

# The Influence of Color on the Perception of Scene Gist

Monica S. Castelhana  
University of Massachusetts at Amherst

John M. Henderson  
University of Edinburgh

In 3 experiments the authors used a new contextual bias paradigm to explore how quickly information is extracted from a scene to activate gist, whether color contributes to this activation, and how color contributes, if it does. Participants were shown a brief presentation of a scene followed by the name of a target object. The target object could be consistent or inconsistent with scene gist but was never actually present in the scene. Scene gist activation was operationalized as the degree to which participants respond “yes” to consistent versus inconsistent objects, reflecting a response bias produced by scene gist. Experiment 1 demonstrated that scene gist is activated after a 42-ms exposure and that the strength of the activation increases with longer presentation durations. Experiments 2 and 3 explored the contribution of color to the activation of scene gist. The results revealed that color has an influence across a wide variety of scenes and is directly associated with scene gist.

*Keywords:* scene perception, scene gist, conceptual representation, perceptual properties

Despite the complexity of visual input, it has been well established that a 100-ms exposure is sufficient to extract a scene’s basic-level category or scene gist (Intraub, 1980; Metzger & Antes, 1983; Oliva & Schyns, 1997; Potter, 1975, 1976; Potter & Levy, 1969; Schyns & Oliva, 1994). This finding makes it difficult to conceive that scene gist is acquired from the serial identification of component objects (Friedman, 1979; Marr, 1982; Marr & Nishihara, 1978). It is more probable that scene gist is acquired from global, scene-based information that is quickly extracted and is predictive of gist. The current study is an exploration of how one of the global properties, color, contributes to the extraction of scene gist. In addition, a new paradigm is introduced as a more sensitive probe of the timing of visual information extraction that leads to the activation of scene gist. It is hoped that with this new

paradigm, some of the limitations of past studies will be avoided and the inconsistent findings can be resolved.

The type of methodology used across previous studies that have examined scene gist varies greatly. Many past studies have relied on naming or judgment tasks that require participants to explicitly label scenes. Judgment tasks that have open-ended responses often result in highly variable responses for any given image, and require the experimenter to make a subjective judgment about whether the given response is an acceptable label for a particular scene. Verification tasks require participants to judge whether a given label matches a particular scene. These tasks face the same uncertainty of labeling found with the judgment tasks. They require that the experimenter assign a label to the scene that will correspond to the participant’s label; otherwise, a higher error rate will result. This has led many researchers to limit the types of scenes in experiments to a subset of categories that are easily labeled by participants. The limited set of stimuli that have prototypical properties for their category may be processed faster or easier than scenes that are not readily labeled, and so a mischaracterization of scene gist perception may result.

To avoid the problems associated with naming and judgment tasks, some researchers have opted to use a go/no-go detection task (Delorme, Richard, & Fabre-Thorpe, 2000; Fabre-Thorpe, Delorme, Marlot, & Thorpe, 2001; Thorpe, Fize, & Marlot, 1996; Van Rullen & Thorpe, 2001). Scenes are flashed for an extremely brief duration (20–40 ms), and participants are asked to detect the presence of a target object (e.g., animal, food item, human, or vehicle). It is argued that the detection task is a superior method because it is better able to capture early processes; however, it remains unclear what type of processing is necessary to complete this task. Recent research suggests that instead of processing scene gist, participants are basing their responses on specific visual features that correspond with target identity (e.g., animal features; Evans & Treisman, 2005). To successfully complete the task, one need not interpret the scene gist. For instance, to detect that there is a foxlike animal in a scene, one need not process the scene as a forest or field or cage. Therefore, it is unclear whether the visual

---

Monica S. Castelhana, Department of Psychology, University of Massachusetts at Amherst; John M. Henderson, Department of Psychology, University of Edinburgh, Edinburgh, Scotland.

This research was supported by a Michigan State University graduate school fellowship to Monica S. Castelhana on behalf of the National Science Foundation (NSF) Integrative Graduate Education and Research Traineeship program (Grant ECS-9874541). Further support was provided by NSF Grant BCS-0094433 and Army Research Office Grant W911NF-04-1-0078. The opinions expressed in this article are those of the authors and do not necessarily represent the views of the Department of the Army or any other governmental organization.

This research was completed in fulfillment of Monica S. Castelhana’s doctoral degree, and we thank committee members Erik Altmann, Thomas Carr, Fred Dyer, and Aude Oliva for their valuable input, support, and advice on this project and Keith Rayner and Kristin Weingartner for comments on drafts of this article. We also thank Tonisha Shanks, Jennifer Garman, and Devon Witherell for data collection help; Dan Gajewski, Aaron Pearson, James Brockmole, and Mareike Wieth for helpful discussions; and Gary Schrock and Christine Miscisin for technical support.

Correspondence concerning this article should be addressed to Monica S. Castelhana, who is now at Department of Psychology, Queen’s University, Kingston, Ontario K7L 3N6, Canada. E-mail: monica.castelhana@queensu.ca

features being used are linked to the scene concept. Taken as a whole, these methodologies have a limited ability to assess how quickly the information necessary to activate scene gist is extracted from a scene owing to response variations and strategies that participants adopt to complete the task more efficiently.

The variation in methodologies across studies is likely to have arisen from different interpretations of “scene gist.” As a concept, scene gist is elusive, and its definition varies widely (Friedman, 1979; Oliva, 2005; Potter, 1976, 1999). We attempt to define it here by beginning with the photographic scenes used in this study. All scenes were views of real-world environments that were scaled to a human size, corresponding to the type of views that people experience in their everyday lives. For these types of scenes, people have amassed knowledge of typical object contents, spatial layout, functions that can be performed in or around the scene, and other semantically related information. Thus, when enough information is extracted, the scene’s basic-level category, the information regarding the expected scene content, and functions that can be performed within the scene are all activated; the activation of this information is collectively referred to as the scene gist. For instance, if a kitchen is presented for the minimum duration required for the necessary amount of visual information to be acquired, then it follows that sufficient information was extracted for a person to be able to reason not only that a kitchen was presented but that a stove, refrigerator, and toaster are likely to be present and that a lawnmower, bed, and mailbox are not likely to be present. In fact, this *response bias* toward affirming the presence of semantically consistent objects and rejecting objects if they are inconsistent with scene gist has been found in past studies (Hollingworth & Henderson, 1998, 1999).

In the current study, then, the definition of *scene gist* includes the inferences made about a particular scene, including expected component objects and expected layout. This definition is akin to the scene’s conceptual representation, often described as the semantic or descriptive depiction of a scene (for further elaboration, see Oliva, 2005; Potter, 1999). Because we do not distinguish here between *scene gist* and a scene’s initial *conceptual representation*, these terms are henceforth used interchangeably. The purpose of the present study was to examine the visual properties of a scene that lead to the activation of scene gist, which in turn supports inferences and predictions about the scene.

To investigate what types of information lead to the activation of scene gist, recent studies have focused on scene structure as an important global property. Scene structure refers to the “shape” of a scene, which is defined by the different surfaces and their spatial layout within the scene (Burton & Moorhead, 1987; Oliva & Torralba, 2001; Torralba & Oliva, 2002). Scene structure is a potentially important global property because it is both quickly extracted and found to be predictive of scene gist (Oliva & Schyns, 1997; Párraga, Brelstaff, & Troscianko, 1998; Renninger & Malik, 2004; Schyns & Oliva, 1994; Torralba & Oliva, 2003; Vailaya, Jain, & Zhang, 1998). Studies have also found that this type of information is processed rapidly and is available to the scene-processing system after only a brief exposure. Studies using hybrid scenes (i.e., low spatial frequency information from one scene is combined with high spatial frequency information from another scene, creating two possible interpretations from the same image) have shown that identification of scene gist was based on low spatial frequency information for very brief exposures and on high

frequency information with longer exposures (Schyns & Oliva, 1994). Further research showed that this low-to-high use of frequency information is modulated by task constraints (Oliva & Schyns, 1997). There is also evidence that there are structural statistical regularities in scene images, and these statistical regularities can be used to make higher level categorical distinctions, such as natural versus man-made environments (Torralba & Oliva, 2003; Vailaya et al., 1998). Taken together, these results suggest that a scene’s structure conveys a great deal of information that can be rapidly acquired by the system.

The findings of scene structure notwithstanding, one question that still remains concerns the contribution of color to gist perception. In addition to the processing of luminance information, pathways within the visual system are devoted to the processing of color (Edwards, Xiao, Keysers, Földiák, & Perrett, 2001; Gegenfurtner, 2003; Gegenfurtner & Kiper, 2003; Livingstone, 1988; Livingstone & Hubel, 1984, 1988). The studies on scene structure reviewed above were limited to black-and-white depictions of scenes, and so it is not clear from these studies whether color plays an additional role in the rapid perception of scene gist. Thus, the rapid extraction of color may provide an additional important global scene property for the initial perception of scene gist.

In the area of scene perception, few studies have investigated the role of color in rapid scene gist perception (Delorme et al., 2000; Goffaux et al., 2005; Oliva & Schyns, 2000), and the findings have been inconsistent across these experiments. In one study, Delorme et al. (2000) found that when scenes were presented rapidly, color had a small effect on performance for scenes containing food items and no effect for scenes containing animals. They also found that the small effect on food items coincided with longer RT responses and concluded that the initial processing of the stimuli did not involve color information. This study used the go/no-go task and has the same limitations described above. It is not clear how much of the overall scene was processed as opposed to looking for distinctive visual details that would signal the presence of the target object. The presence of an effect of color when participants were trying to detect food items suggests that in this case color may have been an important feature to process in order to complete the task. However, it remains unclear whether detecting the target in the scene required the processing of scene gist at all. In contrast, Oliva and Schyns (2000) found that color information had an effect on scene processing, but only in certain circumstances. More specifically, categorization of natural scenes was slower and less accurate when color information was removed compared with man-made scenes, which showed no difference between the colored and monochrome versions. These results were replicated in a recent study measuring event-related potentials for the natural scene stimuli (Goffaux et al., 2005). In each case, the authors concluded that the early processing of the scene must use color information (regardless of the scene type) to produce an effect when natural scenes are viewed. However, these studies did not show the effects for color with man-made scenes, and it remains unclear how the system would “decide” to use color information early on in the case of natural but not man-made scenes, before their identity is known. Thus, although color information seems to be important in some cases (e.g., perceiving food or natural images), it is not clear how or whether color plays a role in the initial activation of scene gist regardless of scene type.

