (to remove semantic content while maintaining low-level visual content), making four possible background categories: upright beach, upright office, inverted beach, and inverted office. Objects were also stock photographs taken from the Internet by two independent undergraduate research assistants. They were told to collect images of objects (appearing alone against white backgrounds) that they would expect to see at either a beach or in an office. Eight

of the objects were chosen for use in the experiment based on overlap in seeking out that object category (e.g., both assistants chose "computer" as an office item). The final experiment used images of the following 8 objects: an office chair, a desktop computer with LCD monitor, a copier, a laser printer, a beach ball, a beach chair, a toy sand pail and shovel, and a beach umbrella. The images were each cropped and scaled to fill a square, 200 x 200 pixel box (approximately 3.75˚ of visual angle), and contrast was reduced to 40% of the original in order to increase the difficulty of the object recognition task.

Design

The experiment was designed so that each object would appear on each category of scene context 10 times over the course of the experiment (8 objects x 4 categories x 10 repetitions = 320 trials total). The specific background for each trial was chosen randomly from the set of scene images.

Procedure

The experiment proceeded as follows (figure 1.1). First, the scene context appeared on screen for 300 ms with a medium gray box containing a fixation cross superimposed at the center, where the object would appear. A target object was then flashed on the screen for one refresh (approximately 11 ms) with a translucent layer of random white noise overlaid on top in order to increase the difficulty of the perception task. Finally, a visual mask of white noise appeared in the object's place for 300 ms before all stimuli disappeared from the screen. White noise for the object recognition task and the mask were generated separately on each trial by assigning a random RGB value to each pixel inside the box where the objects appeared.

After each trial, subjects reported which object they saw in an 8-alternative forced choice task. An array of the 8 possible objects appeared (randomly distributed on each trial) in a circle around central fixation. Subjects used the mouse to click on the object they believe they saw on that trial and received accuracy feedback. Subjects were encouraged to answer as accurately as possible, but a 3-second time limit encouraged them to guess even when they were unsure of the correct answer.

Experiment 1 began with a 64-trial staircasing procedure, QUEST (Watson & Pelli, 1983) that ensured uniform task difficulty across subjects. The task proceeded as described above, but the objects always appeared in the context of a phase-scrambled version of the scenes used in the final experiment. The QUEST algorithm operated on the opacity of the white noise overlaid on top of the target objects, increasing or decreasing their visibility. At the end of 64 trials, the QUEST algorithm set the opacity to a level at which the subject would perform at approximately 50% accuracy (chance = 12.5%). Following the staircasing procedure, 32 trials were used to check (after the experiment) subjects' accuracy to ensure they actually performed at approximately 50% accuracy at that visibility level. The rest of the experiment trials used the noise overlay opacity set by the QUEST algorithm.

Subjects then completed 320 trials of the experiment at their own pace with breaks every 64 trials. The paradigm remained the same, but now used upright and inverted scene images in the background. Twenty practice trials preceded both the QUEST and main experiment blocks.

Results & Discussion

This first experiment recorded only accuracy data from the subjects (figure 1.2). A 2 (consistency) x 2 (inversion) repeated-measures ANOVA revealed no significant main effect of

semantic consistency, (F[1,18] = 0.59, p = .45), or inversion of the scene (F[1,18] = 1.29, p = .27), nor did subjects show an interaction between consistency and inversion (F[1,18] = .01, p = .92). Initial analyses seemed to show no significant differences between the different conditions.

During debriefing, some subjects commented that they became increasingly unaware of the backgrounds, reporting a feeling of tunnel vision for the objects at fixation alone. It may be possible that any priming effects of the background require at least awareness of the scene category. Because subjects claimed their focus on the object alone narrowed as the experiment progressed, we analyzed data for the first and second halves of the experiment separately.

Splitting the data set in half into two epochs, subjects showed a marginally significant interaction between consistency and inversion in the first half (see figure 1.3) of the experiment (F[1,18] = 3.92, p= .06), but not in the second half (F[1,18] = .01, p = 0.92). The interaction appears to reflect a larger inversion effect in consistent scenes (r = .18) than in inconsistent scenes (r = -.06), such that inverting a consistent scene hurts performance, whereas inverting an inconsistent scene does little to accuracy. Such a finding would be consistent with the hypothesis that semantically consistent scenes aid object recognition, and possibly that semantically inconsistent scenes hurt recognition.

Although these data suggest a potential semantic consistency effect, we should not jump to conclusions just yet. First, analyzing the data in halves severely reduces the size of the data set. Second, the interaction was only marginal. Third, a 2 (consistency) x 2 (inversion) x 2 (epoch) repeated-measures ANOVA found that epoch did not significantly interact with consistency (F[1,18] = 0.292, p = 0.59), inversion (F[1,18] = 0.001, p = 0.98), or the interaction of the two (F[1,18] = 2.501, p = 0.13).

To address the variable awareness of the backgrounds that subjects reported, Experiment 2 introduced an intermittent probe task that tests background scene categorization. Such a task encourages subjects to devote attention to the background scenes throughout the the experiment. We also amended the task to more closely follow the word superiority effect paradigm, using only two objects.

EXPERIMENT 2

The second experiment addressed the issues raised by experiment 1 and switched to a 2-alternative forced choice task in order to more closely follow the word superiority effect paradigm. The task remained largely the same but with the addition of a background probe task.

Subjects

A new set of nineteen undergraduate students from the University of Illinois participated in the study for course credit as part of the introductory psychology subject pool. As before, an experimenter obtained informed consent from each subject and explained the task. Subjects again participated in a debriefing and survey after the experiment.

Stimuli

The experiment now only used 2 of the original 8 object images: the computer and the printer. The stimuli were chosen because subject accuracy for identifying those objects (computer, 40%; printer, 34%) in experiment 1 came closest to the target accuracy of the QUEST procedure (50, where chance was 12.5%). Subjects responded to a 2-AFC task with key-presses ("z" or "/") corresponding to each of the objects.

In an attempt to improve masking, visual masks were no longer made of random noise. The computer generated a random mosaic for each trial made up of 10x10 pixel tiles sampled randomly from both of the object images.

Rather than using offices and beaches, the current experiment now used cities and offices. The office scenes from the previous experiment were re-used, and 6 city scenes replaced

the beaches. The change was motivated by a desire to see if a consistent category effect would still appear in the data despite both scene categories belonging to a superordinate category of "man-made environments."

Procedure

Experiment 2 was identical to experiment 1 except for three changes. First, it used a 2-AFC task rather than 8-AFC. Second, a probe task encouraged subjects to pay attention to the backgrounds. On 1/8 of the trials, regardless of condition, a 2-AFC probe would ask subjects to report which category of scene appeared in the background (city or office). Answers were assigned randomly to either "z" or "/", the same response keys as the primary discrimination task, and subjects responded with a key-press. Probes were distributed randomly throughout the experiment to make them unpredictable. Finally, a floor effect in accuracy during pilot testing motivated slight changes in timing. Scene context appeared for 300 ms before the onset of the object, which appeared for two screen refreshes (approximately 22 ms), followed by the object mask for 300 ms. The next screen asked subjects which object they saw, displaying the two choices and their key mappings until a key-press response. At that point, the experiment gave accuracy feedback and moved on to the next trial. If subjects did not respond for 3 seconds, the trial "timed out" and was counted as incorrect. Response times were recorded as the time between the onset of the response screen and the moment the subject made a key press.

Results & Discussion

One subject's data were excluded from the analyses due to chance-level responding on the main discrimination task (48% accuracy overall). We analyzed accuracy, sensitivity, and

response time across subjects. Response time analyses excluded trials in which the subject responded incorrectly, the 3-second time limit elapsed, or RT fell outside of 3 standard deviations from the mean RT for a given subject. A repeated-measures, 2 (category) x 2 (inversion) ANOVA found no significant main effects of scene category ($F[1,17] = 0.045$, $p = .84$) or inversion ($F[1,17] = 0.274$, $p = 0.61$) on RT, nor was there a significant interaction ($F[1,17] = 3.263$, $p = .09$).

Overall background probe accuracy across subjects was 88.47%, and subjects reported general awareness of the backgrounds for the duration of the experiment. A repeated-measures, 2 (category) x 2 (inversion) ANOVA showed no main effect of scene category ($F[1,17] = 0.15$, $p = .70$) or inversion ($F[1,17] = 0.22$, $p = .64$) on background probe accuracy, suggesting subjects could recognize scenes well regardless of category or inversion. The data did not show evidence of a speed accuracy trade-off, as probe RT did not show any significant main effects of semantic consistency ($F[1,17] = 3.151$, $p = 0.09$) or inversion ($F[1,17] = 1.794$, $p = 0.20$) or an interaction ($F[1,17] = 0.705$, $p = 0.41$) Unfortunately, this means our inversion condition may not truly serve as a semantically neutral baseline. Still, the data show a significant interaction of category and inversion ($F[1,17] = 10.12$, $p = .005$). Inversion seemed to affect subjects' ability to correctly identify offices (within-subject $t[17] = -3.01$, $p = .008$), but not cities (within-subject $t[17] = 1.44$, $p = .17$).

Strangely, subjects were more accurate for identifying inverted offices than upright offices. This may have been due to a significant response bias against responding "office" in the upright condition; a bias appeared in both C (mean: .306; $t[17] = 2.52$, $p = .02$) and B'' (mean: 0.383; $t[17] = 2.80$, $p = .01$), which differed significantly from response bias in their respective inverted-scene conditions ($t[17] = 2.4$, $p = .03$; $t[17] = 2.12$, $p = .05$ for C and B'', respectively),

which showed no such bias (t[17] = -0.19, p = .85; t[17] = -0.05, p = .95 for C and B'', respectively). The reason for such a bias is open to speculation. Perhaps subjects overcompensated for the fact that the target objects came from offices, and so felt extra wary about confusing the scene content with the target object content. The current data cannot speak directly to exactly why subjects behaved in this way.

Raw accuracy for the object discrimination task showed neither a main effect of category (F[1,17] = 2.65, p = .12) or inversion (F[1,17] = 1.41, p = .25), nor any significant interaction between the variables (F[1,17] = 1.77, p = .20). Hit rate and false alarm rate for the sensitivity calculations were determined by treating one object as the "signal" and the other as "noise." Hits were trials when the subject responded "computer" to the presentation of a computer, while false alarms were trials when the subject responded "computer" to a printer. In cases where performance was at ceiling (i.e., hit rate = 1, false alarm rate = 0), values were adjusted according to a standard correction (Stanislaw and Todorov, 1999) that reflects missing or false alarming to the equivalent of ½ of a trial (in this experiment, 1/160). Two subjects' false alarm rates in the "city" background condition required such a correction. Biederman et al. (1982) and Hollingworth and Henderson (1998) used different methods for calculating sensitivity (d' and A', respectively), so we applied both measures to the data for the sake of comparison.

