Global image properties do not guide visual search

Michelle R. Greene Brigham and Women's Hospital, USA, & Harvard Medical School, USA

Jeremy M. Wolfe Brigham and Women's Hospital, USA, & Harvard Medical School, USA

While basic visual features such as color, motion, and orientation can guide attention, it is likely that additional features guide search for objects in real-world scenes. Recent work has shown that human observers efficiently extract global scene properties such as mean depth or navigability from a brief glance at a single scene (M. R. Greene & A. Oliva, 2009a, 2009b). Can human observers also efficiently search for an image possessing a particular global scene property among other images lacking that property? Observers searched for scene image targets defined by global properties of naturalness, transience, navigability, and mean depth. All produced inefficient search. Search efficiency for a property was not correlated with its classification threshold time from M. R. Greene and A. Oliva (2009b). Differences in search efficiency between properties can be partially explained by low-level visual features that are correlated with the global property. Overall, while global scene properties can be rapidly classified from a single image, it does not appear to be possible to use those properties to guide attention to one of several images.

Keywords: visual search, gist, scene recognition, global property, natural images, guided search, attention

Citation: Greene, M. R., & Wolfe, J. M. (2011). Global image properties do not guide visual search. *Journal of Vision, 11*(6):18, 1–9, http://www.journalofvision.org/content/11/6/18, doi:10.1167/11.6.18.

Introduction

A great deal has been learned about visual search using laboratory displays of isolated items on homogeneous backgrounds. In particular, we know that attention can be guided by basic visual attributes such as the color, orientation, and motion of the target (Wolfe, 1994). However, the efficiency of search in real-world scenes is unlikely to be explained by effective guidance by these attributes alone. Although such features are likely to play some role in guiding search in real-world scenes, they do not completely account for the efficiency of these searches. Search for objects in scenes seems to be very efficient (Wolfe et al., 2008; though measures of "efficiency" depend on the problematic process of counting the numbers of objects in scenes). In contrast, search through isolated objects is quite inefficient (Vickery, King, & Jiang, 2005) though guidance by basic features should be similar in the two situations.

Of course, scenes have attributes that random arrays of objects lack, such as lawful spatial layout. In this paper, we investigate the role of *global scene properties* (Greene & Oliva, 2009a, 2009b) in guiding search in images of real-world scenes. Global properties are features of scene environments that describe an image's spatial layout, affordances, or surface properties. Some examples include the *mean depth* of a scene, the degree to which an agent could *navigate* freely in the scene, the proportion of manufactured elements in the scene (*naturalness*), and the presence of movement, or *transience* in a scene's elements. Global properties are global in the sense that they cannot be accurately predicted from local regions in the image. They are attributes of an image, and not necessarily the entire visual field. In other words, in order to determine whether a certain scene affords navigation, it is necessary to take the entire layout of the image into account, independent of the size of the image. In contrast, a local task, such as object detection, can be done from only examining a small piece of the image. It is also of note that there can be multiple levels of global and local scene analyses. Consider a street scene containing a pedestrian. The scene's spatial layout is a global scene property, and the pedestrian in that scene is a local property. However, when considering the pedestrian, the shape of this person is more global than one of his parts (say an eye).

Human observers can very rapidly perceive global properties. For example, observers can reliably discriminate a panoramic, large-depth scene from an image showing a close-up view of a single surface even if the scene is presented for a mere 26 ms followed by a mask (Greene & Oliva, 2009b). Moreover, these properties are sufficient to predict a scene's basic-level category (e.g., beach, forest, etc.; Greene & Oliva, 2009a). Getting a rough description of a scene's spatial layout and surfaces could help preferentially allocate attention to regions of a scene likely to contain an object of interest, therefore reducing the effective set size of the scene. If this is a beach, we know where to look for deck chairs.

doi: 10.1167/11.6.18 Received January 10, 2011; published May 24, 2011 ISSN 1534-7362 © ARVO

The rapid extraction of global scene properties could occur in a number of different ways. Properties could be processed "preattentively" in the sense that is similar to color—a red item among green items will "pop-out" and can be detected in a time that is essentially independent of the number of green items. If global scene properties are processed in a similar manner, we would expect that the time required to determine the presence of a *natural* scene image would be independent of the number of *urban* scene images presented. Alternatively, it could be that observers are limited to the perception of one set of global properties at a time. Thus, observers might rapidly determine the depth, transience, and naturalness of *this* image but slowly find a property in one of several images if they might be required to shift their attention in order to determine the properties of another image. Experiment 1 tested these possibilities using visual search methods. Observers searched for a relatively small scene image with a target global property (e.g., high transience) among similarly sized distractor images without that property (e.g., low transience). For four global scene properties: naturalness, mean depth, navigability, and transience, we found that search was inefficient, arguing against an ability to process the global properties of multiple images in parallel.

Experiment 1

Methods

Materials

Both poles of each global property (e.g., high transience and low transience) were tested as targets. In order to ensure that observers were searching for a target property and not a particular image, 99 images exemplifying each global property pole were selected as targets. The large number of images ensured that participants could not overlearn a few images but had to search for the global property of interest. A larger image set also guarded against confounding low-level image features for just a few images. All scene images were used in Greene and Oliva (2009a, 2010) and had been rated by independent observers as belonging to a particular property pole. Images were in full color and were 256 by 256 pixels in size.