In the current study, we investigated the influence of color on initial scene gist perception by measuring the influence of gist on biases in object judgments. In the contextual bias paradigm, response bias is used as a measure of how quickly after a scene's onset sufficient information is extracted to activate its gist. Participants are simply asked to indicate the presence of a target object whose consistency with the scene has been manipulated. It is important to note that the target objects are never present in the scenes. Rather, responses that the target is present reflect false alarms (i.e., response biases). The logic of the paradigm is that if the visual information acquired from a brief presentation of the scene is sufficient to activate scene gist, then responses should be influenced by this knowledge and reveal a bias. One advantage of using the contextual bias paradigm over the methodologies used in past studies is that it does not require that the participants and experimenters be in complete agreement about the correct basic-level category or name for a particular scene. This new paradigm circumvents this problem by using probes that all can agree are likely (or unlikely) to appear in a particular type of scene. In addition, this methodology is an improvement in estimating the information required for scene gist acquisition over a fast classifying or detection method. These methods may misrepresent how quickly scene gist is acquired owing to alternative strategies adopted by participants that rely on feature detection of a target object, therefore producing timing estimates that are too low. The contextual bias paradigm provides a measure that is dependent on the scene being understood but does not require providing a prespecified target and therefore should provide a more accurate measure of how quickly scene gist is acquired.

In the current study, three experiments were conducted to address (a) how much time is required to extract sufficient information for the activation of scene gist, (b) whether color has any influence on gist activation, and (c) if so, what role color plays in gist activation. Experiment 1 (a–c) was designed to investigate the validity of the contextual bias paradigm. Experiment 2 addressed whether color had any influence on scene gist by systematically varying the properties of color and structure of the scene. Finally, Experiment 3 addressed the question of how color influences the activation of scene gist by altering the informativeness of color.

### Experiment 1

Experiment 1 investigated the nature of the response bias produced by scenes presented for different durations. The participants' task was to indicate whether a named target object was present in the previously viewed scene. The target objects were never present in the scenes, and participants were not informed that the target objects were missing. Therefore, this task was not a detection task but rather measured response bias only (measures collected are equivalent to false alarms and correct rejections). As the initial conceptual representation becomes more complete, it should bias participants to respond "yes" more often to consistent than to inconsistent target objects. The response bias is measured as the difference in "yes" responses to consistent and inconsistent objects. The degree of this difference indicates the strength of gist activation.

Experiment 1a investigated a broad range of durations from 20 ms to 250 ms. Experiment 1b was designed to investigate how early the response bias can be measured by displaying scenes in a more fine-grained range of durations than those in Experiment 1a,

ranging from 20 ms to 50 ms. Experiment 1c investigated the nature of this increase in response bias by looking at a later time frame, with a range of durations from 50 ms to 100 ms. According to past findings, gist information should be available with presentation times of 100 ms (see Potter, 1975, 1976; Schyns & Oliva, 1994). If scene gist can be activated with shorter presentation times, then the effect of object consistency should be seen with presentation durations under 100 ms.

### Method

*Participants.* Ninety Michigan State University undergraduates (24 in Experiment 1a, 30 in Experiment 1b, and 36 in Experiment 1c) participated in the experiment. All participants received credit toward an introductory psychology course.

*Apparatus and stimuli.* The stimuli were full-color photographs taken from a number of sources (books, calendars, the Internet, and personal photos). The scenes selected were mostly man-made, with a few scenes containing both natural and man-made characteristics (e.g., parks, playgrounds, and gardens). There were a total of 80 scenes presented in Experiment 1a, 80 presented in Experiment 1b, and 96 presented in Experiment 1c. For each scene a consistent and an inconsistent object were selected.<sup>1</sup>

The scenes were presented on a Dell P78 Trinitron 16-in. (41.1-cm) monitor driven by a GeForce3 NVIDIA Pro super video graphics adapter card. The refresh rate was set at 100 Hz for Experiment 1a and at 120 Hz for Experiments 1b and 1c. The scenes had a resolution of 800 × 600 pixels and subtended 30° × 22.5° of visual angle viewed from 61.5 cm away.

*Design.* All experiments (1a–1c) had two target conditions, consistent and inconsistent, which were displayed with equal frequency across conditions to all participants (50% consistent and 50% inconsistent). The only factor that varied across Experiments 1a–1c was duration. In Experiment 1a, there were four duration conditions (20, 50, 100, and 250 ms). In Experiment 1b, the duration conditions were 25, 33, 42, 50, and 250 ms (the 250-ms condition was included in order to ensure that the participants were actually performing the task). In Experiment 1c, participants viewed the scenes for 50, 58, 75, 83, 92, and 100 ms. Thus, the

<sup>1</sup> A pilot study was conducted in which the scenes were shown for 250 ms each and followed by a target word (either consistent or inconsistent with the scene). The instructions in the pilot study were the same as in Experiment 1, in that participants were asked to indicate whether the target object was present in the scene. Scenes were shown for 250 ms (enough time for the gist to be extracted but not for an eye movement to be executed, thus revealing the bias effect without an active search for the target object). Ten participants were shown 460 scenes (in either of the target conditions), and bias effects were calculated. Scenes with a response bias less than one standard deviation from the mean had their consistent and inconsistent objects changed, and the experiment was rerun with an additional 10 participants. In all, 20 participants were included in the pilot study, and for each of the experiments reported in this study, scenes were taken from this pool of photographs with their respective target objects. Although not explicitly controlled, some objects could appear as consistent and inconsistent targets across different scenes. The main goal of the pilot was to gather a large collection of scenes for which the responses to consistent and inconsistent objects showed a reliable bias effect at 250 ms (known to be more than enough time to extract the scene's concept).

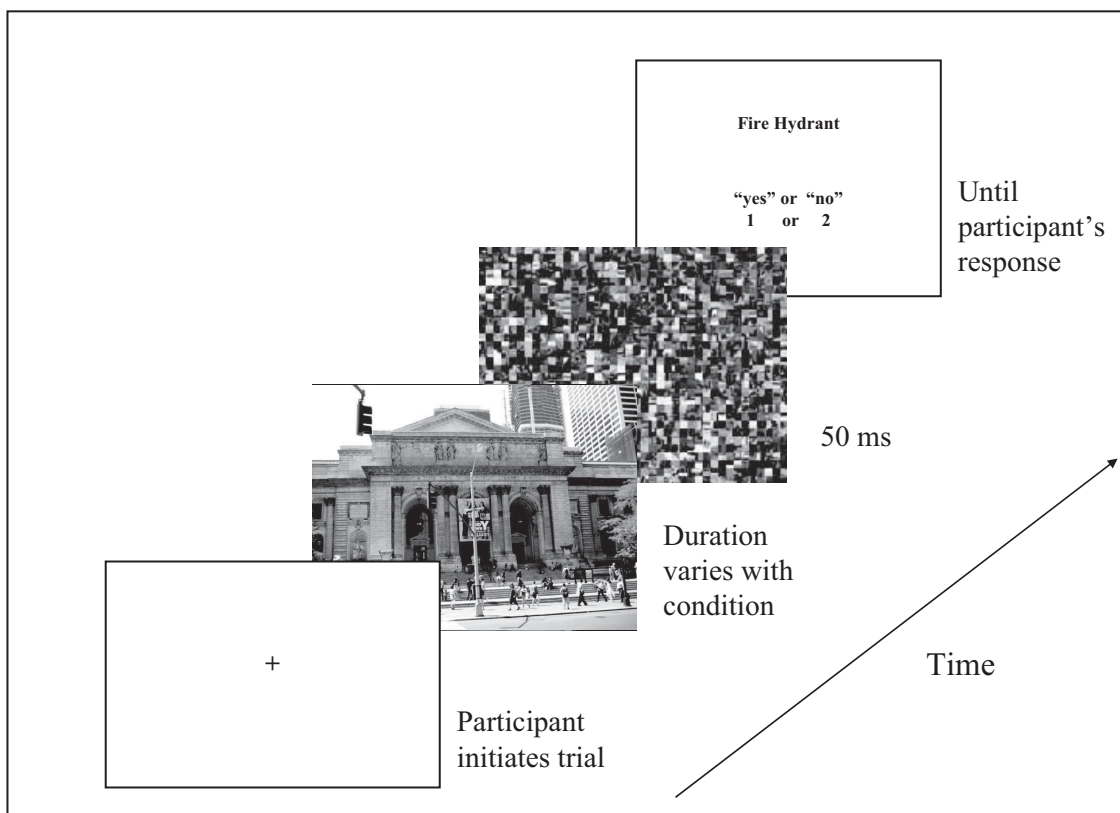
design was  $2 \times 4$  for Experiment 1a,  $2 \times 5$  for Experiment 1b, and  $2 \times 6$  for Experiment 1c.