Sensitivity was calculated both as d' (the measure used by Biederman et al., 1982) and A' (the measure used by Hollingworth and Henderson, 1998) for the sake of comparison to previous research. Sensitivity measures were again analyzed with a repeated-measures, 2 (category) x 2 (inversion) ANOVA. Analysis of d' showed no significant main effects (category: F[1,17] = 0.460, p = .51; inversion F[1,17] = 0.002, p = .96) or interaction (F[1,17] = 0.778, p = .39). A', on the other hand, yielded a significant main effect of category (F[1,17] = 4.638, p = .05);

subjects could better discriminate the office objects on an office background than on a city background. Although inversion showed no main effect ($F[1,17] = 2.76$, $p = .11$) or interaction ($F[1,17] = 2.20$, $p = .16$) with scene category, subjects' ability to identify the inverted scenes may suggest the upright and inverted conditions were functionally equivalent in this particular task.

The data here point to a semantic facilitation effect between objects and scenes; unlike previous studies, however, the facilitation cannot be explained as a mere artifact of response bias. Both objects had the same degree of semantic association with offices and appeared with equal likelihood on any given trial, so bias towards responding with one or the other would not translate into an apparent recognition benefit. Combined with the results from experiment 1, it would appear that object recognition benefits from the presence of a semantically consistent scene context, and not just at the level of response bias. Under these conditions, semantically consistent backgrounds aided perception of the target objects.

Still, we should address some limitations of the current design. First, inversion did not completely remove semantic information about the scenes, so we do not have a semantically devoid but visually equivalent baseline measure of object recognition. We do not know if semantic consistency actually facilitates object recognition or if semantic inconsistency hurts it relative to baseline. Second, the office backgrounds sometimes contained computers and printers (though did not feature them prominently) whereas the cities did not. The facilitation effect might therefore come from visual priming of the target objects themselves rather than priming of higher-order semantic categories.

Finally, reports from the post-experiment debriefing suggest that not all subjects actually engaged in true object recognition. Some subjects reported fixating only parts of the image and

using diagnostic features to discriminate between the objects. For example, one subject reported looking for a patch of blue in the top left corner of the object box because it signaled the presentation of the computer screen. Experiments 3 & 4 aim to address these limitations through modifications to the object stimuli and their presentation.

EXPERIMENT 3

We endeavored to address the limitations of the previous experiment with slight changes to the design. To test for generalizability of the scene superiority effect, we had subjects perform object discriminations for two semantic categories instead of just one. Next, we added a baseline background condition, which used 100% phase-scrambled versions of our scene stimuli to provide a truly semantically empty background.

In order to gain more control over the experimental stimuli, we used a 3D modeling program, Google SktechUp, to create both scene and object images. Doing so provided two distinct advantages. First, we could ensure the target objects never appeared in the backgrounds, so any facilitation we find would have to result from higher-order semantic priming of scene category. Second, we could generate more stimuli by creating images of the objects and scenes from multiple angles. Having more stimuli (combined with random jitter of object placement) should reduce subjects' ability to fixate a location on the image and perform the object discrimination on the basis of a single, diagnostic feature alone.

Unfortunately, the new stimuli provided new challenges. Manipulating visibility through stimulus timing appeared to yield either ceiling or floor effects in subjects' data. Presentation durations are quantized by the computer screen's refresh rate, and while 2 refreshes often proved too short for subjects to see the objects, performance often reached ceiling in as few as 3 refreshes. As a result, the current experiment was designed to be equally easy for all subjects to perform accurately in the hope of finding an effect in response time. The objects appeared on screen simultaneously and remained on screen until the subject responded with a key press.

Subjects

Twenty-nine University of Illinois undergraduate students participated in the study for course credit as part of the introductory psychology subject pool. Once again, an experimenter explained the task when subjects came to the lab and engaged subjects in a debriefing session afterwards.

Stimuli

Experiments 3 and 4 used scenes and objects generated in Google SketchUp, a free 3D modeling program (available at http://sketchup.google.com). A cel-shading graphic effect made images generated with the program appear similar to cartoon drawings of scenes and objects. Models came from the 3D Warehouse, an online database of user-submitted models created with SketchUp. Scene categories and objects were chosen based on the availability and quality of user-submitted models. Scenes were taken from the top user-rated search hits for "kitchen" and "bedroom."

Each scene category had 2 corresponding objects in the experiment. A teddy bear and an alarm clock were chosen as "bedroom items," while a stand mixer and a toaster served as "kitchen items." The scenes and objects were chosen carefully so that target objects never appeared in the scene.

Objects were presented in a manner to allow easy localization while preventing subjects from fixating diagnostic locations and features on every trial. Objects were superimposed over the backgrounds near central fixation during the experiment with random jitter that shifted the object 0-100 pixels (approximately 0-3.2˚) horizontally and vertically. Unlike the previous

experiments, the objects appeared as if free-floating in the scene rather than inside a gray box within the image.

Each scene category used 4 room models viewed from 3 angles each, making 12 possible scene images per category. Similarly, each object had 3 viewing angles associated with it. Phase-scrambled versions of the scenes served as semantically meaningless control backgrounds.

Design

The experiment was split into four blocks: two within-category discrimination blocks (teddy bear vs. alarm clock; toaster vs. mixer) and two between-category discrimination blocks (i.e., one bedroom object vs. one kitchen object). Target objects in the between-category discrimination were chosen randomly, so each subject only performed two of the four possible between-category discriminations. Block order was counterbalanced between subjects. Scene category probes were distributed randomly within each block to make them unpredictable and ensure subjects maintained awareness of the backgrounds throughout the experiment.

The objects could appear on one of three types of background scene with equal likelihood: bedrooms, kitchens, or a phase scrambled version of an image from the stimulus set. Phase scrambled backgrounds were sampled randomly from the full set of both phase-scrambled bedroom and kitchen scenes. Each block had subjects perform the object discrimination task on each of the backgrounds twice, once with one target object and once with the other, making a total of 144 trials per block. Presentation order was randomized within blocks.

Rather than analyze the data in terms of background, trials were categorized under conditions of semantic consistency. The “semantically consistent” condition consisted of trials

on which the target object and scene background belonged to the same semantic category, while the "semantically inconsistent" condition was just the opposite.

Procedure

Each block began by showing subjects which two objects they would discriminate for the following trials, and to which keys the objects were mapped. Twenty practice trials then allowed subjects to grow accustomed to the key mapping. Each trial began with a central fixation cross that appeared for 500 ms. The scene and object appeared together and remained on-screen until the subject responded with a key press. If the subject did not respond within 3 seconds, the trial would time out and was counted as an incorrect response. Response time consisted of the interval between the onset of the scene and object stimulus and the moment of the subject's key press. The experiment then provided accuracy feedback before the next trial began. As in the previous experiment, a random probe appeared on approximately 1/6 of the trials, asking subjects to report (with a key-press) the category of the scene background with the same time limit and feedback as the recognition task.

Results & Discussion

Data were analyzed using a 2 (semantic consistency) x 2 (discrimination type) repeated-measures ANOVA. Subject data was converted into difference scores that related each subject's performance on the two semantic consistency conditions to the scrambled baseline condition (i.e., consistent – scrambled, inconsistent – scrambled). Thus, the data represent each subject's deviations from baseline performance (object recognition without semantic context). A positive

RT difference score indicated a slowing of responses, while a positive accuracy difference score indicated a benefit.

Outliers were removed from the RT as they were in experiment 2; analyses excluded trials on which the subject responded incorrectly, timed out, or had a response time that fell outside 3 standard deviations of the subject's mean RT. The data showed no main effects of semantic consistency ($F[1,28] = 2.53$, $p = .12$) or discrimination type ($F[1,28] = .08$, $p = .79$), nor was there an interaction between the variables ($F[1,28] = .01$, $p = .91$). Accuracy was at ceiling as we expected (mean overall accuracy = 97.8%), but surprisingly, subjects still showed a significant effect of semantic consistency ($F[1,28] = 5.42$, $p = .03$). Semantically consistent scenes offered a slight boost to accuracy relative to baseline (an increase of 0.3% in accuracy across discrimination types), while semantically inconsistent scenes led to a slight decrement (a decrease in accuracy of 0.4% across discrimination types).

Experiment 3 showed continued to show a (small but reliable) benefit to object recognition from semantically consistent scenes. Because the extent of the effect could have been obscured by a ceiling effect, experiment 4 aimed to reduce subject accuracy by increasing the difficulty of the task. Additionally, because experiments 2 and 3 seemed to suggest semantically consistent context scenes facilitate within-category object discrimination, we wanted to see if context could go so far as to facilitate discrimination between two exemplars of the same object (e.g., discriminating one alarm clock from another).

Although the experiments thus far suggest enhancement occurred at the perceptual rather than response stage of the object recognition task, the nature of the perceptual enhancement is unclear. There are at least two possibilities. The scene context may prime representations of object categories, allowing discrimination between two different categories of objects, but not

different exemplar of the same category of object. On the other hand, a semantically consistent context may actually make the target object easier to perceive more generally, improving performance regardless of the type of discrimination. Experiment 4 addresses these alternatives.

EXPERIMENT 4

Experiment 4 used many of the same stimuli as experiment 3, but made several changes to the paradigm. First, two of the blocks asked subjects to perform within-category discriminations as before (teddy bear vs. alarm clock; toaster vs. mixer, herein "object discrimination"), but now the other two blocks had subjects perform discriminations between object exemplars (two types of alarm clock or two types of stand mixer, herein "exemplar discrimination"). If semantically consistent scenes truly enhance overall perception of the target object, we should see benefits in both conditions.

More importantly, the experiment returned to the rapid-presentation paradigm of experiments 1 and 2. Visibility was manipulated by adjusting the target object's opacity, but unlike the previous experiments, the current study did not begin with a staircasing procedure for each subject. To keep the experiment short, we ran four subjects in a version of the experiment that used QUEST to determine an appropriate opacity level for each type of discrimination to produce approximately 75% accuracy. Those data then determined the visibility settings for the rest of the subjects.

Subjects

Twenty-two University of Illinois undergraduate students participated in the study for course credit as part of the introductory psychology subject pool. Once again, an experimenter explained the task at the beginning of the experiment and engaged subjects in a debriefing session afterwards.

Stimuli

Objects used for the exemplar discriminations were chosen on the basis of model availability in the SketchUp 3D Warehouse.

Like experiment 3, the objects appeared near the middle of the screen with random jitter 0-100 pixels in the vertical and horizontal directions from center fixation. The objects were superimposed directly on top of the image, giving the appearance that they were free-floating near the center of the screen.