The experiment was run with MATLAB using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997).

Observers

Sixteen observers took part in Experiment 1. All were between the ages of 18–55, had normal or corrected-to-normal vision, and had no history of eye or muscle disorders. All provided informed consent and were compensated $10 for their time.

Design and procedure

Each of the four global properties (naturalness, mean depth, navigability, and transience) was tested in two experimental blocks, with each pole of a property tested as targets in an independent block for a total of eight blocks. In other words, in the two blocks testing *navigability,* highly navigable scenes were the targets in one block and highly non-navigable scenes were the targets in the other. In all blocks, distractor images were images from the opposite pole of the target property (for example, urban scenes for natural targets). Figure 1 shows examples of targets and distractors for all experimental conditions. The order of blocks was randomized and counterbalanced across participants.

Participants were seated in a dimly lit room, 57.4 cm from a 21-inch CRT monitor. At the beginning of each block, participants received an instruction screen describing the target global property pole and providing examples of target and non-target images. Following the instruction screen, observers completed 170 trials for each target, with the first 10 trials for practice. Targets were present in 50% of trials, and each trial had between 1 and 4 images. The center of each image was 3.2 degrees away from a central fixation point. The background was mid-gray. Images remained on the screen until observers gave a "present" or "absent" response with a button press. Observers were instructed to respond as quickly and accurately as possible for each trial. Performance feedback was given after every trial. The experiment lasted approximately 45 min.

Results

Trials with reaction times under 200 ms or over 4000 ms were discarded from analysis. One observer had >20% rejected trials and was discarded from analysis. For the 15 remaining observers, the rejected trials constituted less than 15% of total trials.

Reaction time

Figures 2a–2d show mean reaction time as a function of set size for each of the four global properties. A main finding is that search for each of the global properties was inefficient. All slopes were significantly greater than zero (all $t(14) > 2.75$, all $p < 0.05$), with a mean target-present slope of 52.3 ms/item (range 10.1–96.8) and a mean target-absent slope of 77.4 ms/item (range 19.2–138.2). There was a wide variation between target types. The urban–natural distinction yielded the most efficient search with *urban* among *natural* producing an average slope of 16 ms/item. *Mean depth* supported the least efficient search with *small* among *large depth* producing a slope of 128 ms/item for target-present trials. These differences

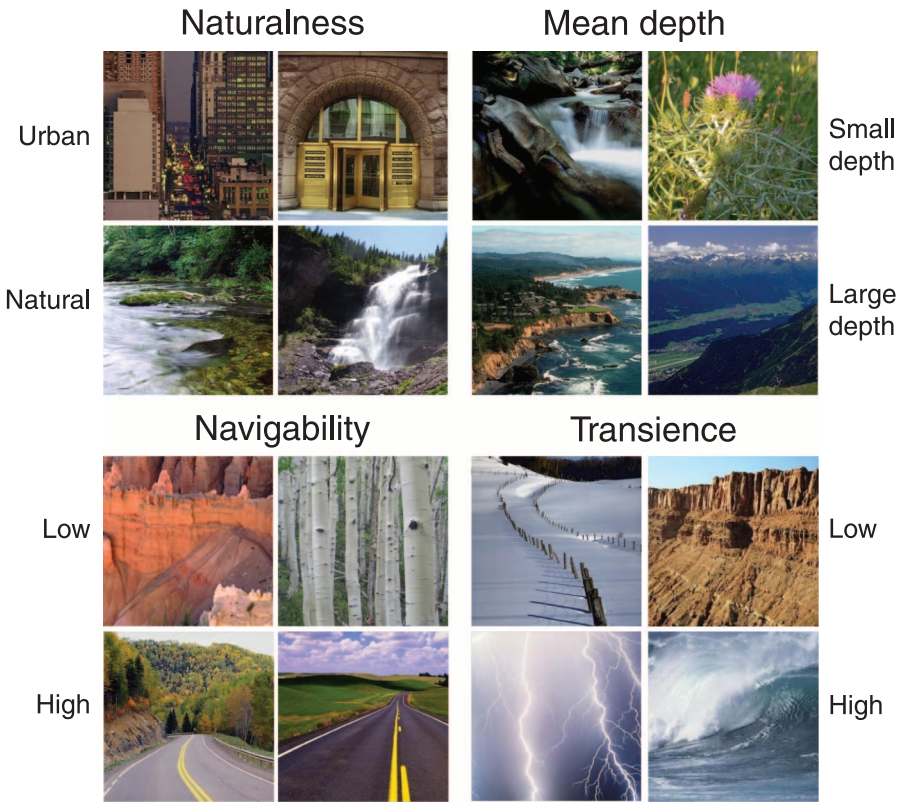

Figure 1. Sample scenes for each target global property pole. Each property is shown in a 2 × 2 square with 2 examples of the property's low pole shown in the top row and 2 examples of the property's high pole shown in the bottom row.

produced a significant main effect of property on reaction time ($F(3,42) = 50.4$, $p < 0.001$).