**Procedure.** The experimenter explained the sequence of events for each trial, as depicted in Figure 1. At the beginning of a trial, participants fixated a center fixation cross displayed for 2,000 ms. A photograph of a scene was presented for between 20 ms and 250 ms, depending on the duration condition, which varied across Experiments 1a–1c. Each participant viewed all conditions and viewed each scene once. Scenes were counterbalanced for each condition across participants and were presented in a random order (determined by the program for each individual). The presentation of the scene was followed by a visual mask for 50 ms. The mask was composed of a jumble of scene sections taken from the collection of scenes being shown in the experiment that were then further scrambled so that no individual section was recognizable. This allowed for the same variation of colors and textures, but without any interpretable visual information. Following the mask, the target object name was displayed at the center of the screen until the participant responded. The target condition was varied across trials and for each scene was either semantically consistent or inconsistent. The target was never present in the scene. Therefore, the participants' responses were based on what they knew about the image and never based on having viewed that object in the scene. The participants responded 1 for yes and 2 for no with a button press (these labels were always presented as a reminder on each response screen below the word). Although no

participant saw the same scene twice, all scenes and target types (consistent vs. inconsistent) were counterbalanced across participants using a Latin square design. The experiment took approximately 10–20 min to complete.

### Results

Across Experiments 1a–1c, results were calculated and analyzed using the same method. For each duration condition, the proportion of “yes” responses was calculated for both the consistent and inconsistent target conditions. The bias effect was considered to exist for a given duration when the proportions of “yes” responses was significantly higher for the consistent target condition than for the inconsistent target condition (meaning that participants were more biased to respond “yes” when the target object was consistent than when it was not consistent). If scene gist is acquired within a given presentation time, then the consistent targets should produce more “yes” responses than the inconsistent targets; this is referred to as the bias effect. The bias effect was assessed with omnibus analysis of variance (ANOVA) followed by planned comparisons testing whether “yes” responses were greater for the consistent targets over the inconsistent targets for each duration condition using a one-tailed *t* test. Scenes were considered to be *recognized* when the bias effect was statistically significant. To avoid possible Type I errors due to multiple comparisons at each level of the duration condition, we used a Bonferroni correction so that all



**Figure 1.** Trial sequence. The photographs and mask were shown in full color. For this particular scene, the target object was *fire hydrant* for the consistent condition and *tea set* for the inconsistent condition.

comparisons collectively had an alpha of .05. This correction was applied for every experiment reported hereafter. The specific correction made for each experiment (based on the number of comparisons) is reported below. Figure 2 illustrates the results for Experiments 1a–1c.

Figure 2a depicts the proportion of “yes” responses by duration according to target conditions for Experiment 1a. An omnibus

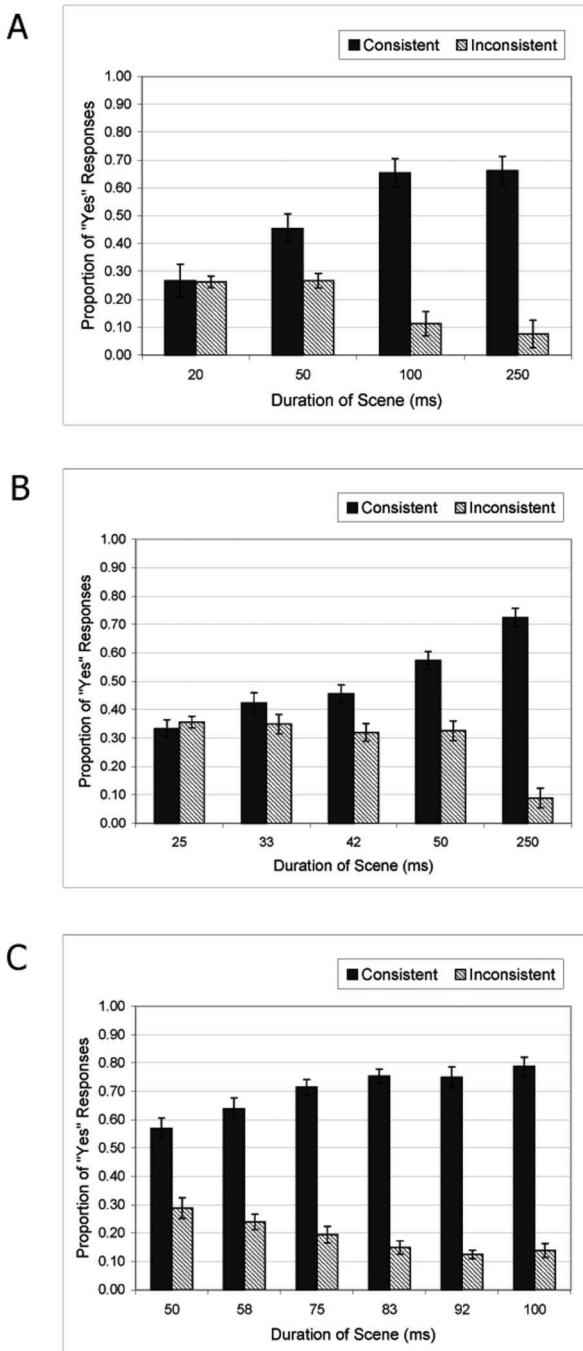


Figure 2. Proportion of “yes” responses to consistent and inconsistent target objects for each duration condition in Experiment 1: (A) Experiment 1a; (B) Experiment 1b; (C) Experiment 1c. Error bars represent standard errors of the means.

ANOVA revealed that there was a main effect of target condition,  $F(1, 23) = 138.37, p < .01, MSE = 0.038$ , partial  $\eta^2 = .86$ , in which mean “yes” responses were significantly higher for consistent targets than for inconsistent targets; a main effect of duration condition,  $F(3, 69) = 4.73, p < .01, MSE = 0.031$ , partial  $\eta^2 = .17$ , which revealed an overall increase in “yes” responses as duration increased; and a significant interaction between target condition and duration,  $F(3, 69) = 56.09, p < .01, MSE = 0.017$ , partial  $\eta^2 = .71$ , in which the relative difference between targets increased as a function of increase in duration. Further analyses showed a significant linear trend for an increase in the bias effect (calculated as the consistent object mean minus the inconsistent object mean) as duration increased,  $F(1, 23) = 87.78, p < .01, MSE = 0.06$ , partial  $\eta^2 = .79$ . To examine at which point the rate of “yes” responses for the consistent objects was greater than the rate for inconsistent objects (thus, the bias emerged), we conducted a priori comparisons with a Bonferroni correction ( $\alpha = .0125$  for each comparison), which revealed significant differences at durations of 250 ms,  $t(23) = 10.27, p < .01$ ; 100 ms,  $t(23) = 12.15, p < .01$ ; and 50 ms,  $t(23) = 5.69, p < .01$ . Scenes presented at 20 ms showed no significant difference,  $t(23) = 0.12, ns$ .

Results for Experiment 1b are shown in Figure 2b. An omnibus ANOVA revealed that the pattern of results paralleled the results of Experiment 1a. Specifically, there was an overall effect of target condition,  $F(1, 29) = 67.10, p < .01, MSE = 0.052$ , partial  $\eta^2 = .70$ ; a main effect of duration condition,  $F(4, 116) = 2.60, p < .05, MSE = 0.041$ , partial  $\eta^2 = .08$ ; and a significant interaction between target condition and duration,  $F(4, 116) = 33.21, p < .01, MSE = 0.032$ , partial  $\eta^2 = .53$ . As was found in Experiment 1a, there was also a significant linear trend for an increase in bias over increasing presentation durations,  $F(1, 29) = 50.59, p < .01, MSE = 0.07$ , partial  $\eta^2 = .59$ . Planned comparisons ( $\alpha = .01$  for each comparison) revealed significant differences at durations of 250 ms,  $t(29) = 16.14, p < .01$ ; 50 ms,  $t(29) = 5.89, p < .01$ ; and 42 ms,  $t(29) = 2.50, p < .01$ . Scenes presented at 33 ms,  $t(29) = 1.25, ns$ , and 25 ms,  $t(29) = -0.67, ns$ , showed no significant difference.

Results for Experiment 1c are shown in Figure 2c. An omnibus ANOVA revealed the same pattern of effects as in the previous two experiments. There was an overall effect of target condition,  $F(1, 35) = 476.19, p < .01, MSE = 0.06$ , partial  $\eta^2 = .93$ ; no main effect of duration condition,  $F(5, 175) = 0.46, ns, MSE = 0.026$ , partial  $\eta^2 = .01$ ; and a significant interaction between target condition and duration,  $F(5, 175) = 17.60, p < .01, MSE = 0.022$ , partial  $\eta^2 = .34$ . As in Experiments 1a and 1b, there was also a significant linear trend for an increase in bias over increasing presentation durations,  $F(1, 29) = 104.20, p < .01, MSE = 3.56$ , partial  $\eta^2 = .78$ . Planned comparisons ( $\alpha = .008$  for each comparison) revealed significant differences across all duration conditions: 100 ms,  $t(35) = 16.32, p < .001$ ; 92 ms,  $t(35) = 18.92, p < .001$ ; 83 ms,  $t(35) = 18.19, p < .001$ ; 75 ms,  $t(35) = 12.73, p < .001$ ; 58 ms,  $t(35) = 9.74, p < .001$ ; and 50 ms,  $t(35) = 5.97, p < .008$ .

## Discussion

The first experiment was designed to demonstrate the efficacy of the contextual bias paradigm and to investigate how quickly information is extracted for the activation of scene gist as measured by the presence of a response bias. The response bias was calculated as the difference between the proportion of “yes” re-

sponses to consistent target objects and the proportion of “yes” responses to inconsistent target objects. Experiment 1 clearly established that the contextual bias paradigm can be used to determine how quickly the information needed to activate gist is extracted from a scene. Experiment 1a showed that whereas a 20-ms exposure was not sufficient, after a 50-ms presentation, enough visual information was extracted to activate scene gist. The results also revealed an increase in the size of the bias effect as scene duration lengthened. Therefore, consistent with Schyns and Oliva (1994; Oliva & Schyns, 1997), our findings show that scene gist is able to influence behavior with display durations shorter than the 100-ms estimate given in some previous studies (Intraub, 1980; Metzger & Antes, 1983; Potter, 1975, 1976; Potter & Levy, 1969). It is interesting to note that participants in this experiment were not told until debriefing that the objects were not actually present in any of the scenes. When asked during the debriefing period whether they had suspected this during the experiment, participants typically reported that they had not noticed.