In order to bring subject accuracy down from ceiling levels, object opacity was adjusted for each block. To determine appropriate opacity levels, four subjects completed the experiment as described below, but a QUEST staircasing algorithm determined the percent opacity level at which each subject would perform with 75% accuracy for each block. The final experiment used the average of the between-subjects mean and median opacity for each block, resulting in the following opacity levels: alarm clock vs teddy bear, 58%; mixer vs toaster, 53%; alarm clocks A vs B, 66%; mixers A vs B, 78%.

Object masks were generated for each block by overlapping the six possible target object images set to 16% opacity while keeping any black line-defined edges at 100% opacity. The result resembled overlapping outline drawings filled with translucent color.

Design

The experimental design was identical to experiment 3, with the exception of the discrimination types. Subjects completed four blocks (counterbalanced between subjects), each of which had them perform one of the four possible discriminations of interest: two bedroom items, two kitchen items, two alarm clocks, and two mixers.

Procedure

Like the previous experiment, each block began with instructions indicating the discrimination subjects would make and the response key mapping. After twenty practice trials, subjects took part in the actual task. A fixation cross appeared on screen for 500 ms, followed by the presentation of the scene and target object together for 2 screen refreshes (approximately 22ms). The objects were masked for 22 ms, then a response screen asked subjects which object they saw, presenting the two options and their key mappings. The screen disappeared when the subject made a response or "timed out" after 3 seconds had elapsed. Response time was recorded as the time from offset of the scene and object mask until the moment subjects pressed a response key. The computer provided accuracy feedback then began the next trial.

Results & Discussion

Two subjects were excluded from the analyses because their accuracy on the background probe was near chance (48% and 50%, respectively). Although we have not directly addressed whether conscious awareness of the backgrounds is truly necessary for them to affect object recognition, the subjects' relatively poor performance suggests that they at least approached the task differently from other subjects or possibly misunderstood the directions.

Like experiment 2, the data were initially analyzed as deviations from baseline performance using a 2 (discrimination type) x 2 (consistency) repeated-measures ANOVA. Analyzing the data relative to each subject's baseline performance for each type of discrimination was particularly important for this experiment if we are to compare the different discrimination types directly. By subtracting out baseline performance, we hoped to factor out

the differences between the discrimination conditions unrelated to the effects of background semantics.

A significant main effect of discrimination type was found in discrimination accuracy ($F[1,19] = 5.09$, $p = .03$), and d' ($F[1,19] = 7.06$, $p = .01$). The effect did not quite reach significance for A' ($F[1,19] = 3.82$, $p = .07$). In general, it appeared that discrimination between two different objects was more affected by the backgrounds than discrimination between two exemplars of the same object. This result, however, may be an artificat of the different opacity levels in the exemplar-discrimination condition (mean: 72% opacity) and the object-discrimination condition (mean: 55.5%). While this was meant to make the two discriminations equally difficult, it also introduces the possibility that low-level visual features of the background interfered with perception of the target object.

The nature of the analysis also confounded effects on the semantically consistent/inconsistent conditions with effects on baseline performance alone. Therefore, we analyzed the data for the semantic background conditions separately from the baselines. All significant effects and interactions disappeared when the data for the consistent and inconsistent conditions were analyzed alone rather than subtracted from the baseline, scrambled-scene condition. This finding pointed to an effect on baseline performance, which we tested with a within-subject t-test. Accuracy and sensitivity showed no significant differences in the baseline scrambled-background conditions (accuracy: $t[19] = 0.522$, $p = .61$; d': $t[19] = 1.331$, $p = .20$)

Baseline RT, however, was significantly different for the two discrimination conditions ($t[19] = 2.07$, $p = .053$); when collapsed across the different scene conditions, RT was significantly longer for discriminating two different objects than for two exemplars of the same object ($t[19] = 2.22$, $p = .04$). Collectively, these results suggest a speed-accuracy trade-off.

Although the scene context seemed to have a bigger effect on discriminating two different objects than two exemplars of the same object, subjects were also taking longer to respond in the different-objects condition. The current data do not disentangle the amount of processing devoted to the stimulus from the effect of semantic consistency.

Another—potentially related—explanation appears in the probe accuracy and sensitivity data. Subjects demonstrated a marginally significant main effect of discrimination type on accuracy ($F[1,22] = 3.94$, $p = .06$) and a significant main effect on d' ($F[1,22] = 4.25$, $p = 0.053$). People were more accurate at reporting the backgrounds in the object-discrimination condition than the exemplar-discrimination condition. During debriefing, many subjects reported greater difficulty with discriminating the exemplars of the same object than discriminating two different objects, though overall accuracy in both conditions was roughly equal ($t[19] = 0.009$, $p = .99$). Therefore, they likely devoted all their attention to the discrimination task in the same-object exemplars condition, perhaps to the exclusion of processing the scene background. Responses may have been faster in that condition because subjects focused entirely on the discrimination task without always maintaining a representation of the scene in working memory, as well.

If we assume subjects were consciously processing the backgrounds while discriminating objects but not exemplars (as suggested by the probe accuracy and d' results above), then we might assume that variance in the RT data is connected to processing the backgrounds in the former condition but not the latter. Furthermore, we would expect the amount of processing the background receives to be positively correlated with the size of the consistency effect for a subject; more processing should improve performance in the consistent condition and hurt performance in the inconsistent condition, widening the gap between the two. Therefore, we should see a correlation between the size of the consistency effect and RT in the object-

discrimination condition, but not the exemplar-discrimination condition. Consistent with this hypothesis, RT is correlated with the size of the consistency effect for both d' ($r = 0.44$, $p = .054$) and accuracy ($r = 0.46$, $p = .039$) in the object-discrimination condition, but not in the exemplar-discrimination condition ($r = 0.13$, $p = 0.57$; $r = 0.24$, $p = .31$ for d' and accuracy, respectively).

Thus, experiment 4 did not find a clear-cut effect of semantic consistency between the background and target objects; however, we may have evidence that the degree of processing (and, in turn, awareness) of the backgrounds modulates the semantic consistency effect. Future research can confirm this experimentally. Other follow-up work can further refine the paradigm and bolster our findings. First, because the objects were superimposed over the scene, they sometimes made violations of physical support, position, and size (Biederman et al., 1982). The cost that such violations likely had on object recognition may have swamped out or interacted with the effects of scene semantics in experiment 4. Another refinement might be to increase the number of object images in the experiment to test how well the effect of scene semantics generalizes and affects perception of novel stimuli.

GENERAL DISCUSSION

Four experiments provide some evidence for a benefit to object recognition from semantically consistent scenes. Experiment 1 showed a modest benefit to object recognition, but primarily early in the study, before attention presumably narrowed onto the object and ignored the context scene. Experiment 2 provided the strongest evidence for the scene superiority effect we sought at the beginning of this study, showing an increase in sensitivity when discriminating two office-related objects in the context of an office scene. Experiment 3 demonstrated an accuracy benefit of semantic scene contexts to a variety of object discriminations. Finally, experiment 4 suggests that conscious awareness of scene gist is necessary for it to influence object recognition.

With respect to object recognition in scenes, this study provides some support to accounts of object recognition that propose an interactive relationship with scene categorization. Unlike past studies, experiment 2 demonstrated a perceptual enhancement to object recognition that cannot have resulted from response bias alone. That raises the question, however, of why this experiment would find an effect of scene context on sensitivity when previous experiments that controlled for bias (Hollingworth and Henderson, 1998) did not.

Simple differences in timing may be enough to explain the disparity. The objects appeared for short, 22 ms durations with reduced visibility in this experiment compared to a 200-ms, clear view of the image in others. Furthermore, the objects in this experiment appeared inside a box that separated it from the rest of the scene, which itself appeared alone at the beginning of each trial. This separation may have allowed subjects to carry out the scene categorization and object recognition processes more discretely than they would have if the

object had actually been truly part of the scene. Rather than carry out a single recognition process that integrates information between the object and scene in parallel, subjects may have completed the scene categorization, which then had time to prime the relevant object set before object recognition. If so, then scene information might influence object recognition only when the scene information is already encoded and not when both must be encoded simultaneously. To the extent that previously encoded scenes influence object recognition, we might expect semantic consistency effects for new objects in highly familiar scenes but not for new objects in novel scenes, a testable hypothesis.

There is reason to believe our experiments simulate the natural processes of object recognition in context. It has been long established that scene categorization occurs extremely rapidly, on the order of around 100 ms (Potter, 1975). Semantic priming from the scene category can occur even without local object information; people can extract scene gist information from band-pass filtered images that contain only low spatial frequency information (Schyns and Oliva, 1994). Importantly, such a feat is only possible with natural scene photographs, which have a diverse spatial frequency spectrum—line drawings are nothing but high spatial frequency information. Therein lies what may be the critical difference between our experiment and those of Hollingworth and Henderson (1998). Scene gist and semantic priming simply may not make a difference to object recognition in 200 ms with line drawings. Experiment 2, on the other hand, gave subjects ample time to extract scene gist and have that information influence the object recognition process. Maximizing the effect of scene context on object recognition may require natural scene and object photographs, with time to process the scene gist before attempting object recognition.

For the sake of comparison, the stimuli in experiments 3 and 4 came closest to line drawings, as the models used a cel-shaded graphics effect to give the appearance of a cartoon scene. Notably, experiment 3 found a semantic consistency effect on accuracy, but subjects were given as much time to respond as they needed, and RTs averaged around 630 ms. Experiment 4 appears more consistent with Hollingworth and Henderson (1998), both in terms of timing and results; semantic consistency between the target object and background scene had no effect on performance. Even when the between-object data were analyzed alone, they failed to show a significant effect of context. On the one hand, the failure to replicate the findings of experiment 2 with different stimuli may be due to effects of task switching associated with the multiple discriminations in experiment 4, and the physical violations created by the superimposed objects. On the other hand, the relatively sparse visual information contained in the cel-shaded scenes compared to natural scene photographs may have been processed more slowly or provided less information than would be needed to affect object recognition. (The natural scenes in experiment 2 were given a 500 ms head start that experiment 4 did not offer the cel-shaded scenes.)

Comparison to the Word Superiority Effect

The parallels between the Word Superiority Effect paradigm and our studies require more discussion. Although our paradigm used similar logic, having subjects discriminate between two stimuli that could both fit into a scene context, the constraints and degree of association between the target and context differ for words and scenes. For example, the words in the Word Superiority paradigm are what linguists call "minimal pairs." The target letters are both necessary and sufficient for determining the meaning of the word in which they appear; the same

cannot be said of our objects. In fact, no single object is perfectly diagnostic of scene category, as arguably no objects are both necessary and sufficient to determining scene category the same way the letters determine word meaning.