A second finding was that there were significant search asymmetries (Treisman & Souther, 1985). Search for urban targets among natural produced shallower slopes than natural among urban ($t(28) = 2.08$, $p < 0.05$), search for highly navigable images was more efficient than for non-navigable images ($t(28) = 5.09$, $p < 0.001$), and search for images with a high degree of transience was more efficient than search for low-transience images ($t(28) = 2.86$, $p < 0.01$). Search slopes did not significantly differ between large- and small-depth targets ($t(28) = 1.14$, $p = 0.27$).

Accuracy

Despite producing such inefficient search slopes, participants were very accurate in their searches. Overall d' was 3.31 (false alarm rate: 3.7%). By property pole, d' ranged from 3.1 for *large depth* to 4.4 for *natural*. There was no main effect of property pole on search accuracy ($F(1,14) < 1$).

Critically, set size did not significantly interact with accuracy ($F(9,126) < 1$), suggesting that participants did not become less accurate with increasing numbers of images.

Discussion

As noted earlier, global scene properties are readily identifiable in a fully attended display. Indeed, observers in Greene and Oliva's (2009b) study could reliably classify these specific images after 19–45 ms of viewing time. Nevertheless, this experiment demonstrates that such global properties do not support efficient search and, therefore, do not appear to be sources of guidance. The current results are not in dispute with Greene and Oliva (2009b). The ability to rapidly classify a single, attended property does not imply that the property should guide search. To take a simple example, the digits "2" and "5" can each be rapidly identified in isolation, but search for a 2 among a field of 5s leads to inefficient search (Kwak, Dagenbach, & Egeth, 1991). Furthermore, although the categorization of a single object can be very rapid (Grill-Spector & Kanwisher, 2005), search for a particular object among others is not efficient (Biederman, Blickle, Teitelbaum, & Klatsky, 1988; Vickery et al., 2005). Likewise, search for material type is inefficient (Wolfe & Myers, 2010) despite rapid classification of single materials at the center of attention (Sharan, Rosenholtz, & Adelson, submitted for publication).