Experiment 1b investigated the onset of the bias after shorter exposure durations. Results revealed that 42 ms of exposure was sufficient to elicit a response bias. Additionally, the results replicated Experiment 1a in that the response bias for the 250-ms condition was much greater than that found for the 42- and 50-ms conditions (as evidenced by the significant linear trends in both), indicating that the activation of scene gist increased as exposure duration increased. The nature of this pattern was further explored in Experiment 1c. A strong linear trend was also found, revealing that the response bias increased monotonically in strength up to and including 100 ms. These results suggest that activation was increasing with longer stimulus presentations, but no maximum in the amount of activation was uncovered within the duration conditions tested.<sup>2</sup>

It is clear from the findings of Experiment 1 that the information needed to activate gist can be extracted very rapidly from a scene. As discussed in the introduction, although scene structure has been implicated as a potentially important global property in scene perception, the role of color is not as clear. In Experiment 2, we investigated the possible influence of color on the activation of scene gist.

## Experiment 2

Few studies have examined the effect of color on the activation of scene gist. In addition, findings have been inconsistent, and variations in methodology and in the definition of scene perception make the effect of color difficult to determine. Some studies have shown no effect of color (Delorme et al., 2000), whereas others have shown that color can affect scene perception for particular types of scenes (Goffaux et al., 2005; Oliva & Schyns, 2000). In Experiment 2, we investigated the effect of color on the activation of scene gist across a wide variety of scene types by presenting scenes as either colored or monochrome photographs.

There are two views on what role color can play in the perception of scenes. On the one hand, it may be that color influences scene gist through a direct semantic association between color and gist (Gegenfurtner & Rieger, 2000; Price & Humphreys, 1989; Spence, Wong, Rusan, & Rastegar, 2006; Tanaka, Weiskopf, & Williams, 2001). Other researchers posit that the effect of color is negligible, and so initial representation is based on edge-based

information (Biederman, 1988; Biederman & Ju, 1988; Delorme et al., 2000; Grossberg & Mingolla, 1985). If color does contribute to the activation of scene gist, then the contextual bias effect should be larger for colored scenes than for monochrome scenes. Conversely, if color is not used in the early stages of processing leading to the activation of relevant semantic information, then there should be no difference between colored and monochrome scenes.

However, an alternative explanation is that the initial processing of scene gist does involve color information, but because structure information is sufficient to activate gist, the contribution of color is masked (Price & Humphreys, 1989). In this case, the visual system is so efficient at extracting necessary information from the structure of the scene that the level of performance is at ceiling and the addition of color cannot improve the activation of scene gist any further. This hypothesis suggests that only when structural information is not as efficiently extracted (causing the visual system to no longer be at its peak level of efficiency) would a contribution of color be revealed.

In order to investigate the contribution of color independent of that derived from scene structure, we filtered half of the scenes presented in this experiment to remove high-level spatial frequency information while keeping most medium- and low-level spatial frequency information. With some of the structural information removed, the scene structure becomes more difficult to extract and thus may no longer be sufficient to activate scene gist.

<sup>2</sup> Generally speaking, there are two possible reasons for the increase in the response bias with increased duration. The first assumes that the recognition of a scene occurs as an all-or-none process, in which a scene is recognized when information supporting that scene gist reaches a certain threshold. The other possible explanation is that scene recognition is continuous. In this case, a scene is recognized incrementally with increased presentation durations because more supporting visual information is available. Across the three experiments reported so far, there was a gradual increase in response bias across the range of durations. An assumption that all scenes are processed at equal rates and reach similar levels of activation at the same time would lead to the conclusion that the increase in the response bias is due to a gradual accumulation of activation strength in all scenes simultaneously. However, given the variety in complexity and type of information available from one scene to the next, it is unlikely that this assumption is true. It is more likely that scenes are processed at different rates and reach activation of their gist information at different times. For instance, the results in Experiment 1b show that scene gist can be retrieved after 42 ms of exposure, but this may not be the case for all scenes. Taking this into account, it seems that the gradual increase in the response bias could be due to both an increase in the number of scenes that reached some level of scene gist activation and an increase in the activation level as more visual information is available with increasing durations. The actual mechanism responsible for the response bias (e.g., as either a threshold or continuous accumulation of information) is of little consequence to the pattern of the bias effect. In either case, the bias is affected by an increasing availability of information about the scene. Deciding which of the two models is correct is not within the scope of this study and does not affect the predictions. Therefore, for the purpose of the present article, we arbitrarily adopt the view that scene recognition occurs as a continuous activation of related information that can increase over time. Adopting this view does not mean that we support this view exclusively but rather that we have sided with one view in order to outline the predictions of the current investigation more clearly.



Figure 3. Example of the stimulus color conditions used in Experiment 2: (A) normal–colored condition; (B) normal–monochrome condition; (C) blurred–colored condition; (D) blurred–monochrome condition.

Without the full contribution of scene structure, it is possible that an influence of color will be revealed.

Therefore, the effect of color on the activation of scene gist was assessed by manipulating the availability of both the color and the quality of structure information. First, it was expected that with a degraded structure, blurred scenes would show a reduced bias effect overall when compared with normal scenes. Second, if color contributes to the activation of scene gist and this influence is indeed masked by structure, the bias effect should be larger for the blurred colored scenes than for the blurred monochrome scenes, and no color effect should be seen between the normal color and normal monochrome scenes. On the other hand, if color is not extracted and processed in the initial activation of scene gist, then there should be no difference between the color conditions for either the normal or the blurred scenes.

### Method

**Participants.** Eighty Michigan State University undergraduates participated in this experiment for credit in an introductory psychology course.

**Apparatus and stimuli.** With the addition of two conditions (color and quality), another 140 scenes were added to the experiment for a total of 400 scenes.<sup>3</sup> Each photograph had four versions: normal colored, normal monochrome, blurred colored, and blurred monochrome. For the color condition, monochrome versions of the photographs were created by transforming the photograph from RGB to  $L^*a^*b^*$  color mode and then discarding the chromatic components  $a^*$  and  $b^*$  of the colored scenes, leaving only  $L^*$  (luminance information).<sup>4</sup> Thus, this method of creating monochrome photographs maintains the same luminance values as in the original color photograph (Oliva & Schyns, 2000). For the quality condition, the blurred scenes were created by using a low-pass filter at 1 cycle/degree of visual angle (corresponding to 17 cycles/image and therefore keeping most of the mid- and low-level spatial

frequency information). Figure 3 shows an example scene in each of the quality and color conditions.

**Design.** In Experiment 2, color (color or monochrome) and quality (normal or blurred) conditions were introduced. In addition, the duration condition included the following levels: 20, 50, 80, 100, and 250 ms. Therefore, the experiment had a four-factor, within-subject design: quality, color, target, and duration ( $2 \times 2 \times 2 \times 5$ ).

**Procedure.** The experimental procedures were identical to those of Experiment 1a, with the following exceptions. Participants were shown 400 photographs in one of four experimental conditions (varying color and quality). The experiment took approximately 25 to 40 min to complete. Participants were encouraged to take breaks while performing the experiment, and all participants took at least two breaks.

### Results

Two types of analyses were carried out. The first analysis examined the level of exposure needed to produce a response bias. The second analysis assessed the contribution of color; planned comparisons were conducted on each duration and difference score combination (with a Bonferroni correction). Difference scores were calculated for each participant by subtracting the target conditions (consistent and inconsistent) from each other for each duration and color condition separately. In this way, the differences between the biases of color and monochrome scenes are easier to interpret.

Table 1 shows the means and standard deviation for all conditions. An omnibus ANOVA revealed that there was a significant

<sup>3</sup> These additional scenes were taken from the original piloted set.

<sup>4</sup> We thank Aude Oliva for providing the MATLAB code that performed the transformations on the photographs.

Table 1  
Mean (Standard Deviation) Proportion of “Yes” Responses for Experiment 2

Condition	Normal		Blurred	
	Colored	Monochrome	Colored	Monochrome
250 ms				
Consistent	.88 (.12)	.85 (.15)	.81 (.19)	.74 (.24)
Inconsistent	.12 (.13)	.13 (.16)	.19 (.16)	.19 (.17)
100 ms				
Consistent	.85 (.17)	.81 (.15)	.68 (.21)	.61 (.21)
Inconsistent	.16 (.14)	.13 (.16)	.26 (.17)	.28 (.18)
80 ms				
Consistent	.83 (.17)	.82 (.17)	.59 (.22)	.53 (.20)
Inconsistent	.19 (.16)	.16 (.15)	.26 (.17)	.30 (.19)
50 ms				
Consistent	.75 (.19)	.69 (.19)	.40 (.21)	.41 (.22)
Inconsistent	.24 (.27)	.25 (.17)	.26 (.16)	.30 (.22)
20 ms				
Consistent	.37 (.23)	.34 (.23)	.25 (.20)	.27 (.22)
Inconsistent	.25 (.22)	.29 (.24)	.26 (.21)	.26 (.21)

main effect of quality,  $F(1, 79) = 91.64, p < .01, MSE = 0.0378$ , partial  $\eta^2 = .54$ ; color,  $F(1, 79) = 4.99, p < .05, MSE = 0.0925$ , partial  $\eta^2 = .60$ ; target,  $F(1, 79) = 1,372.33, p < .01, MSE = 0.093$ , partial  $\eta^2 = .95$ ; and duration,  $F(4, 316) = 103.12, p < .01, MSE = 0.04$ , partial  $\eta^2 = .57$ . There were several significant two-way interactions that included those between target and quality,  $F(1, 79) = 404.93, p < .01, MSE = 0.032$ , partial  $\eta^2 = .84$ ; quality and duration,  $F(4, 316) = 15.22, p < .01, MSE = 0.022$ , partial  $\eta^2 = .84$ ; target and color,  $F(1, 79) = 13.65, p < .01, MSE = 0.029$ , partial  $\eta^2 = .15$ ; and target and duration,  $F(4, 316) = 374.80, p < .01, MSE = 0.024$ , partial  $\eta^2 = .83$ . There was also a significant three-way interaction between quality, target, and

duration,  $F(4, 316) = 24.18, p < .01, MSE = 0.0248$ , partial  $\eta^2 = .24$ . Finally, the analysis revealed that there was a significant four-way interaction between quality, color, target, and duration,  $F(4, 316) = 3.52, p < .01, MSE = 0.0202$ , partial  $\eta^2 = .04$ . No other interactions were significant.