Furthermore, word context constrains the set of possible letters at the target location much more than a scene constrains the set of objects that can appear in it. Few objects can exist in only one type of scene, and any given scene category could contain myriad object combinations. Perhaps to maximize the effect of scene context on object recognition, we need to find a minimal pair in scenes—two scene categories that are differentiated by two (or at least very few) objects. If the scene superiority effect truly exists and functions like the word superiority effect, the diagnosticity of the objects should modulate the effect size. Cars and trucks on a highway and stoves and sinks in a kitchen might yield a more impressive effect of context on discrimination than the stimuli we used here.

Broader Implications

Our study also raises issues of relevance to the broader scene perception and attention literature. Many studies suggest that contextual priming occurs automatically without focused attention or awareness. People can perform rapid object detection in a rapid serial visual presentation of scenes (Intraub, 1981), even when attention is engaged in a demanding primary task (Li, VanRullen, Koch, and Perona, 2002). People also can learn to associate complex properties of global contexts with targets (contextual cueing), and the context facilitates the deployment of spatial attention and object recognition; subjects locate and identify targets faster after learning to associate certain contexts with spatial and identity information. Learning and

accessing these contingencies appears to occur implicitly without conscious awareness (Chun and Jiang, 1998).

One challenge to these ideas of automatic processing came from a study that had subjects detect animals and vehicles in an RSVP stream (Evans and Treisman, 2005). Subjects showed an attentional blink (taken as evidence of binding features into a coherent object representation) when they had to identify the animals or vehicles they saw, but not when they simply detected the objects' presence. While object detection may not require focused attention, encoding and accessing more detailed information—such as location and identity, as in the current study—for explicit recall does.

Our results further challenge the claim that scene context automatically and implicitly influences object recognition. Subjects' awareness of the scene context appeared to be correlated with the effect it had on object recognition. Unlike the previous studies supporting automatic context processing, our experiments did not require subjects to search for or localize their targets, nor was there any reliable correlation between the objects and certain backgrounds. The current study strips down the subject's task solely to object recognition that simply occurs in the environment of a scene. It would appear that aspects of the scene representation that potentially drive the effect on object recognition, such as semantic priming of scene category and the activation of associated objects, require attention and explicit processing of context.

Oftentimes, discussions of implicit and explicit information processing in cognitive science bring up theories about "what" and "where" visual pathways in the brain. Although their exact purposes are debated, the ventral pathway appears to be associated with explicit (i.e., consciously accessible) processes related to identifying what an object in the visual field is, whereas the dorsal pathway carries out implicit processes determining spatial relationships

between objects and how to act on them (Ungerleider and Haxby, 1994; Goodale and Milner, 1992). The importance of explicit processing in our study may be consistent with this theoretical framework. Subjects need not process any spatial information in our experiments beyond localizing the object to the center of the screen. The nature of the task largely obviates the need for the implicit functions of the dorsal pathway. Instead, information processing relevant to the task occurs largely in the explicit ventral pathway. Thus, our findings depended upon modulation of consciously accessible processes, such as awareness of context scene semantics,

Conclusions

Our experiment provides the first behavioral evidence for a semantic consistency effect of context on object recognition that cannot be attributed to response biases. In addition, the data suggest conscious processing of scene context is necessary for accessing aspects of scene representations that aid in object recognition. Our paradigm provides an excellent tool for investigating the interaction between object recognition and scene processing without the biases inherent in other priming measures.

FIGURES

Figure 1.1. Experiment 1 paradigm schematic

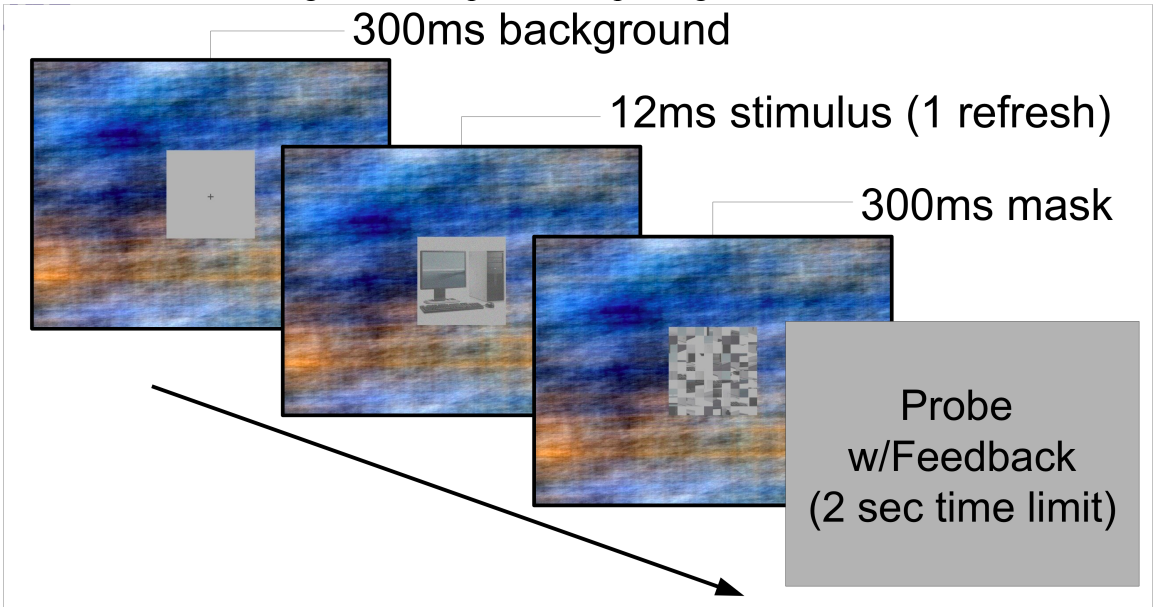


Figure 1.2 accuracy data (full)

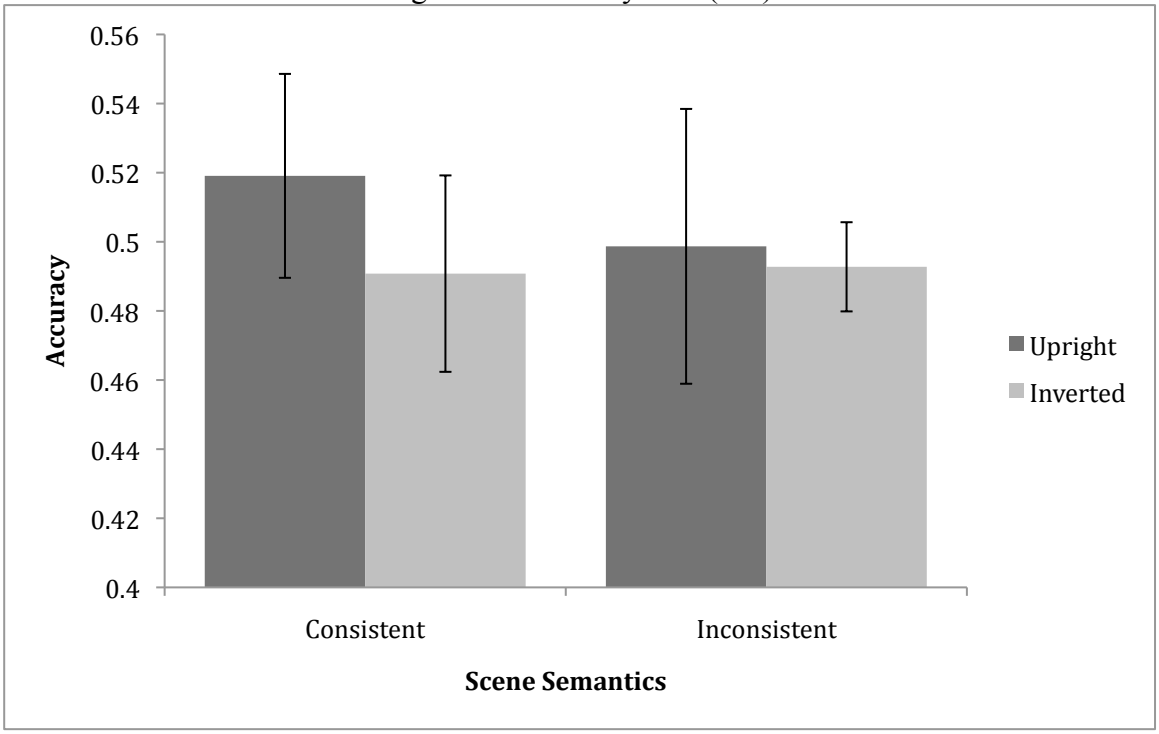


Figure 1.3 accuracy data (first half of experiment only)