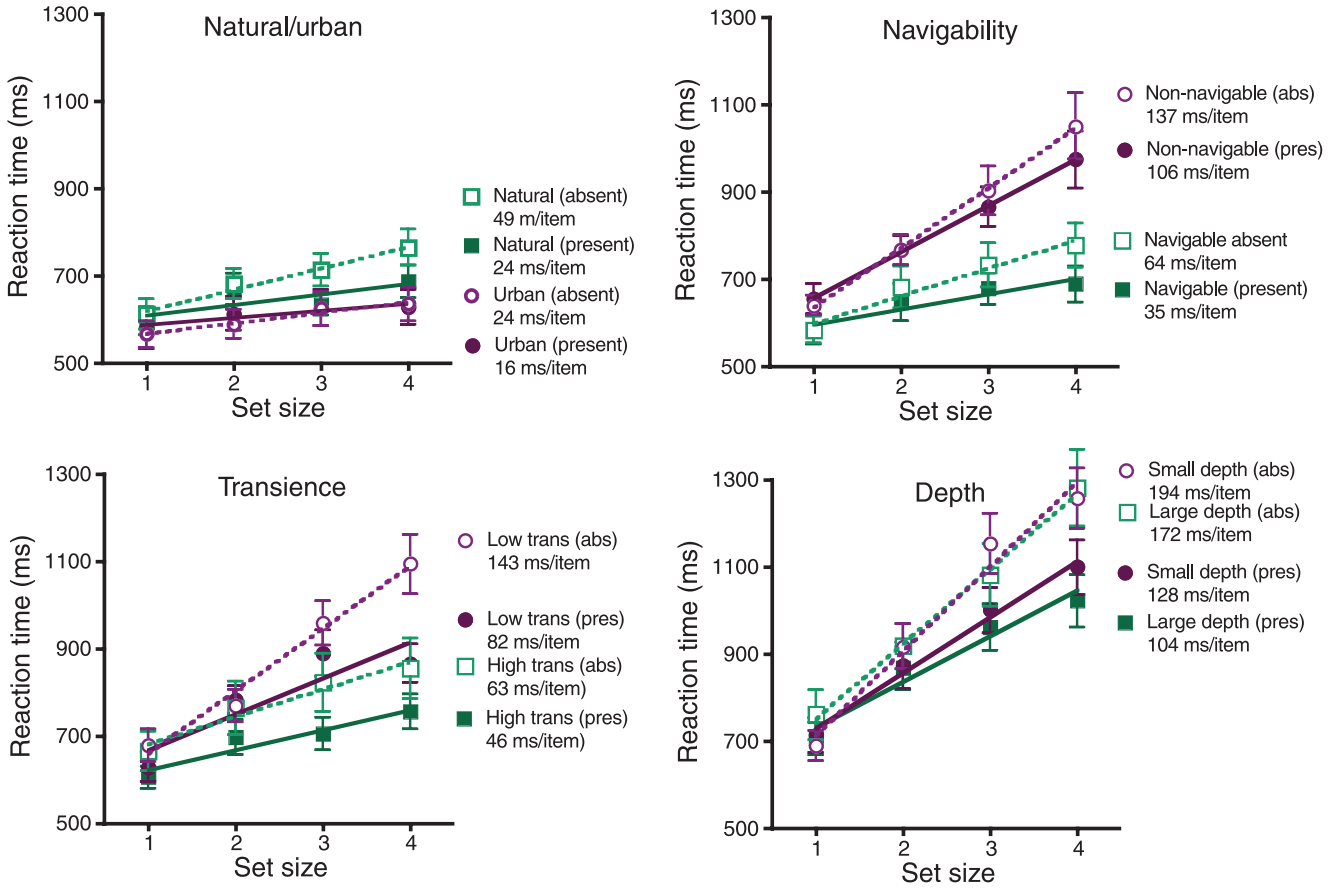

Figure 2. Reaction time as a function of set size for the four global properties: (a) Naturalness, (b) navigability, (c) transience, and (d) mean depth. Circles show the case where one pole served as target. Squares show the other pole. Target-present trials are represented with solid lines and target-absent trials with dashed lines. Error bars represent ±1 *SEM*.

We cannot attribute search inefficiency to a failure to understand the global scene properties. The high level of accuracy in the search shows that once an image was attended, the properties were perceived and then correctly classified. Nor is the inefficiency due to a need to fixate each item in order to categorize it. Given 3–4 fixations per second, we do not begin to suspect a need for fixation until target-present slopes are over 100 ms/item and this only occurred for the *mean depth* conditions.

The inefficiency of global property searches suggests that such properties cannot be extracted in parallel from multiple images. In this case, what determines the rate of global property search? One hypothesis would be that the slope of the RT × set size function should be related to the time required to identify a global property in a single scene. An estimation of this time (viewing duration required for 75% correct classification) was made by Greene and Oliva (2009b). The relationship of the identification time and the average slope is shown in Figure 3. Although a more stringent test of this relationship would compare identification times and slopes for the same, individual observers, these between-observer, averaged results suggest that search slopes are not strongly related to the time required to identify each item ($r = -0.10$, $p = 0.90$). Identification of *mean depth* conditions was not much slower than identification of *natural* and *urban* scenes. However, search in *mean depth* conditions yielded slopes many times less efficient than search in *naturalness* conditions.

Nor is it obvious how to explain the search asymmetries. It is easier to find urban among natural, highly navigable among non-navigable, and highly transient among static than vice versa. Were one global property pole to support efficient search, then we might follow the argument that it is easier to find the presence of this "basic feature" than to find its absence (Wolfe, Klempen, & Dahlen, 2000). For example, it is easier to find a moving stimulus among static distractors than vice versa because it is easier to detect the presence of motion than its absence (Dick, Ullman, & Sagi, 1987). However, this logic does not hold when the easier of the pair of search slopes is inefficient. In this case, all that can be said is that it is easier to search through one type of scene when it is the distractor than the other. In this case, that would mean

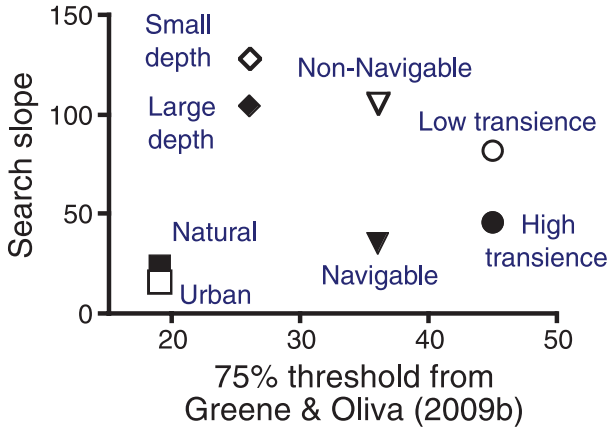

Figure 3. Average search slope from Experiment 1 as a function of 75% threshold for identifying a global scene property from Greene and Oliva (2009b). Greene and Oliva only tested one of the two poles in each category. Those values are shown with filled symbols. The other pole is shown with open symbols.

that search for urban is easier than search for natural because it is easier to reject natural stimuli when looking for urban than vice versa.

More striking than the search asymmetries are the differences between global scene properties. In particular, search for urban and natural targets produced search slopes that were shallower than those produced by other property targets. In Experiment 2, we examined the extent to which low-level image features account for the relative ease of search for natural and urban scenes.