Of theoretical interest are the effects of color as a function of quality at each display duration. To simplify this analysis, we analyzed difference scores. Figure 4 shows the difference scores plotted as a function of quality and color by the duration condition. Planned comparisons were conducted between the color and monochrome scenes at each duration ( $\alpha = .005$ ) with paired-sample two-tailed  $t$  tests.

For scenes in the normal condition there was no difference between color and monochrome for all durations: 250 ms,  $t(79) = 1.34, ns$ ; 100 ms,  $t(79) = 1.41, ns$ ; 80 ms,  $t(79) = 1.41, ns$ ; 50 ms,  $t(79) = 2.12, ns$ ; 20 ms,  $t(79) = 1.43, ns$ . For the blurred condition, however, there was a significant effect of color at durations of 80 ms,  $t(79) = 2.94, p < .01$ , and 100 ms,  $t(79) = 2.93, p < .01$ , and a marginal effect at 250 ms,  $t(79) = 2.40, p = .019$ . In these cases, there was an advantage for the colored scenes over the monochrome scenes. There was no difference between the colored blurred scenes and monochrome blurred scenes at durations of either 50 ms,  $t(79) = 0.66, ns$ , or 20 ms,  $t(79) = -0.72, ns$ .

Discussion

Experiment 2 was designed to explore the effects of color on normal and structurally degraded (blurred) scenes. These effects were examined for presentation durations ranging from 20 ms to 250 ms. As evidenced by a main effect of quality and as would be expected, blurred scenes showed a smaller bias effect than the normal scenes. Results also revealed that response bias to normal photographs did not vary with color content. In contrast, when blurred scenes were shown, color did have an effect on the acti-

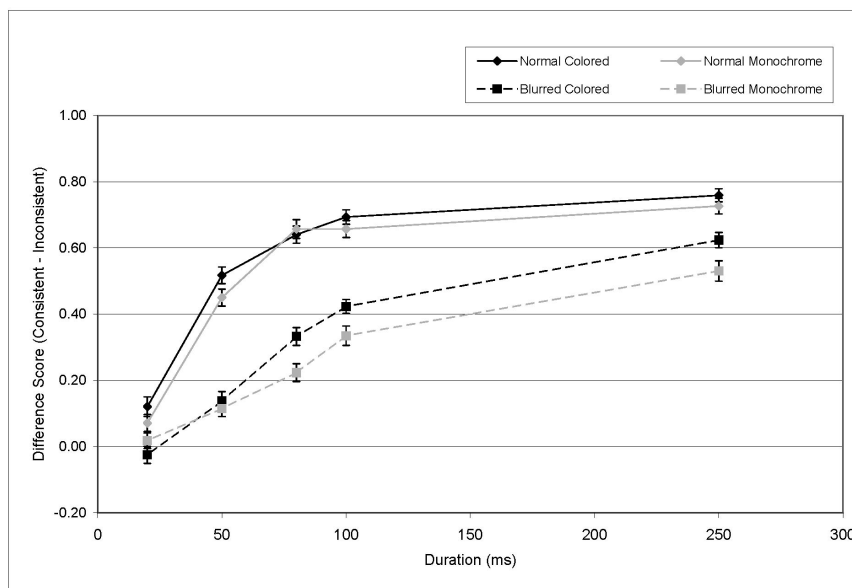


Figure 4. Difference scores (consistent mean minus inconsistent mean) for each color and duration condition combination in Experiment 2. Error bars represent standard errors of the means.



vation of scene gist: A greater response bias was found for color scenes than for monochrome scenes. Furthermore, the results suggest that the effect of color occurred later. At a presentation duration of 50 ms, there was a definite (and not surprising) bias effect present for both normal and blurred scenes but no effect of color. The effect of color only appeared at 80 ms, which is well after the 42-ms onset of the bias effect seen in Experiment 1b. This is an interesting effect that we return to in the General Discussion.

Taken together, the results suggest that color plays a role in activating scene gist. However, on the basis of these results, we cannot determine what specific role color is playing. The two possibilities are (a) that color is able to activate scene gist directly and (b) that boundaries are more difficult to extract from blurred scenes, and added color improves segmentation of the scene. Experiment 3 was designed to test between these two possibilities.

### Experiment 3

Color could play two roles in the activation of scene gist: Color could help with scene segmentation, or color could directly act as a cue for scene gist owing to its association with particular scene concepts in memory. To tease apart these possibilities, Experiment 3 investigated the influence of color on the activation of scene gist for structurally degraded scenes (i.e., for blurred scenes) by measuring performance for abnormally colored scenes in comparison to normally colored or monochrome scenes. The use of abnormally colored scenes can reveal how color contributes to the processing of scene gist because abnormal hues provide segmentation information without any link to the relevant gist information.

As discussed above, previous studies have shown that scene structure can activate scene gist. If color contributes to scene gist activation by providing additional segmentation information, then the specific hues present in a scene should not be critical. As long as hue is present, the system should be able to use color differences to help locate boundaries. The predictions of the results are

straightforward: For blurred scenes, if color acts solely as a segmenter, the response biases for abnormally colored scenes should be greater than for monochrome scenes and should be just as strong as for normally colored scenes. We refer to this as the *segmentation hypothesis*.

Alternatively, color may directly contribute to the activation of scene gist if it is directly associated with scene category in memory. In this case, color effects were seen for the blurred scenes in Experiment 2 because when the quality of structure information was lessened, color information contributed directly to increase the activation of scene gist. Thus, if color plays a direct role in the activation of scene gist, then the effect of showing abnormal colors in a scene should adversely affect activation. We refer to this as the *color association hypothesis*. Whether the abnormal color scenes produce a cost that is equivalent to, less than, or greater than the cost produced by the monochrome scene cannot be predicted from this hypothesis because the proposed mechanisms of interference for each of these manipulations is different (i.e., no hues vs. misleading hues).

### Method

*Participants.* One hundred twenty Michigan State University undergraduates participated in this experiment for credit in an introductory psychology course.

*Apparatus and stimuli.* A subset of 300 scenes was randomly selected from the set used in Experiment 2. There were three versions of each scene: normal color, monochrome, and abnormal color. Figure 5 shows an example of a scene in each of the three color conditions. The methods used to produce the abnormal color version are identical to those used by Oliva and Schyns (2000). The photographs were transformed into  $L^*a^*b^*$  color space and the information for  $a^*$  and  $b^*$  channels was swapped and inverted, thereby producing hues opposite each other within the  $L^*a^*b^*$

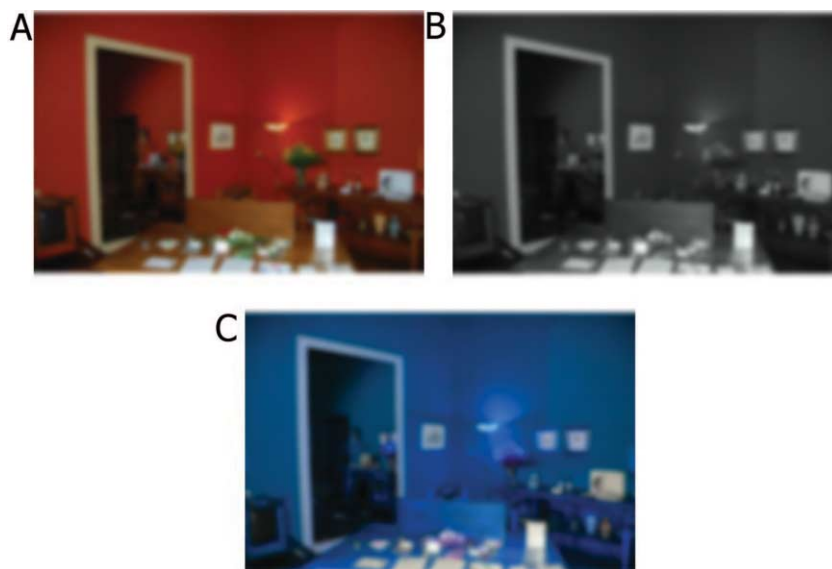


Figure 5. Example of stimulus color conditions used in Experiment 3: (A) normal color condition, (B) monochrome color condition, and (C) abnormal color condition.

Table 2  
Mean (Standard Deviation) Proportion of “Yes” Responses for Experiment 3

Condition	Colored	Abnormal	Monochrome
250 ms			
Consistent	.86 (.13)	.78 (.17)	.82 (.15)
Inconsistent	.18 (.14)	.22 (.17)	.20 (.16)
100 ms			
Consistent	.78 (.15)	.67 (.21)	.71 (.17)
Inconsistent	.30 (.17)	.31 (.18)	.30 (.17)
80 ms			
Consistent	.69 (.18)	.60 (.19)	.64 (.19)
Inconsistent	.33 (.18)	.32 (.18)	.16 (.17)
50 ms			
Consistent	.49 (.22)	.45 (.21)	.48 (.21)
Inconsistent	.34 (.19)	.34 (.19)	.34 (.20)
20 ms			
Consistent	.35 (.25)	.33 (.25)	.35 (.25)
Inconsistent	.32 (.24)	.31 (.22)	.31 (.24)

color space. The scenes were then blurred using the same methods described in Experiment 2.