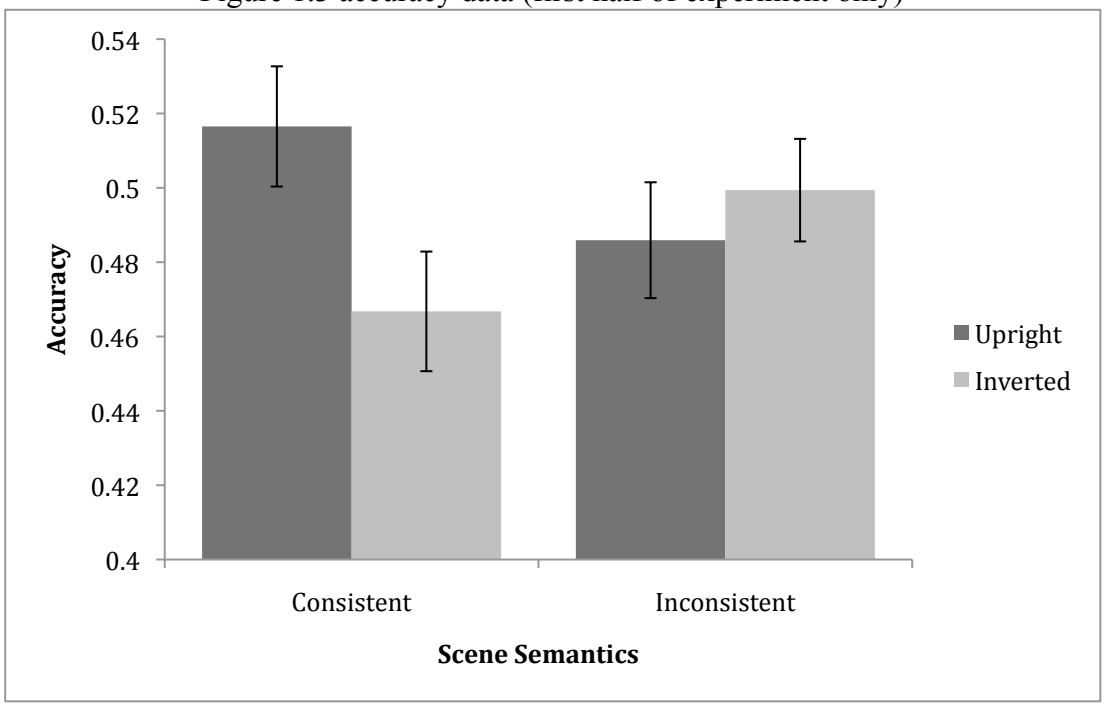


Figure 2.1. Experiment 2 probe Accuracy

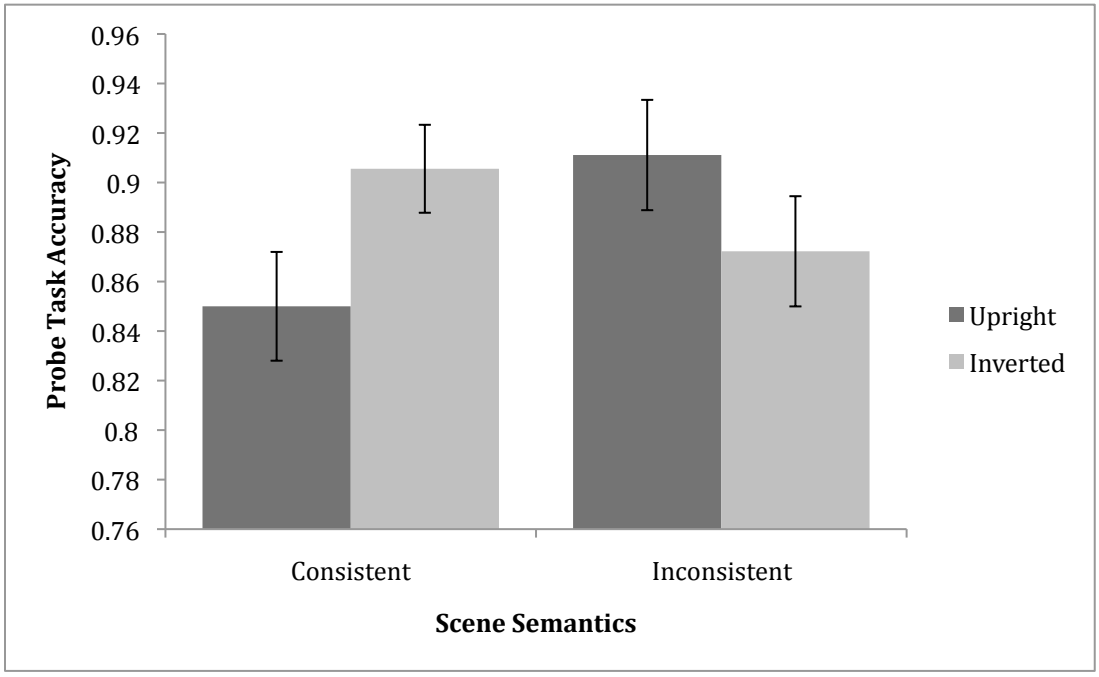


Figure 2.2. Experiment 2 A' data

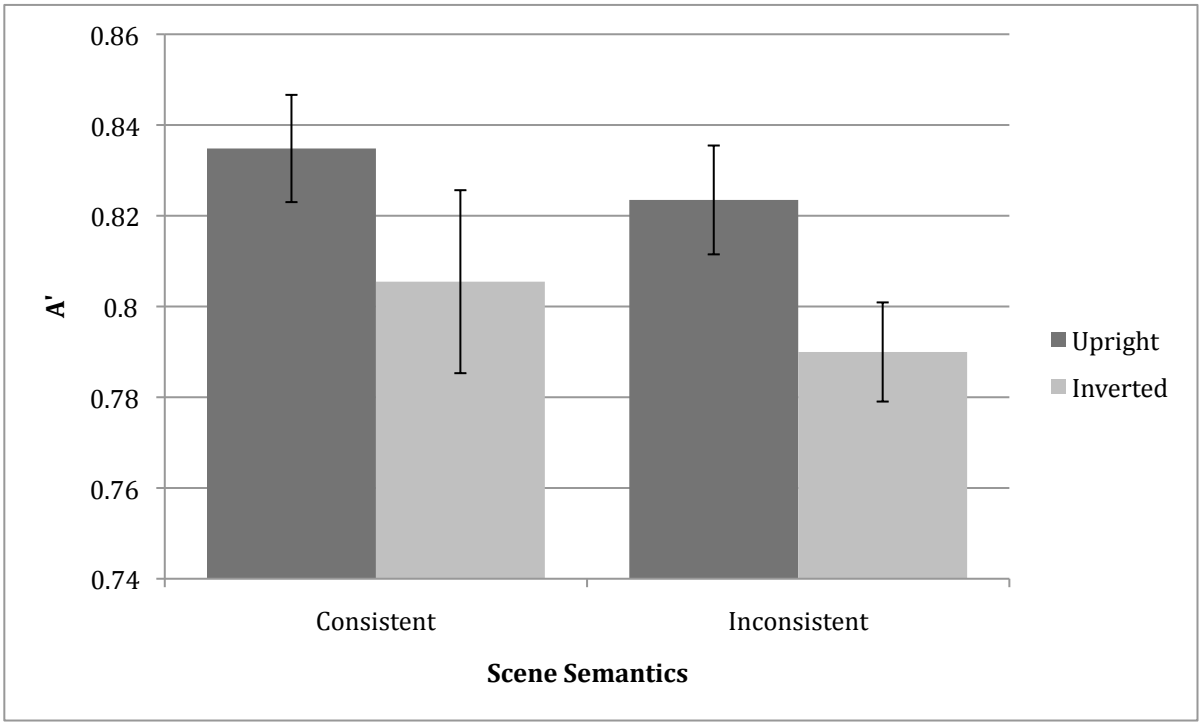


Figure 3.1. Paradigm schematic

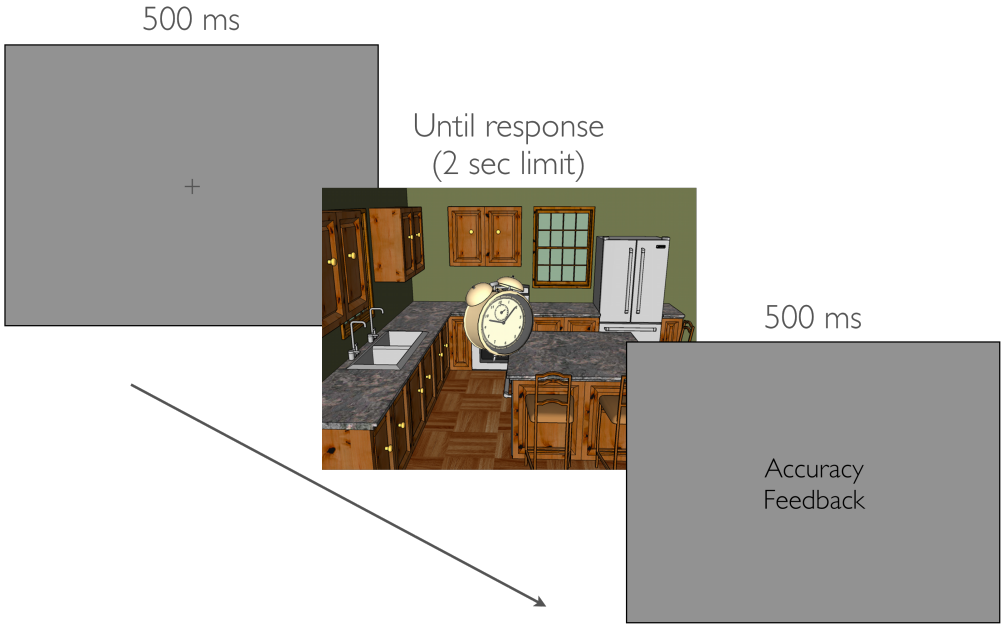


Figure 3.2. Experiment 3 Deviation from accuracy in scrambled background condition, separated by scene semantics and discrimination type