Experiment 2

Although the search slopes for natural and urban scenes were not particularly efficient, they were markedly more efficient than the other searches of Experiment 1. With complex stimuli such as scene images, the suspicion must always be that some more basic attribute is driving performance. To give a trivial example, if all natural scene images were green and all urban scene images were red, it would be unsurprising and uninformative to find efficient search for natural among urban. No such blatant confound exists, but Torralba and Oliva (2003) have demonstrated that natural and urban scenes can be distinguished on the basis of their global amplitude spectra alone. Kaping, Tzvetanov, and True (2007) showed that adaptation to spectral statistics from natural and urban images produced robust aftereffects in the perception of natural and urban scenes, and Joubert, Rousselet, Fabre-Thorpe, and Fize (2009) demonstrated that removal of diagnostic Fourier amplitude information diminished accuracy and increased reaction time on a rapid natural–urban categorization task. Accordingly, in Experiment 2, we tested the hypothesis that global amplitude spectra supported the comparatively easy search for natural and urban scenes. To remove this cue, we presented observers with images whose amplitude spectra were the average of all natural and urban scenes from Experiment 1 (mean amplitude condition). In a separate block, to test the sufficiency of the global amplitude spectrum cue, we presented observers with phase-randomized images, containing *only* the amplitude spectra of the originals. The logic of Experiment 2 is given as follows: if the Fourier amplitude spectrum is necessary (or, at least, helpful) in efficient search for natural among urban scenes and vice versa, then search efficiency will be diminished for the mean amplitude condition relative to the conditions using color and grayscale images with normal amplitude spectra in Experiments 1 and 2. If Fourier amplitude, by itself, is sufficient to support this efficient search, then phase-randomizing the stimuli should not markedly disrupt search even though it turns scenes into seemingly content-free textures.

Methods

Materials

All stimuli were created from the natural and urban scenes used in Experiment 1. Four groups of images were used: full-color original images, the same images presented in grayscale, phase-randomized images, and images whose amplitude spectrum had been replaced with the average amplitude spectrum from all 198 natural and urban scenes. Examples of all four groups are shown in Figure 4.

Observers

Nine observers took part in Experiment 2. All were between the ages of 18–55, had normal or corrected-to-normal vision, and had no history of eye or muscle disorders. All provided informed consent and were compensated $10 for their time. None of the participants from Experiment 2 had taken part in Experiment 1.

Design and procedure

Participants completed eight experimental blocks, searching for natural scene targets in 4 blocks and urban scene targets in the other 4. Each observer viewed each of the four types of images in separate blocks: color, grayscale, phase randomized, and amplitude averaged. As in Experiment 1, distractor images were from the opposite global property pole as the target (urban for natural and vice versa). The order of conditions was randomized and counterbalanced across participants.

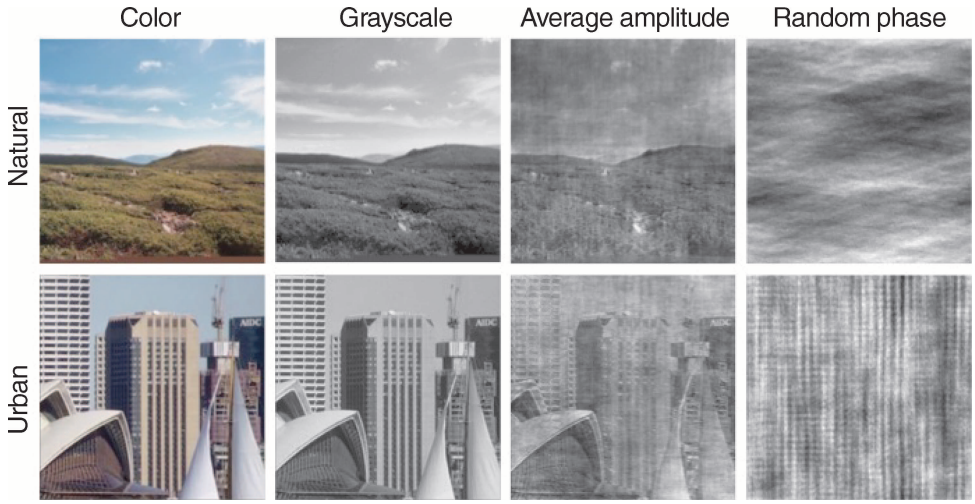

Figure 4. Examples of (top) natural and (bottom) urban images used as targets for Experiment 2.

The procedure for Experiment 2 was otherwise identical to Experiment 1.

Results and discussion

Trials with reaction times under 200 ms or over 4000 ms were discarded from analysis; 1.3% of total trials were rejected (less than 10% of trials from each observer).

Reaction time

Figure 5 shows target-present reaction times as a function of set size for all conditions of Experiment 2. Two points can be made here. First, Experiment 1 can be replicated. Search for natural scene images among urban or vice versa is not efficient even if it is more efficient than search for the other global scene properties tested in the first experiment. Second, the grayscale conditions shows that the relatively efficient search for urban and natural scene images in the first experiment was not based on a color signal ($t(34) = 1.45$, $p = 0.15$). If anything, grayscale images produced slightly more efficient search with urban grayscale targets producing a slope of just 11 ms/item. This finding is in agreement with others showing little contribution of color to rapid scene understanding (Delorme, Richard, & Fabre-Thorpe, 2000; Fei-Fei, Van Rullen, Koch, & Perona, 2005), although the use of color depends on stimuli and task (Castelhano & Henderson, 2008; Oliva & Schyns, 2000).