*Design.* Three factors were varied for each photograph: color, target, and duration ( $3 \times 2 \times 5$ ).

*Procedure.* The procedure was similar to that of Experiment 2, with the exception that participants were shown blurred scenes in three color conditions (normal, abnormal, and monochrome).

**Results**

The analyses are organized as follows. First, the response bias patterns were analyzed for all color conditions. Then, the color effects were analyzed by comparing the difference scores across

color conditions as described in previous experiments. The means for all conditions are presented in Table 2.

An omnibus ANOVA with color, target, and duration ( $3 \times 2 \times 5$ ) as within-subject factors revealed that there was a significant main effect of color,  $F(2, 238) = 12.52, p < .01, MSE = 0.023$ , partial  $\eta^2 = .10$ ; target,  $F(1, 119) = 1,265.66, p < .01, MSE = 0.066$ , partial  $\eta^2 = .91$ ; and duration,  $F(4, 476) = 83.92, p < .01, MSE = 0.054$ , partial  $\eta^2 = .41$ . In addition, the interactions between color and target,  $F(2, 238) = 15.63, p < .01, MSE = 0.027$ , partial  $\eta^2 = .12$ , and between target and duration,  $F(4, 476) = 332.02, p < .01, MSE = 0.030$ , partial  $\eta^2 = .74$ , were significant. No other interactions were significant.

Of specific theoretical interest was the effect of color between the abnormal condition and the other two color conditions. To further simplify these analyses, we calculated difference scores by subtracting the inconsistent target means from the consistent target means for each color and duration condition combination. Figure 6 shows the difference scores plotted as a function of color and duration conditions.

An omnibus ANOVA on the differences with color and duration ( $3 \times 5$ ) as within-subject factors revealed that there was a significant main effect of color,  $F(2, 238) = 15.62, p < .01, MSE = 0.053$ , partial  $\eta^2 = .12$ , and duration,  $F(4, 476) = 331.98, p < .01, MSE = 0.059$ , partial  $\eta^2 = .74$ . The interaction between color and duration was not significant,  $F(8, 712) = 1.03, p > .10, MSE = 0.048$ , partial  $\eta^2 = .01$ .

To investigate the main effect of color and to test the predictions of the competing hypotheses, we conducted two simplified ANOVAs comparing normal color with abnormal color scenes and monochrome with abnormal color scenes. For each of these simplified analyses of color effects, planned comparisons were made between the color conditions for each duration condition ( $\alpha = .005$  for each comparison).

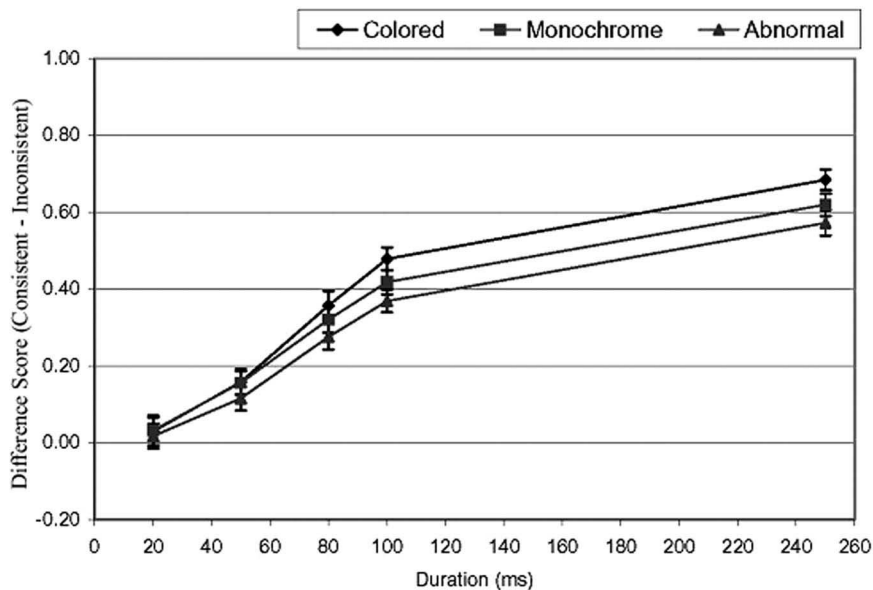


Figure 6. Difference scores (consistent means minus inconsistent means) for the normal color, monochrome, and abnormal color scene conditions in Experiment 3. Error bars represent standard errors of the mean.

First, the abnormal color scenes were compared with the monochrome scenes. A simplified ANOVA showed that there was a main effect of color,  $F(1, 119) = 8.46, p < .05, MSE = 0.052$ , partial  $\eta^2 = .07$ , and a main effect of duration,  $F(4, 476) = 220.25, p < .01, MSE = 0.054$ , partial  $\eta^2 = .65$ , but no interaction,  $F(4, 476) < 0.10, ns, MSE = 0.049$ , partial  $\eta^2 = .002$ . Planned comparison  $t$  tests were conducted and showed no effect of color across duration conditions: 250 ms,  $t(119) = 1.48, ns$ ; 100 ms,  $t(119) = 1.95, ns$ ; 80 ms,  $t(119) = 1.55, ns$ ; 50 ms,  $t(119) = 1.51, ns$ ; 20 ms,  $t(119) = 0.32, ns$ .

Second, a simplified ANOVA was conducted to compare the effects of normal and abnormal color scenes. The analysis revealed a significant main effect of color,  $F(1, 119) = 32.64, p < .01, MSE = 0.051$ , partial  $\eta^2 = .22$ , and duration,  $F(4, 476) = 258.45, p < .01, MSE = 0.052$ , partial  $\eta^2 = .69$ , and a significant interaction,  $F(4, 476) = 2.71, p < .05, MSE = 0.047$ , partial  $\eta^2 = .02$ . Further planned comparisons revealed an advantage of the normal color over the abnormal color for certain duration conditions. There was a higher response bias for the normal color than for the abnormal color for 250 ms,  $t(119) = 4.04, p < .001$ ; 100 ms,  $t(119) = 4.39, p < .001$ ; and 80 ms,  $t(119) = 3.14, p < .005$ . There was no difference between the normal color condition and the abnormal color condition at durations of either 50 ms,  $t(119) = 1.69, ns$ , or 20 ms,  $t(119) = -0.30, ns$ .

One interesting aspect about creating abnormal color scenes by swapping and inverting the color hues is that it is possible that some scenes in the abnormal color condition would still fall within or close to their normal range of possible color hues. For example, many man-made scenes have a greater variety of possible colors than natural scenes (Oliva & Schyns, 2000). Because the normal hues have an association to the correct color cue and because some of these abnormal color scenes may be close to the normal range,

it is possible that a difference exists in the amount of interference that is produced depending on how different the abnormal colors are from a scene's normal colors.

To further explore the degree to which abnormal colors can differentially interfere with the activation of scene gist, we conducted a secondary analysis in which the abnormal scenes were divided into high- and low-abnormal categories. The abnormal color scenes were rated by a separate group of participants ( $n = 20$ ) on a 7-point Likert-type scale indicating the strangeness of the colors for each particular scene (1 = *normal*; 7 = *extremely strange*). Participants were shown the 300 images from this experiment; half were shown in normal color and half were in abnormal color. Each participant saw each scene once, and the color condition for each scene was counterbalanced across participants.

The abnormal scenes with an average rating between 1.1 and 5.0 were designated as low abnormal, and those scenes with an average rating of higher than 5.0 were designated as high abnormal. There were 165 low-abnormal scenes (average rating: 3.77) and 135 high-abnormal scenes (average rating: 5.84). Figure 7 shows an example of a low- and high-abnormal scene representing the average rating for each range.

Response bias difference scores were calculated for each abnormal group for all duration conditions. Figure 8 shows the difference scores for normally colored, low-abnormal, and high-abnormal scenes as a function of duration. A within-subject ANOVA was conducted for the normal color and abnormal color conditions with color and duration ( $3 \times 5$ ) as factors and revealed that there was a significant main effect of color,  $F(2, 212) = 19.00, p < .01, MSE = 0.083$ , partial  $\eta^2 = .15$ , and duration,  $F(4, 424) = 171.64, p < .01, MSE = 0.101$ , partial  $\eta^2 = .62$ , and a significant interaction,  $F(8, 848) = 2.51$ ,



*Figure 7.* Example of low-abnormal and high-abnormal stimuli used in Experiment 3. Panels A and B show a scene in its normal and low-abnormal colors, respectively. The average rating for the group was 3.77. Panels C and D show a scene in its normal and high-abnormal colors, respectively. The average rating for the group was 5.84.

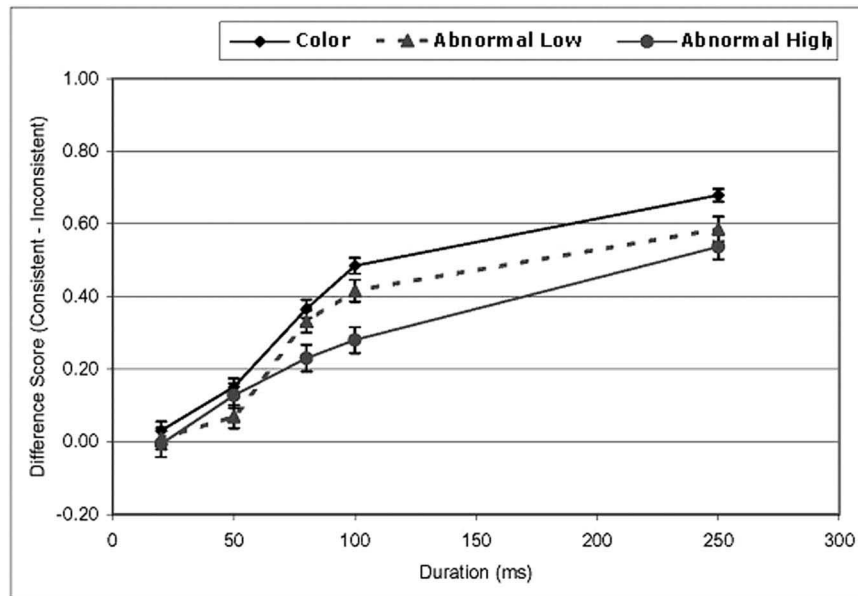


Figure 8. Difference scores (consistent means minus inconsistent means) for the colored, low-abnormal, and high-abnormal scene conditions. Error bars represent standard errors of the means.