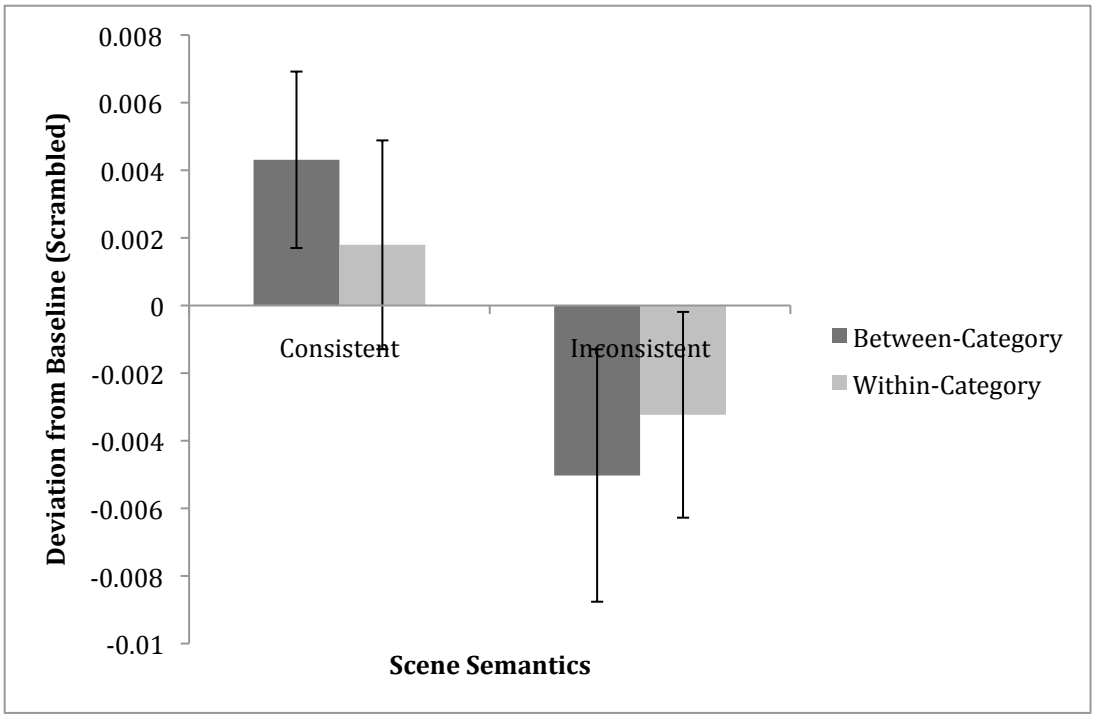


Figure 4.1. Experiment 4 paradigm schematic

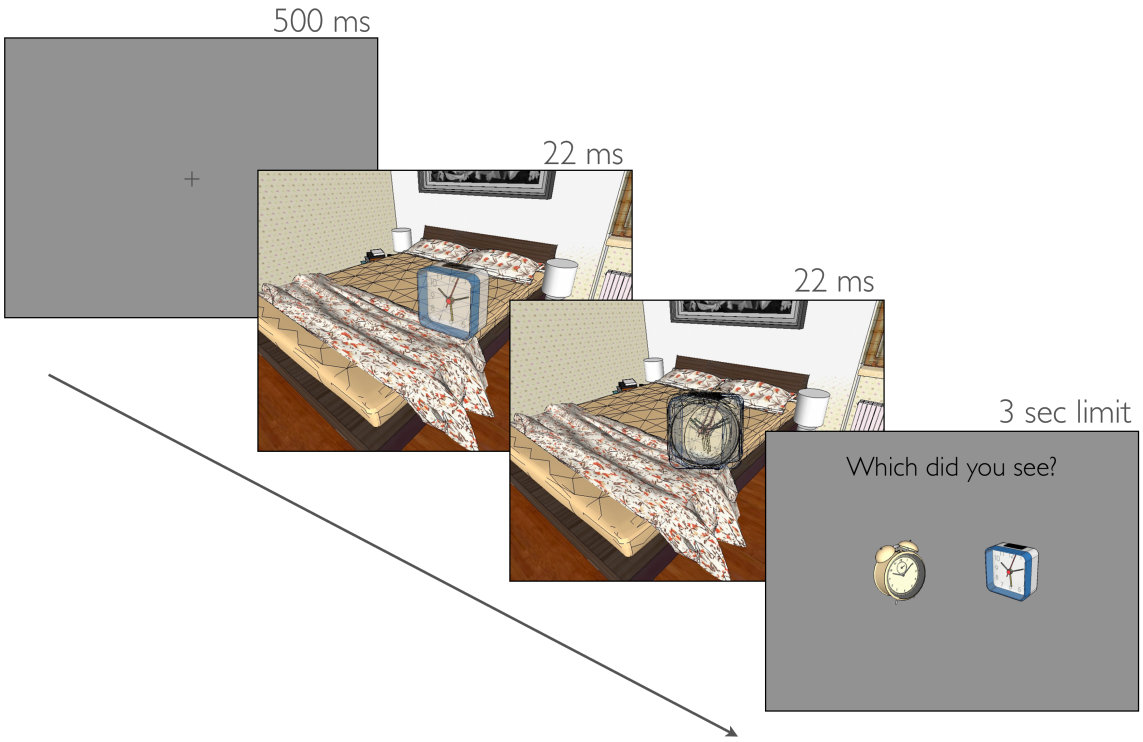


Figure 4.2. Experiment 4 Accuracy deviations by discrimination type and consistency

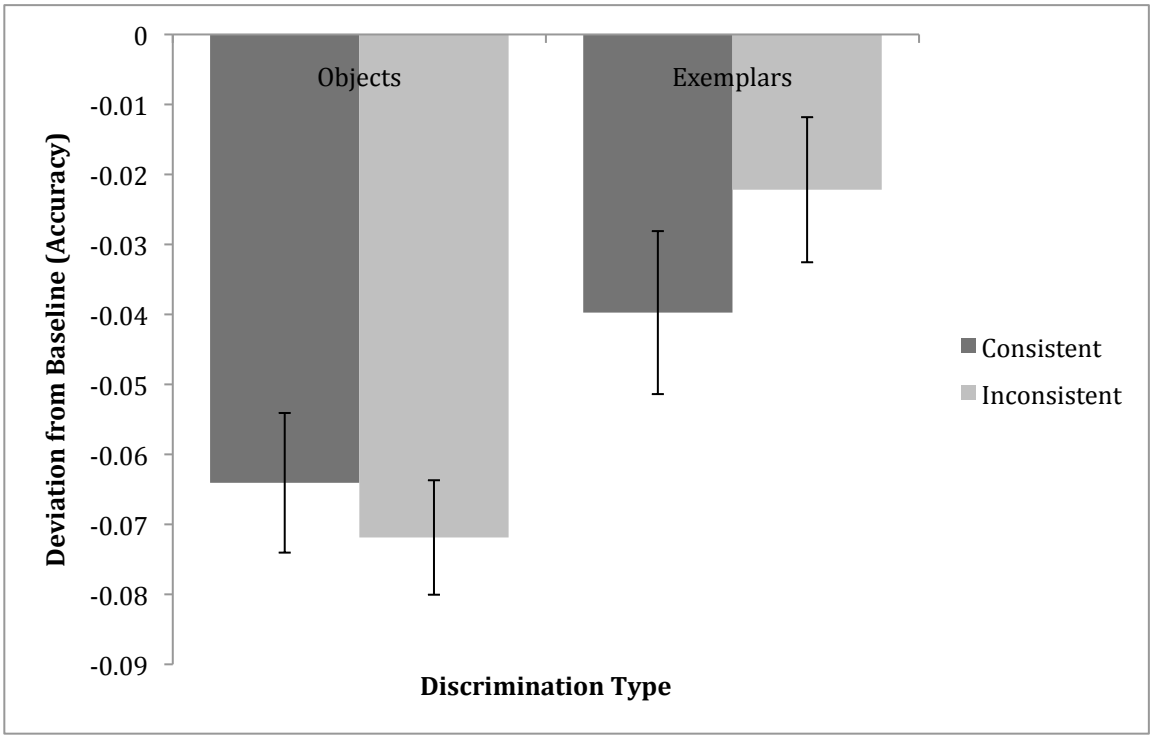


Figure 4.3. Experiment 4 raw accuracy

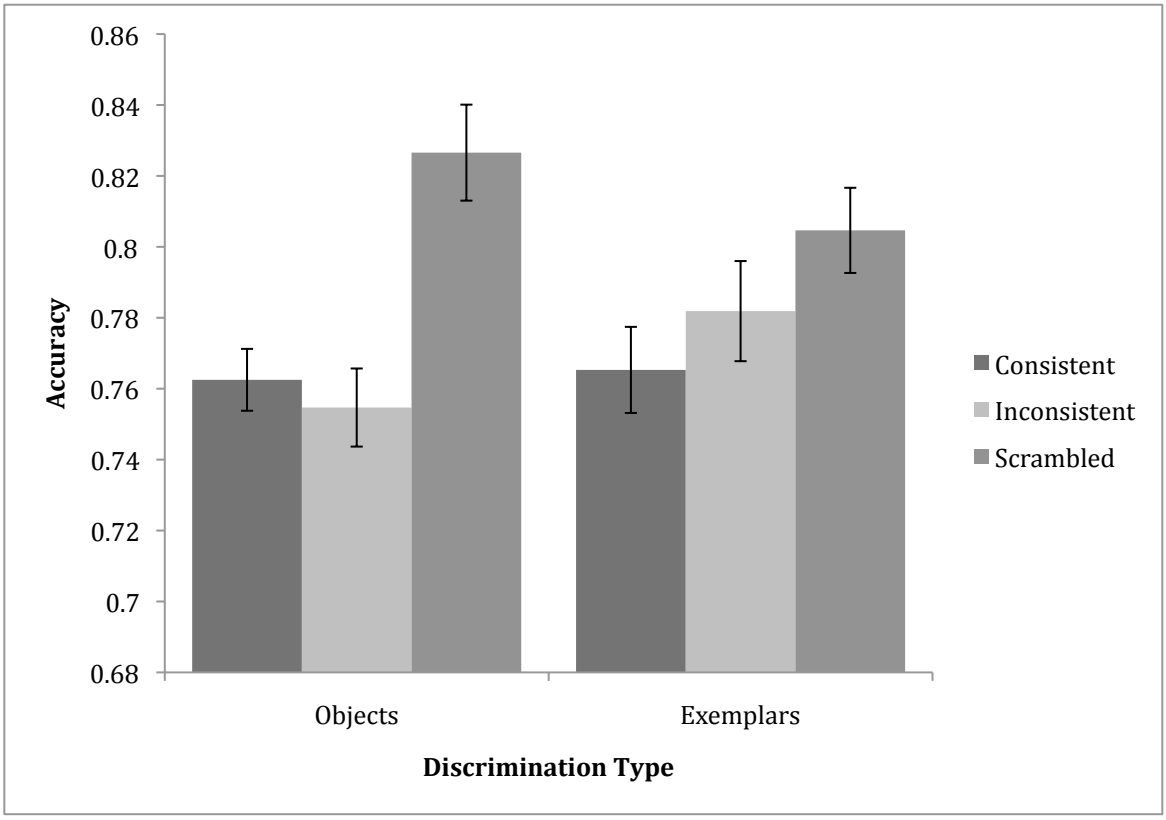


Figure 4.4. Experiment 4 deviation from baseline d'

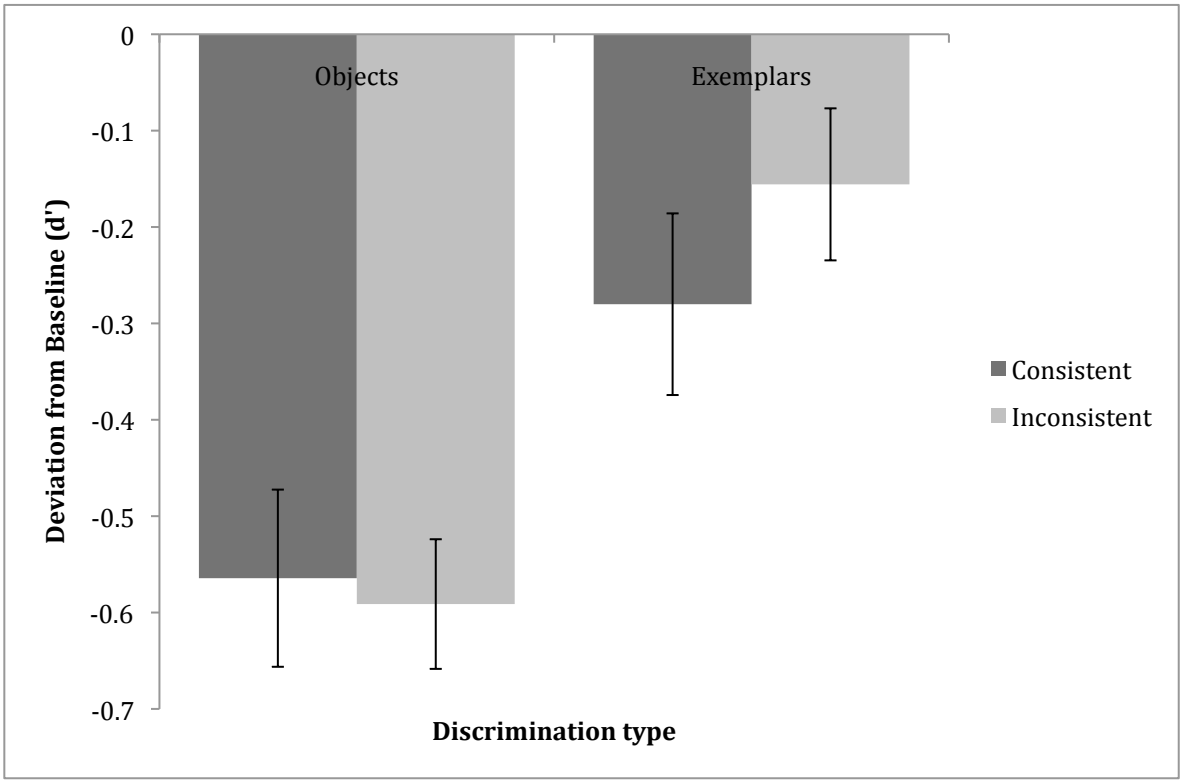


Figure 4.5. Experiment 4 Raw d'