While the color signal does not seem to have been of much use, observers used the Fourier amplitude signal. When Fourier amplitude information was rendered non-diagnostic in the average amplitude condition, search slopes increased significantly to 55.4 ms/item ($t(34) = 2.28$, $p < 0.05$), suggesting that amplitude plays a role in

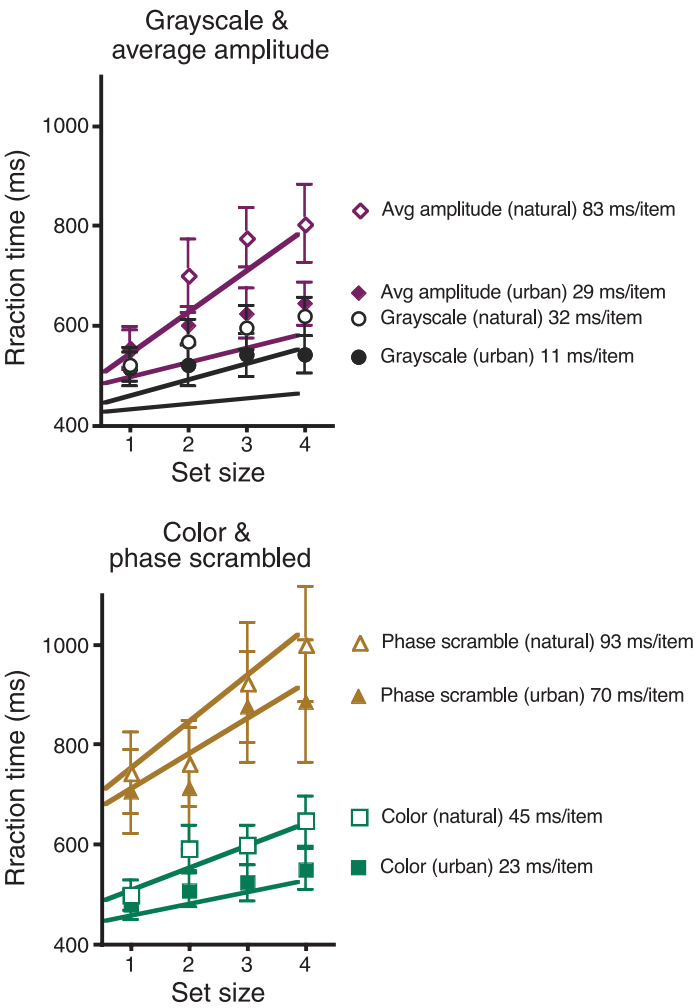

Figure 5. Target-present reaction time as a function of set size for all conditions in Experiment 2. Natural scene targets are shown with open symbols and urban scene targets are shown with closed symbols. Error bars represent ±1 *SEM*.

	Color		Grayscale		Mean amplitude		Phase random	
	Natural	Urban	Natural	Urban	Natural	Urban	Natural	Urban
Miss	3.7	2.6	3.2	2.2	5.0	3.2	22.9	20.2
F.A.	1.9	2.1	2.2	2.2	2.5	3.2	26.6	32.4

Table 1. Miss and false alarm rates (percentage) for all conditions in Experiment 2.

search for natural and urban scenes. The relative efficiency of naturalness search, when compared to the other global properties of Experiment 1, is partially mediated by differences in the global amplitude spectra of natural and urban images, as search became less efficient when this information equated between targets and distractors in the average amplitude condition, similar to the results found by Joubert et al. (2009) for fully attended items. Obviously, the amplitude spectrum was not the only usable signal. If it had been the necessary signal, then equating it across urban and natural scenes would have made search impossible.

Search for phase-randomized urban and natural scenes was most inefficient, with an average slope of 64.1 ms/item. Furthermore, the intercept for this condition was larger than that of the color and grayscale conditions ($t(34) = 3.43$, $p < 0.01$), showing that search was slowed overall in this condition. Note that the phase-randomized stimuli lack spatial layout and recognizable objects, suggesting that scene and object recognition mechanisms might be the basis for search when amplitude spectrum information is removed.

In fact, it might seem surprising that participants could perform that phase-scrambled search condition at an above-chance level at all. Without phase information, all semantic information is destroyed, and the images look like textures rather than scenes (see Figure 4). Indeed, the phase spectrum of natural images has long been shown to be more important than the amplitude spectrum for recognizing images (Piotrowski & Campbell, 1982). However, the amplitude spectra information retained in these images has been shown to be sufficient for computational models to classify scenes as natural or urban (Oliva & Torralba, 2001; Torralba & Oliva, 2003) and may contribute to the rapid detection of naturalness in fully attended images (Greene & Oliva, 2009b; Joubert et al., 2009; Joubert, Rousselet, Fize, & Fabre-Thorpe, 2007). Apparently, our observers could make use of that information to categorize stimuli well enough to perform at above-chance levels.

As in Experiment 1, urban targets produced shallower search slopes (24.4 ms/item) than natural (60.9 ms/item) in Experiment 2 ($t(70) = 3.54$, $p < 0.01$).

