A JOURNAL OF THE ASSOCIATION FOR PSYCHOLOGICAL SCIENCE aps

*Research Article*

# The Briefest of Glances

## The Time Course of Natural Scene Understanding

**Michelle R. Greene and Aude Oliva**

*Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology*

**ABSTRACT**—*What information is available from a brief glance at a novel scene? Although previous efforts to answer this question have focused on scene categorization or object detection, real-world scenes contain a wealth of information whose perceptual availability has yet to be explored. We compared image exposure thresholds in several tasks involving basic-level categorization or global-property classification. All thresholds were remarkably short: Observers achieved 75%-correct performance with presentations ranging from 19 to 67 ms, reaching maximum performance at about 100 ms. Global-property categorization was performed with significantly less presentation time than basic-level categorization, which suggests that there exists a time during early visual processing when a scene may be classified as, for example, a large space or navigable, but not yet as a mountain or lake. Comparing the relative availability of visual information reveals bottlenecks in the accumulation of meaning. Understanding these bottlenecks provides critical insight into the computations underlying rapid visual understanding.*

Catching meaning at a glance is a survival instinct, and a uniquely human talent that movie producers manipulate to their advantage when making trailers: By mixing snapshots of meaningful scenes in a rapid sequence, they can convey in a few seconds an evocative story from unrelated pictures of people, events, and places. In the laboratory, now-classic studies have shown that novel pictures can be identified in a 10-Hz sequence, although they are quickly forgotten when new images come into view (Intraub, 1981; Potter, 1975; Potter & Levy, 1969). Although several studies have investigated the availability of visual features over the course of a glance, in the study reported here we investigated the early perceptual availability of a number of semantic features used in scene classification. What types of meaningful information can human observers perceive from the briefest glances at images of novel scenes?

A typical scene fixation of 275 to 300 ms (Henderson, 2003; Rayner, 1998) is often sufficient to understand the gist of an image, namely, its semantic topic (e.g., "birthday party"; Intraub, 1981; Potter, 1975; Tatler, Gilchrist, & Risted, 2003). It takes slightly more exposure to recognize small objects in the scene (Fei-Fei, Iyer, Koch, & Perona, 2007) or to report their locations and spatial relations (Evans & Treisman, 2005; Tatler et al., 2003).

There is also evidence that observers can accomplish sophisticated scene analysis after viewing a novel scene for a single monitor refresh (10–40 ms) without masking. With such a brief exposure, observers can detect how pleasant a scene is (Kaplan, 1992) or whether it is natural or urban (Joubert, Rousselet, Fize, & Fabre-Thorpe, 2007); they can also determine the basic- or superordinate-level categories of a scene (Oliva & Schyns, 2000; Rousselet, Joubert, & Fabre-Thorpe, 2005) or detect the presence of a large object (Thorpe, Fize, & Marlot, 1996; Van Rullen & Thorpe, 2001). Although the extraordinarily high level of performance in these studies may be partially mediated by persistence in iconic memory, high performance is seen on similar tasks using masking paradigms (Bacon-Mace, Mace, Fabre-Thorpe, & Thorpe, 2005; Fei-Fei et al., 2007; Greene & Oliva, 2009; Grill-Spector & Kanwisher, 2005; Maljkovic & Martini, 2005).

Although many studies of natural scene understanding have focused on basic-level categorization or object identification, real-world scenes contain a wealth of structural and functional information whose time course of perceptual availability has not yet been determined. For example, determining how navigable a place is or whether it affords concealment is a perceptual decision with high survival value (Kaplan, 1992). Similarly, how the surfaces in a scene extend in space and change over time may influence how observers would behave in the scene. Properties of spatial layout, such as an environment's mean depth and openness, also influence its affordances (Oliva &

Address correspondence to Michelle R. Greene or Aude Oliva, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 77 Massachusetts Ave., 46-4078, Cambridge, MA 02139, e-mail: mrgreene@mit.edu or oliva@mit.edu.

Psychological Science RESEARCH, THEORY, & APPLICATION IN PSYCHOLOGY AND RELATED SCIENCES

Torralba, 2001). One can run in an open field, but not in a small and enclosed cave. Some materials of natural environments have a high transience, so that the scene changes very rapidly from one glance to the next (e.g., a rushing waterfall or a windy, sand-covered beach), whereas other surfaces (e.g., cliff rocks) have low transience, changing mostly in geological time. Similarly, material properties of surfaces, along with the interplay of atmospheric elements (e.g., water, wind, heat), give a place a particular physical temperature, another global property of the natural environment that strongly influences observers' behavior. All of these properties (and certainly more) combine in forming the understanding of a scene, much as recognition of a person depends on his or her gender, race, and facial expression, or recognition of an object depends on its shape, material, or position.