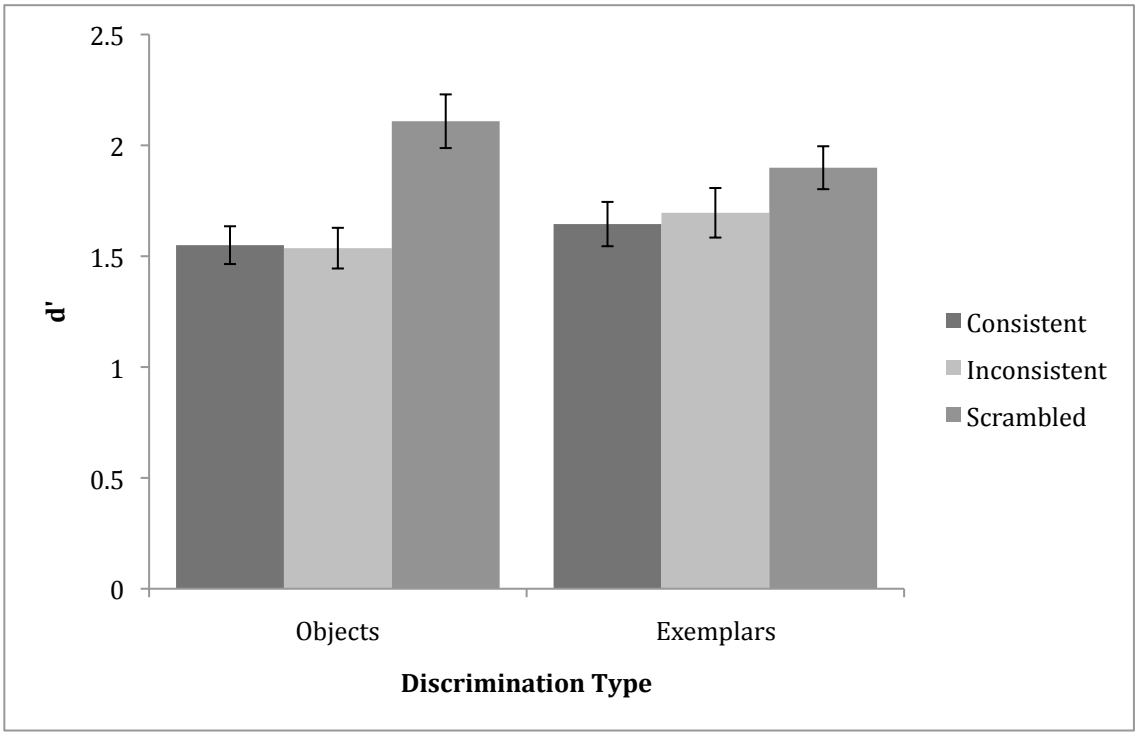


Figure 4.6. Exp. 4 response time by discrimination type

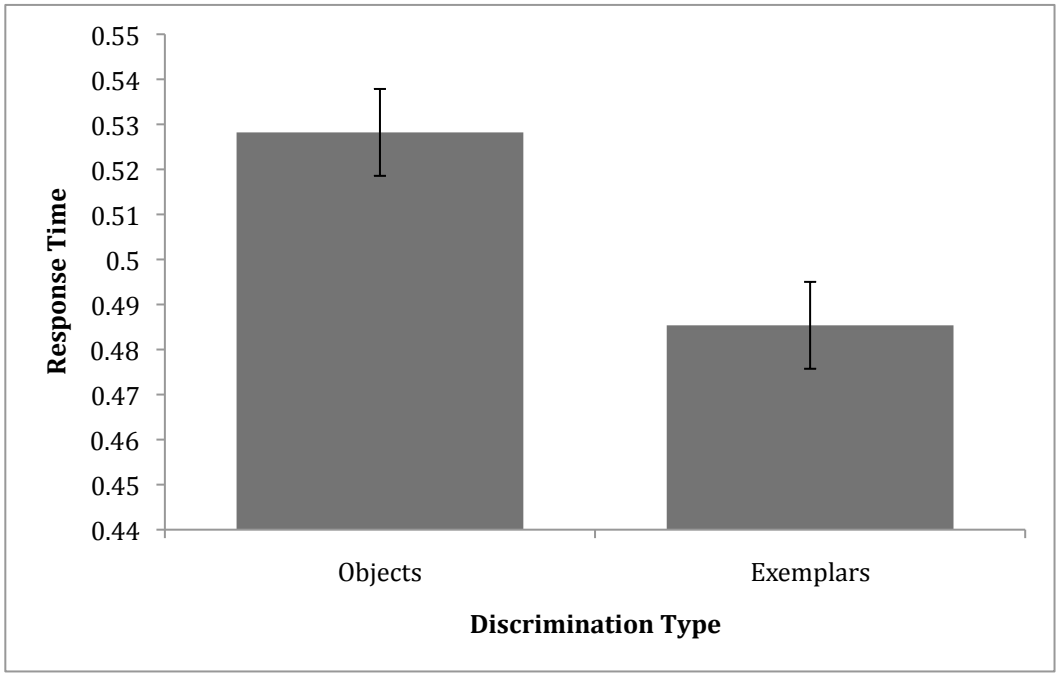


Figure 4.7. Correlation between d' effect size (consistent-background d' – inconsistent-background d') and average RT (by subject) in the object-discrimination condition

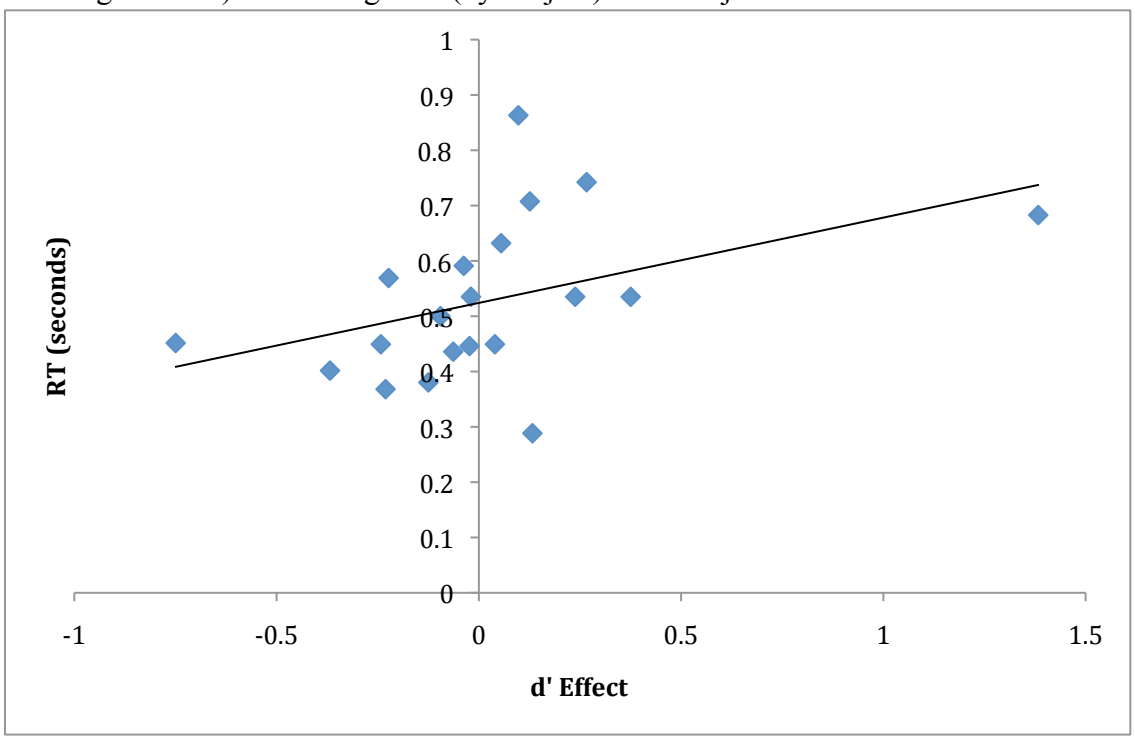


Figure 4.8. Correlation between accuracy effect size (consistent-background accuracy – inconsistent-background accuracy) and average RT (by subject) in the object-discrimination condition