Accuracy

Overall performance was excellent, as in Experiment 1, with an overall d' of 4.01. This level of performance is similar to that seen in the natural and urban targets in Experiment 1 ($d' = 4.28$). The accuracy was not uniform over conditions, however. There was a significant effect of condition in this experiment ($F(3,24) = 67.3$, $p < 0.001$). As shown in Table 1, while color, grayscale, and average amplitude conditions had similar levels of performance (miss and false alarm rates of 2–5%), observers were substantially less accurate in searching for phase-randomized images (miss and false alarm rates of 20–32%, $d' = 1.5$).

There was no significant effect of set size on d' ($F(3,24) = 1.36$, $p = 0.28$), indicating that, as in Experiment 1, accuracy did not decrease with increasing numbers of distractors. Furthermore, blocks of natural target trials produced no more errors than blocks of urban target trials ($F(1,8) = 1.26$, $p = 0.29$).

General discussion

Overall, we have shown that global scene properties do not support efficient visual search. Differences in efficiency between global properties cannot be explained by color, but global Fourier amplitude spectrum appears to play a role.

The large deficits we have observed in performing global property detection with increasing numbers of scenes suggests that classification of global properties may be limited to one image at a time. In some other natural scene tasks such as animal detection, there is little performance decrement when a second image is added (Fei-Fei et al., 2005; Rousselet, Fabre-Thorpe, & Thorpe, 2002). However, in those tasks, performance does suffer at set sizes above two (Rousselet, Thorpe, & Fabre-Thorpe, 2004). Drewes, Trommershauser, and Gegenfurtner (2011) argue that animal detection can occur over at least 8 locations in the same scene, simultaneously, though this may be showing that animal detection, like the properties used in the present study, is a global process within a single scene. Van Rullen, Reddy, and Koch (2004) found inefficient visual search for scenes with animals with set sizes up to 16 images (40 ms/item). In the present data, there is a cost even for the second scene as can be seen in the RT differences between trials with set sizes 1 and 2.

Prior work has shown that global scene properties such as mean depth, navigability, naturalness, and transience can be rapidly classified by human observers in fully

attended displays (Greene & Oliva, 2009b). This rapidly extracted information may be used to help get semantic category information from a scene (Greene & Oliva, 2009a, 2010). From there, the contextual information provided by the semantic category may then help guide search for objects in scenes (Neider & Zelinsky, 2006; Torralba, Oliva, Castelhano, & Henderson, 2006). One could imagine a "global precedence" effect (Navon, 1977) in which a set of global scene properties is processed first and guides subsequent processing. We have argued elsewhere for a dual pathway account in which scene information guides search in collaboration with basic attributes such as color or motion (Wolfe, Vo, Evans, & Greene, 2011). A search for a friend in a red coat would be guided by scene context and by redness at the same time.

The results of our two experiments indicate that the first step in this chain, the act of rapid classification of global scene properties, may be limited to one scene at a time. In our normal interactions with the world, it may not be much of a handicap if search for global properties is generally inefficient. After all, while it may be valuable to rapidly understand the current scene, it is rarely important, outside of the laboratory, to rapidly understand two or more scenes at the same time.

The fact that we are, typically, in one scene at a time might lead one to ask if the present experiments have any "ecological validity." The task is, after all, artificial on several levels. Observers are making decisions about multiple images. Those images are very small compared to the immersive nature of real-world environments. Finally, localizing (in addition to detecting) a path in a scene seems to be critically important in our interactions in the world.

To begin with the last issue, visual search experiments tend to produce similar results with localization or presence/absence responses (e.g., Saarinen, 1996). While it is true that the images used in this study occupied only a small piece of the field, the same is true of almost any study of scene perception except those using immersive stimuli. The sizes used here are the same as the sizes used in Greene and Oliva (2009a, 2009b). Moreover, scene properties have been shown to be quite resistant to changes in scale (Torralba, 2009) even if the spatial frequency content is changing differently when you shrink a picture than when you move farther away (Loftus & Harley, 2005). Finally, there are instances when we do confront multiple scenes: looking for a particular photo among vacation pictures or seeking out a specific painting in an art gallery. Returning to the issue of localization, one might suppose that in these cases, one would like to know *where* the navigable areas are in the scene in front of you. It would be interesting to ask observers to find the navigable path in an otherwise non-navigable scene. Following Drewes et al.'s (2011) finding with animals, it may be that we can efficiently localize a scene property within a single scene even if we cannot efficiently find that property over a set of several images.

Acknowledgments

Thanks to Kimberly Lamarre and Ashley Sherman for assistance in running the experiments, and to Karla Evans for constructive comments on the manuscript.

Commercial relationships: none.
Corresponding author: Michelle R. Greene.
Email: m.greene@search.bwh.harvard.edu.
Address: 64 Sidney St. Ste 170, Cambridge, MA 02139, USA.

References

Biederman, I., Blickle, T., Teitelbaum, R., & Klatsky, G. (1988). Object search in nonscene displays. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 456–467.

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

Castelhano, M. S., & Henderson, J. (2008). The influence of color on the perception of scene gist. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 660–675.

Delorme, A., Richard, G., & Fabre-Thorpe, M. (2000). Ultra-rapid categorisation of natural scenes does not rely on colour cues: A study in monkeys and humans. *Vision Research, 40,* 2187–2200.