$p < .01$ ,  $MSE = 0.089$ , partial  $\eta^2 = .03$ . Post hoc comparison of color conditions (using Tukey's least significant difference test [LSD],  $p < .05$  collectively) revealed that normally colored scenes produced a larger response bias than low- and high-abnormal scenes. In addition, high-abnormal scenes had significantly smaller bias than low-abnormal scenes. In order to look more closely at the onset of the abnormal differences, we conducted post hoc comparisons between normally colored conditions and each of the abnormal conditions at each duration condition. Post hoc Tukey's LSD comparisons revealed that the response bias was greater for the normally colored scenes than for low-abnormal scenes for durations of 100 ms and longer and high-abnormal scenes for durations of 80 ms and longer. Taken together, the results indicate that the greater the abnormal coloration of the scenes is, the more interference with the activation of the scene gist is produced, and that the closer to normal the abnormal colors are, the later the onset of the interference is.

### Discussion

Experiment 3 investigated the role of color when the structural information of scenes has been degraded. The segmentation hypothesis posits that color aids in the extraction of scene regions and assumes that the structural information of a scene is responsible for the activation of the scene gist. In contrast, the color association hypothesis proposes that color is directly associated with scene gist and so plays a direct role in the activation of scene gist. Therefore, when scene hues are changed (as in abnormal color scenes), the correct scene gist cannot be activated through color and the hues that are present could activate an incorrect scene gist. The pattern of results for the comparisons of both the abnormal scenes with the monochrome scenes and the abnormal color scenes with the normal color scenes supported the color association hypothesis. The

bias effect was greater for the normal scenes than for the abnormal scenes for durations of 80 ms and longer, and there was no interaction between the abnormal color and monochrome scenes, although the significant main effect of color suggests that these two color conditions interfere with the activation of scene gist differently, with abnormal colors producing a slightly greater interference. It is clear from this pattern that hue directly contributes to the activation of scene gist. However, it is important to note that although these results support the color association hypothesis, they do not necessarily rule out the segmentation hypothesis. This pattern of results could arise from both the cost of interference from the abnormal hues and the benefit from improved segmentation, with the overall cost being higher. Future investigation will be required to completely rule out the role of color as a segmenter of scene regions.

It was possible with the current set of stimuli to investigate whether abnormal hues that are considered to be highly abnormal produced a greater cost than those that are considered to be closer to normal. A post hoc analysis of the abnormal scenes was conducted in which the abnormal scenes were rated as high or low abnormal by a separate group of participants. These results suggested that the amount of interference produced by abnormal scenes depends on how much the abnormal hues departed from the hues considered normal for a particular scene. The bias effect was greater for the low-abnormal scenes than for the high-abnormal scenes. Therefore, it seems that if the hues are closer to the normal variation of hues for a given scene, the correct scene gist is activated more readily than if those hues are considered extremely abnormal.

It is also interesting to note that the effect of color is not limited to natural scenes. As mentioned earlier, the scenes used in the present study were mostly man-made, with a few scenes containing both natural and man-made characteristics (e.g., parks, play-

grounds, and gardens). The results of the current study show that the effect of color is not limited to natural scenes. This result was predicted (but not shown) by a previous study on color effects on scenes (Oliva & Schyns, 2000). Oliva and Schyns argued that although their measure of response time was probably not sensitive enough to detect color effects in man-made scenes, in order for such an early effect of color on natural scenes to be present, color must have an early effect on both scene types. In a later study, Goffaux et al. (2005) measured event-related potentials (ERPs) evoked by natural scenes in various color conditions. The results showed that for the four categories of natural landscapes tested, the pattern of activation in certain areas supported the findings of Oliva and Schyns (2000). Although ERP is a more temporally sensitive measure of the onset of scene gist activation than response time, only natural (i.e., not man-made) scenes were tested using this method. The results of the present study further support the notion that color has an effect on the processing of scene gist across all scene types as hypothesized in earlier studies, and that the size of this effect is dependent on other factors besides scene category (natural vs. man-made).

Moreover, the effect of the continuum of color abnormality on the activation of scene gist sheds new light on the concept of color diagnosticity. Although the scenes in the current study were not measured according to their diagnosticity (in terms of how consistently they could be associated with a particular color or set of colors), the rating results show that for any particular scene (even one with high variation and, therefore, considered to have low color diagnosticity), the variation of acceptable hues is limited. For instance, kitchens are often colored in a wide range of color hues; however, it is still possible to depict a kitchen in a hue that falls outside its normal range, such as fluorescent yellow. Although previous studies have deemed only natural scenes as color diagnostic (Goffaux et al., 2005; Oliva & Schyns, 2000), these results suggest that as long as the hue change falls within the normal variation, color supports the activation of all scene types.

### General Discussion

The three experiments reported in this study investigated how long visual information must be available to allow activation of scene gist, whether color plays a role in this activation, and if so, what role color plays when its effects are seen. To investigate these questions, we introduced the contextual bias paradigm to assess the degree to which sufficient information has been acquired for the activation of scene gist by measuring the bias created to affirm objects that are consistent and to disconfirm objects that are inconsistent with that gist. This paradigm operationalizes scene gist activation at the conceptual level as a response bias to objects likely or unlikely to appear in a scene. In this way, the contextual bias paradigm avoids the problem of subjectivity in labeling images and is still able to reveal the activation of the conceptual representation of the scene.

In each experiment, the strength of response bias was measured as a function of exposure duration. In Experiment 1, the strength of the bias effect increased with an increase in the duration of scene presentation. The response bias was significant after 42 ms of exposure, with greater affirmative responses to consistent than to inconsistent target objects. The difference

between the responses to each type of target object increased monotonically for exposure durations from 42 ms to 100 ms and reached asymptotic levels between 100 ms and 250 ms. These results suggest that the activation of scene gist builds up over time as more visual information is acquired. Experiment 2 showed that removal of color information from normal scenes produced no effect on the activation of scene gist. However, when structural information was degraded (i.e., the scene was blurred by removing high-frequency information) and so the efficiency with which the visual information could be extracted was slowed, color had an effect on the activation of scene gist. Blurred scenes that were presented in color produced a more pronounced bias effect than those that were presented in monochrome. These results suggest that color does play a role in the activation of scene gist but that it is dependent on the quality of structural information.

In Experiment 2, color may have acted as either a segmenter of scene regions or a direct cue for scene gist. These hypotheses suggest two different architectures for how color information may contribute to the activation of scene gist. The segmentation hypothesis assumes that color contributes only when structure is degraded, because it helps recover some of the boundary edges that are lost when high spatial frequency information is removed. This architecture would involve an indirect route for color to scene gist through structure and would mean that only structure can directly activate scene gist. Alternatively, the color association hypothesis assumes that color itself is associated with scene gist and therefore can act as a direct cue. In this case, the architecture implied would be that both structure and color are associated with scene gist and each can contribute directly to its activation.

In Experiment 3, the role of color in the activation of scene gist was explored using abnormally colored scenes. Abnormally colored scenes support segmentation, but the colors have no association with the gist. The results from this experiment strongly supported the color association hypothesis. Response biases for the abnormal scenes were significantly smaller than for the normal color scenes and did not differ significantly from monochrome scenes.

Additionally, the wrong hues could have potentially cued other scene gists and thus interfered with the activation of the correct scene gist. These results also suggest that there is a continuum on which color can be facilitative or inhibitory depending on how close the colors resemble a matching prototype of the scene type. Future investigations into these questions may allow for further insight on how this continuum of color cues interacts with the activation or inhibition of the correct scene gist.

### *The Influence of Color*

As discussed in the introduction, the results of past studies regarding the role of color in the recognition of scenes have been highly variable (Delorme et al., 2000; Goffaux et al., 2005; Oliva & Schyns, 2000). The results from the present study show that the influence of color interacts with the influence of structure information. The findings also suggest that the architecture of the system provides a direct link between color information and the conceptual representation. Therefore, the contribution of color will depend on properties of the stimulus set and the

relative quality of color information (in terms of how strongly it can activate scene gist). The present study explains the inconsistency of previous findings as dependent on how the system uses color information. Studies examining scene perception have used stimuli that vary greatly in possible strategy use (e.g., identifying the scene vs. the object within the scene) and in the amount of structural information (e.g., the comparison of natural and man-made scenes). In the case of varying amounts of structure information, man-made scenes have more unique shapes that are consistently associated with a certain scene type (e.g., beds in bedrooms, counters in kitchens) as compared with natural scenes (e.g., trees can appear in parks, backyards, forests, etc.). When structural information is insufficient (as we have seen in the present study in the form of blurring), color contributions to the activation of the gist are seen.

The results of Experiments 2 and 3 also showed that color effects appeared relatively late in scene presentation (~80 ms). In past studies, the finding that there is no color contribution in the initial processing stages of visual stimuli was attributed to the slower processing of color. However, in order for a direct link between color and the scene gist information to be initially established, it must be the case that color has some unique contribution to the activation of scene gist above that of structure. Various hypotheses on visual perception work under the assumption that visual information is coded in separate channels containing luminance (i.e., shapes, boundaries, textures) and color information (paralleling the magnocellular and parvocellular pathways from the low-level visual processing areas through the ventral system and into the inferior temporal lobe). Within this framework of separate channels, and knowing that the contributions of color and structure interact, the delayed onset could alternatively be attributed to the unequal weighting of the channels, with one channel favored over another according to which type of information is more useful for a particular task. Therefore, the later onset of the color effect could be explained as being due to the time needed to change the relative strength of input from each channel. Thus, in order for a connection between color and scene gist to be initially formed and strengthened, it is reasonable to assume that under certain conditions the color channel provides more valuable information than structure. Thus, on this view the delayed onset of color observed here is merely a result of the experimental setup and implicit task demands. We are currently conducting experiments to investigate this possibility further.