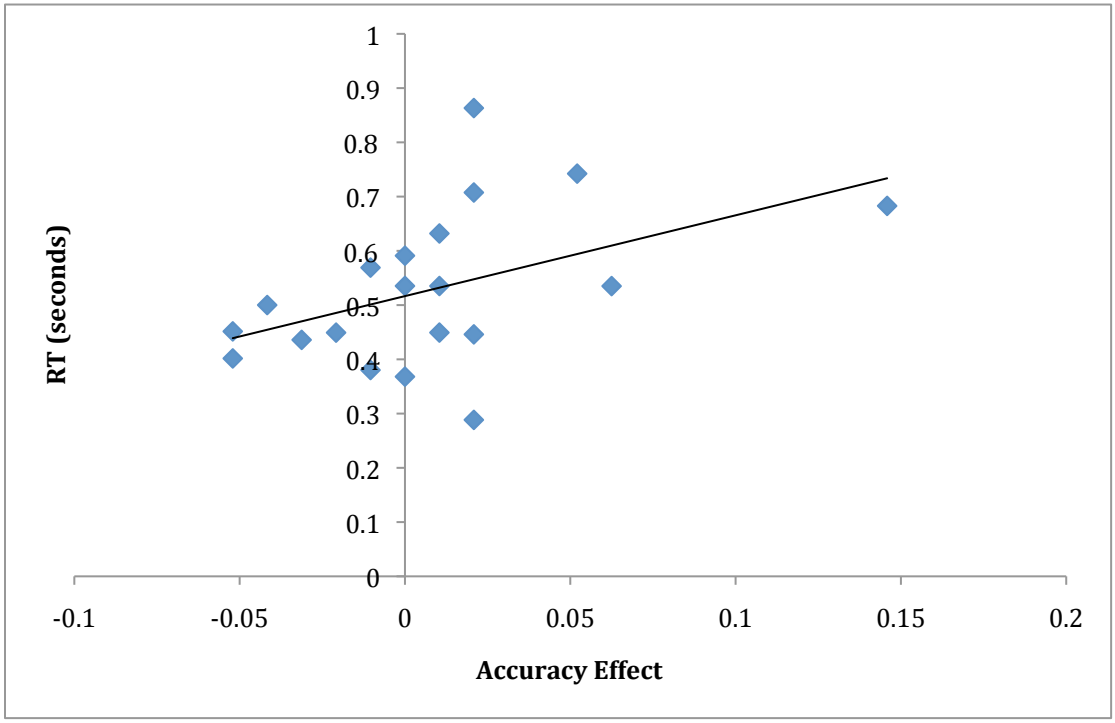


Figure 4.9. Correlation between d' effect size (consistent-background d' – inconsistent-background d') and average RT (by subject) in the exemplar-discrimination condition

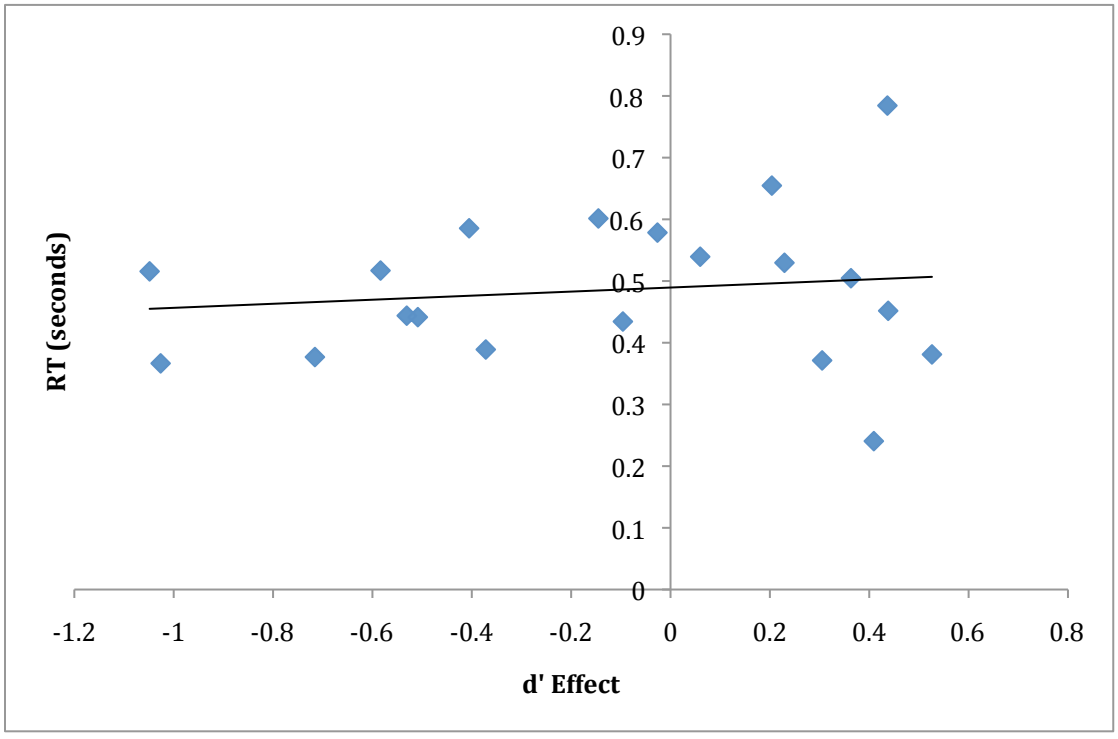
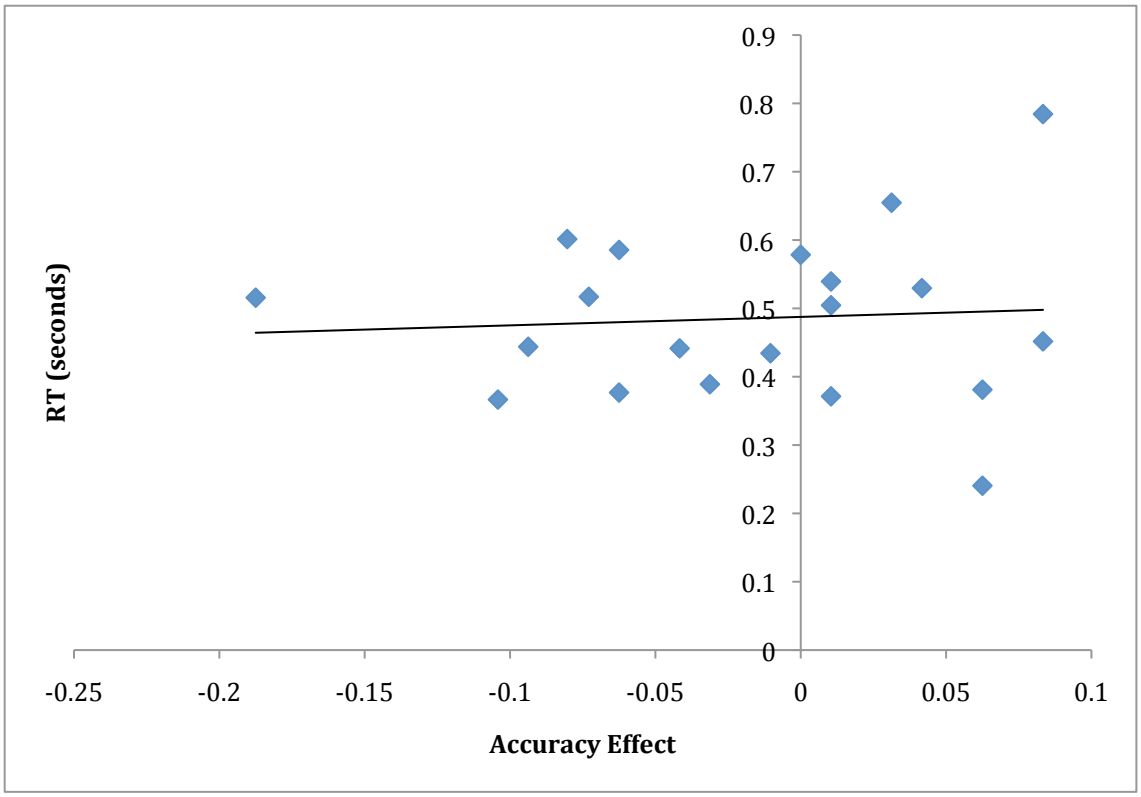


Figure 4.10. Correlation between accuracy effect size (consistent-background accuracy – inconsistent-background accuracy) and average RT (by subject) in the exemplar-discrimination condition



REFERENCES CITED

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

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

Cattell, J.M. (1886). The time taken up by cerebral operations. *Mind, 11(42)*, 220-242.

Chase, C.H. & Tallal, P. (1990). A developmental, interactive activation model of the word superiority effect. *Journal of Experimental Child Psychology, 49*, 448-487.

Chun, M.M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology, 36*, 28-71.

Davenport (2007). Consistency effects between objects in scenes. *Memory & Cognition, 35(3)*, 393-401.

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

Evans K.K. & Treisman, A. (2005). Perception of objects in natural scenes: Is it really attention free? *Journal of Experimental Psychology: Human Perception and Performance, 31(6)*, 1476-1492.

Goodale, M.A. & Milner, A.D. (1992). Separate visual pathways for perception and action. *Trends in Neuroscience, 15(1)*, 20-25.

Green, C. & Hummel, J.E. (2006). Familiar interacting object pairs are perceptually grouped. *Journal of Experimental Psychology: Human Perception and Performance, 32(5),*1107-1119.

Hollingworth, A., & Henderson, J.M. (1998). Does consistent scene context facilitate object perception? *Journal of Experimental Psychology: General, 127(4)*, 398-415.

Henderson & Hollingworth (1999). High-level Scene Perception. *Annual Review of Psychology, 50*, 243-271.

Intraub, H. (1981). Rapid conceptual identification of sequentially presented pictures. *Journal of Experimental Psychology: Human Perception and Performance, 7(3)*, 605-610.

Joubert, O.R., Rousselet, G.A., Fize, D., Fabre-Thorpe, M. (2007). Processing scene context: Fast categorization and object interference. *Visoin Research, 47*, 3286-3297.

Li, F.F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences*, *99(14)*, 9596-9601.

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

Loschky, L.C., Sethi, A., Simons, D.J., Pydimarri, T.N., Ochs, D., & Corbeille, J.L. (2007). The importance of information localization in scene gist recognition. *Journal of Experimental Psychology: Human Perception and Performance, 33(6)*, 1431-1450.

Mack, M.L. & Palmeri, T.J. (2010). Modeling categorization of scenes containing consistent versus inconsistent objects. *Journal of Vision, 10(3):11*, 1-11.

MacMillan, N.A., & Creelman, C.D. (2005). *Detection theory: A user's guide.* Mahwah, NJ: Eribaum.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part I. An account of basic findings. *Psychological Review, 88,* 375-407.

Oliva, A. & Torralba, A. (2006). Chapter 2 – Building the gist of a scene: The role of global image features in recognition. *Progress in Brain Research, 155(2)*, 23-36.

Paap, K.R., Newsome, S.L., McDonald, J.E. & Schvaneveldt, R.W. (1982). An activation-verification model for letter and word recognition: The word-superiority effect. *Psychological Review, 89(5)*, 573-594.

Palmer, S.E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition, 3(5)*, 519-526.

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

Reicher G.M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology, 81(2),* 275-280.

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time- and spatial-scale-dependent scene recognition. *Psychological Science, 5,* 195-200.

Stanislaw, H. & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31(1),* 137-149.

Ungerleider, L.G., & Haxby, J.V. (1994). 'What' and 'where' in the human brain. *Current Opinion in Neurobiology, 4,* 157-165.

Watson, A.B. & Pelli, D.G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics, 33(2),* 113-120.

Wheeler (1970). Processes in word recognition. *Cognitive Psychology, 1,* 59-85.