Dick, M., Ullman, S., & Sagi, D. (1987). Parallel and serial processes in motion detection. *Science, 237,* 400–402.

Drewes, J., Trommershauser, J., & Gegenfurtner, K. R. (2011). Parallel visual search and rapid animal detection in natural scenes. *Journal of Vision, 11*(2):20, 1–21, http://www.journalofvision.org/content/11/2/20, doi:10.1167/11.2.20. [PubMed] [Article]

Fei-Fei, L., Van Rullen, R., Koch, C., & Perona, P. (2005). Why does natural scene categorization require little attention? Exploring attentional requirements for natural and synthetic stimuli. *Visual Cognition, 12,* 893–924.

Greene, M. R., & Oliva, A. (2009a). Recognition of natural scenes from global properties: Seeing the forest without representing the trees. *Cognitive Psychology, 58,* 137–176.

Greene, M. R., & Oliva, A. (2009b). The briefest of glances: The time course of natural scene understanding. *Psychological Science, 20,* 464–472.

Greene, M. R., & Oliva, A. (2010). High-level aftereffects to global scene properties. *Journal of Experimental Psychology: Human Perception and Performance, 36,* 1430–1432.

Grill-Spector, K., & Kanwisher, N. (2005). Visual recognition: As soon as you know it is there, you know what it is. *Psychological Science, 16,* 152–160.

Joubert, O. R., Rousselet, G. A., Fabre-Thorpe, M., & Fize, D. (2009). Rapid visual categorization of natural scene contexts with equalized amplitude spectrum and increasing phase noise. *Journal of Vision, 9*(1):2, 1–16, http://www.journalofvision.org/content/9/1/2, doi:10.1167/9.1.2. [PubMed] [Article]

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

Kaping, D., Tzvetanov, T., & Treue, S. (2007). Adaptation to statistical properties of visual scenes biases rapid categorization. *Visual Cognition, 15,* 12–19.

Kwak, H. W., Dagenbach, D., & Egeth, H. (1991). Further evidence for a time-independent shift of the focus of attention. *Perception & Psychophysics, 49,* 473–480.

Loftus, G. R., & Harley, E. M. (2005). Why is it easier to identify someone close than far away? *Psychonomic Bulletin & Review, 12,* 43–65.

Navon, D. (1977). Forest before the trees: The precedence of global features in visual perception. *Cognitive Psychology, 9,* 353–383.

Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research, 46,* 614–621.

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 global representation of the spatial envelope. *International Journal of Computer Vision, 42,* 145–175.

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

Piotrowski, L. N., & Campbell, F. W. (1982). A demonstration of the visual importance and flexibility of spatial-frequency amplitude and phase. *Perception, 11,* 337–346.

Rousselet, G. A., Fabre-Thorpe, M., & Thorpe, S. (2002). Parallel processing in high-level categorization of natural images. *Nature Neuroscience, 5,* 629–630.

Rousselet, G. A., Thorpe, S. J., & Fabre-Thorpe, M. (2004). Processing of one, two or four natural scenes in humans: The limits of parallelism. *Vision Research, 44,* 877–894.

Saarinen, J. (1996). Target localization and identification in rapid visual search. *Perception, 25,* 305–312.

Sharan, L., Rosenholtz, R., & Adelson, E. (submitted for publication). Material perception in real-world images is fast and accurate.

Torralba, A. (2009). How many pixels make an image? *Visual Neuroscience, 26,* 123–131.

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network, 14,* 391–412.

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review, 113,* 766–786.

Treisman, A., & Souther, J. (1985). Search asymmetry: A diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology: General, 114,* 285–310.

Van Rullen, R., Reddy, L., & Koch, C. (2004). Visual search and dual tasks reveal two distinct attentional resources. *Journal of Cognitive Neuroscience, 16,* 4–14.

Vickery, T. J., King, L., & Jiang, Y. (2005). Setting up the target template in visual search. *Journal of Vision, 5*(1):8, 81–92, http://www.journalofvision.org/content/5/1/8, doi:10.1167/5.1.8. [PubMed] [Article]

Wolfe, J. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review, 1,* 202–238.

Wolfe, J., Alvarez, G., Rosenholtz, R., Oliva, A., Torralba, A., Kuzmova, Y., et al. (2008). Search for arbitrary objects in natural scenes is remarkably efficient [Abstract]. *Journal of Vision, 8*(6):1103, 1103a, http://www.journalofvision.org/content/8/6/1103, doi:10.1167/8.6.1103.

Wolfe, J. M., Klempen, N., & Dahlen, K. (2000). Post-attentive vision. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 693–716.

Wolfe, J. M., & Myers, L. (2010). Fur in the midst of the waters: Visual search for material type is inefficient. *Journal of Vision, 10*(9):8, 1–9, http://www.journalofvision.org/content/10/9/8, doi:10.1167/10.9.8. [PubMed] [Article]

Wolfe, J. M., Vo, M. L.-H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and non-selective pathways. *Trends in Cognitive Science, 15,* 77–84.