#### *Scene Perception, Scene Memory, and Response Bias*

The current study also highlights another aspect about initial scene representations. The results show that after 250 ms (a relatively long exposure compared with the 100-ms durations of prior studies), certain biases that are present concerning the representation of scenes are not diminished but are in fact just as strong as those present after a 100-ms exposure. This could result because the response bias has in fact reached asymptotic levels. Alternatively, it could result from two opposing factors canceling each other and producing a net effect of no difference.<sup>5</sup> These two factors could be understood as follows: (a) As more information about the scene gist is available, there is an

increase in the bias response, and (b) as more time progresses and as more information is gathered about the scene, it becomes apparent that the target object is not in fact in the scene. Although we doubt that this latter effect would happen with the 150-ms time span reported in the current set of experiments, it is an interesting possibility that could occur if participants were allowed more time to visually explore and fixate different parts of the scenes. Thus, the investigative question shifts from how much the person is guessing on the basis of what he or she knows about the scene being viewed to how much is in fact known about the scene.

Past studies on memory for complex scenes have shown that people have an immense capacity for remembering previously viewed scenes, even when the distractors were mirror images of previously viewed scenes (Nickerson, 1965; Shepard, 1967; Standing, 1973). However, researchers have also found that memory for the details of a briefly presented scene is often based on the scene's associated semantic category and is affected by its associated schemas, not the actual details present in the scene (Brewer & Treyns, 1981; Hollingworth & Henderson, 1999; Intraub, 1981). Furthermore, studies examining memory for objects within scenes, focusing on phenomena such as change blindness (O'Regan, 1992; Rensink, 2000; Rensink, O'Regan, & Clark, 1997), incremental change (Hollingworth & Henderson, 2004), and inattention blindness (Mack & Rock, 1998; Simons & Chabris, 1999), demonstrate that online memory for specific details within a scene is not particularly accurate when there is insufficient time for attention to be directed to those details. Taken together, these studies suggest that analyzing the object content of the scene takes quite a bit of time (on the order of a few seconds at least) and suggest that even if viewing time were to be extended, the response bias would still be present. It would be interesting to investigate how the bias changes with extended exposure—whether it would reverse or remain at the asymptotic level that we found between 100 and 250 ms of exposure.

#### *Conclusions*

The present work examined the contribution of color (along with the influence of scene structure) on the rapid activation of scene gist. Previous research has yielded mixed results concerning the contribution of color in the initial identification of objects and scenes. As a result, researchers have been divided as to whether color plays an important role in the initial processing of visual stimuli. The current study found that color does contribute to the activation of scene gist and has a direct connection to the conceptual representation of scenes.

Exploring the possible mechanisms and architecture underlying the interaction between structure and color may help to guide our understanding of how the visual system takes incoming visual information and activates the appropriate conceptual representation. In this way, research into the junction between early perceptual processes and late cognitive processes may illuminate possible strategies (either explicit or implicit in the visual system dynamics) that are used for finding and using the most useful incoming information to process the current task most efficiently.

<sup>5</sup> We thank an anonymous reviewer for this point.

## References

- Biederman, I. (1988). Aspects and extensions of a theory of human image understanding. In Z. Pylyshyn (Ed.), *Computational processes in human vision* (pp. 370–428). New York: Ablex.
- Biederman, I., & Ju, G. (1988). Surface vs. edge-based determinants of visual recognition. *Cognitive Psychology*, *20*, 38–64.
- Brewer, W. F., & Treyns, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology*, *13*, 207–230.
- Burton, G. J., & Moorhead, I. R. (1987). Color and spatial structure in natural scenes. *Applied Optics*, *26*, 157–170.
- Delorme, A., Richard, G., & Fabre-Thorpe, M. (2000). Ultra-rapid categorization of natural scenes does not rely on colour cues: A study in monkeys and humans. *Vision Research*, *40*, 2187–2200.
- Edwards, R., Xiao, D., Keysers, C., Földiák, P., & Perrett, D. (2001). Color sensitivity of cells responsive to complex stimuli in the temporal cortex. *Journal of Neurophysiology*, *90*, 1245–1256.
- 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*, 1476–1492.
- Fabre-Thorpe, M., Delorme, A., Marlot, C., & Thorpe, S. J. (2001). A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. *Journal of Cognitive Neuroscience*, *13*, 171–180.
- Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General*, *108*, 316–355.
- Gegenfurtner, K. R. (2003). Cortical mechanisms of colour vision. *Nature Reviews Neuroscience*, *4*, 563–572.
- Gegenfurtner, K. R., & Kiper, D. C. (2003). Color vision. *Annual Review of Neuroscience*, *26*, 181–206.
- Gegenfurtner, K. R., & Rieger, J. (2000). Sensory and cognitive contributions of color to the recognition of natural scenes. *Current Biology*, *10*, 805–808.
- Goffaux, V., Jacques, C., Mouraux, A., Oliva, A., Schyns, P. G., & Rossion, B. (2005). Diagnostic colours contribute to the early stages of scene categorization: Behavioural and neurophysiological evidence. *Visual Cognition*, *12*, 878–892.
- Grossberg, S., & Mingolla, E. (1985). Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentations. *Perception & Psychophysics*, *38*, 141–171.
- 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.
- Hollingworth, A., & Henderson, J. M. (2004). Sustained change blindness to incremental scene rotation: A dissociation between explicit change detection and visual memory. *Perception & Psychophysics*, *66*, 800–807.
- Intraub, H. (1980). Presentation rate and the representation of briefly glimpsed pictures in memory. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 1–12.
- Intraub, H. (1981). Rapid conceptual identification of sequentially presented pictures. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 604–610.
- Livingstone, M. S. (1988, January). Art, illusion, and the visual system. *Scientific American*, *258*, 78–85.
- Livingstone, M. S., & Hubel, D. H. (1984). Anatomy and physiology of a color system in the primate visual cortex. *Journal of Neuroscience*, *4*, 309–356.
- Livingstone, M. S., & Hubel, D. H. (1988, May 6). Segregation of form, color, movement, and depth: Anatomy, physiology, and perception. *Science*, *240*, 740–749.
- Mack, A., & Rock, I. (1998). *Inattention blindness*. Cambridge, MA: MIT Press.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: Freeman.
- Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London, Series B*, *200*, 269–294.
- Metzger, R. L., & Antes, J. R. (1983). The nature of processing early in picture perception. *Psychological Research*, *45*, 267–274.
- Nickerson, R. S. (1965). Short-term memory for complex meaningful configurations: A demonstration of capacity. *Canadian Journal of Psychology*, *19*, 155–160.
- Oliva, A. (2005). Gist of a scene. In L. Itti, G. Rees, & J. Tsotsos (Eds.), *Neurobiology of attention* (pp. 251–256). San Diego, CA: Elsevier.
- Oliva, A., & Schyns, P. G. (1997). Course blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology*, *34*, 72–107.
- Oliva, A., & Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. *Cognitive Psychology*, *41*, 176–210.
- Oliva, A., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, *42*, 145–175.
- O'Regan, J. K. (1992). Solving the “real” mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology*, *46*, 461–488.
- Párraga, C. A., Breilstaff, G., & Troscianko, T. (1998). Color and luminance information in natural scenes. *Journal of the Optical Society of America A*, *15*, 563–569.
- Potter, M. C. (1975, March 14). Meaning in visual search. *Science*, *187*, 965–966.
- Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *2*, 509–522.
- Potter, M. C. (1999). Understanding sentences and scenes: The role of conceptual short-term memory. In V. Coltheart (Ed.), *Fleeting memories: Cognition of brief visual stimuli* (pp. 13–46). Cambridge, MA: MIT Press.
- Potter, M. C., & Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology*, *81*, 10–15.
- Price, C. J., & Humphreys, G. W. (1989). The effects of surface detail on object categorization and naming. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *41*, 797–827.
- Renninger, L. W., & Malik, J. (2004). When is scene recognition just texture recognition? *Vision Research*, *44*, 2301–2311.
- Rensink, R. A. (2000). The dynamic representation of scenes. *Visual Cognition*, *7*, 17–42.
- Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, *8*, 368–373.
- 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.
- Shepard, R. N. (1967). Recognition memory for words, sentences and pictures. *Journal of Verbal Learning and Verbal Behavior*, *6*, 156–163.
- Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattention blindness for dynamic events. *Perception*, *28*, 1059–1074.
- Spence, I., Wong, P., Rusan, M., & Rastegar, N. (2006). How color enhances visual memory for natural scenes. *Psychological Science*, *17*, 1–6.

- Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology*, *25*, 207–222.
- Tanaka, J. W., Weiskopf, D., & Williams, P. (2001). The role of color in high-level vision. *Trends in Cognitive Sciences*, *5*, 211–215.
- Thorpe, S., Fize, D., & Marlot, C. (1996, June 6). Speed of processing in the human visual system. *Nature*, *381*, 520–522.
- Torralba, A., & Oliva, A. (2002). Depth estimation from image structure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*, 1226–1238.
- Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network: Computation in Neural Systems*, *14*, 391–412.
- Vailaya, A., Jain, A., & Zhang, H. J. (1998). On image classification: City images vs. landscapes. *Pattern Recognition*, *31*, 1921–1935.
- Van Rullen, R., & Thorpe, S. J. (2001). The time course of visual processing: From early perception to decision-making. *Journal of Cognitive Neuroscience*, *13*, 454–461.

Received October 20, 2006

Revision received June 28, 2007

Accepted July 2, 2007 ■