In the present study, we established perceptual benchmarks of early scene understanding by estimating the image exposure thresholds needed to perform two types of tasks: basic-level scene categorization (identifying an image as an ocean, a mountain, etc.) and global-property categorization (classifying spatial and functional properties of a scene, such as whether it is a hot place or a large environment). Different theories suggest different predictions for the results. Prototype theorists might predict that basic-level categories should be available first, as the basic level is privileged in object-naming experiments (e.g., Rosch, 1978). However, formal and experimental work has shown that global-property information is highly useful for basic-level scene categorization (Greene & Oliva, 2009; Oliva & Torralba, 2001), which suggests that global properties might have an early advantage. However, recent work examining the perceptual availability of object information at different levels of categorization showed that although subordinate-level categorizations required more image exposure than basic-level categorizations, knowing what an object was at the basic level did not require more image exposure than knowing that an object (vs. noise) was present (Grill-Spector & Kanwisher, 2005). This latter finding indicates that there may be no substantial threshold differences between the two types of tasks we examined.

In psychophysics, staircase methods have been used to determine human perceptual abilities efficiently (Klein, 2001). We measured presentation-duration thresholds to determine perceptual benchmarks for both global-property classification and basic-level categorization.

## METHOD

### Participants

Twenty participants (8 males, 12 females; ages 18–35 years) completed the psychophysical threshold experiment. They all had normal or corrected-to-normal vision and provided written informed consent. They received $10 for the 1-hr study.

### Stimuli

The stimuli used in this experiment were 548 full-color photographs of natural landscapes (see Fig. 1) selected from a large scene database (Greene & Oliva, 2009; Oliva & Torralba, 2001). The images measured 256 × 256 pixels.

To compare performance in the tasks, it was necessary to have normative rankings of the images' prototypicality as regards basic-level categories and global properties. Prototypicality regarding basic-level categories was assessed in a previous study (Greene & Oliva, 2009): Using a scale from 1 (*atypical*) to 5 (*highly prototypical*), 10 naive observers ranked 500 scenes in terms of how typical they were for each of several different basic-level category labels. For each basic-level category, we selected at least 25 images with a mean rank of 4 or higher. To obtain additional exemplars for each category, we visually matched the ranked prototypes with images from a database of approximately 10,000 natural landscapes. For basic-level categorization, we used prototypical scenes from seven natural landscape categories (*desert*, *field*, *forest*, *lake*, *mountain*, *ocean*, and *river*). In each block of trials, 50 images were from a single target category (e.g., forest), and 50 images were randomly selected from all the other categories (selection was constrained so that roughly equal numbers of images were taken from these categories).

For the global-property tasks, we used images that had been ranked as poles for one of seven global properties (*concealment*, *mean depth*, *naturalness*, *navigability*, *openness*, *transience*, and *temperature*; see Greene & Oliva, 2009, and Table 1 for descriptions). The same 500 natural scenes were ranked along each of the global properties (except *naturalness*) by observers who performed a hierarchical grouping task (at least 10 observers per property), organizing groups of 100 images at a time from lowest to greatest degree of a property (e.g., in the case of mean depth, ranking the images from the most close-up to the farthest view). Images with ranks within the first ($< 25\%$) or last ($> 75\%$) quartiles for a given property were considered typical poles for that property and were used in the current experiment. For the *naturalness* task, the target images were sampled from this pool of natural images, and urban images were added to serve as distractors. For global-property classification, each block contained 50 images from the high pole of a property (targets) and 50 images from the low pole of the same property (distractors).

As far as possible, test images for basic-level categorization and global-property classification were drawn from the same population of pictures. About half of all the images served as both targets and distractors (in different blocks). This helped to ensure that image-level differences were balanced across the experiment.

To produce reliable perceptual benchmarks, it is necessary to effectively limit additional sensory processing following image presentation. To this end, we used a dynamic masking paradigm (Bacon-Mace et al., 2005) consisting of rapid serial visual presentation of a sequence of mask images. The use of multiple

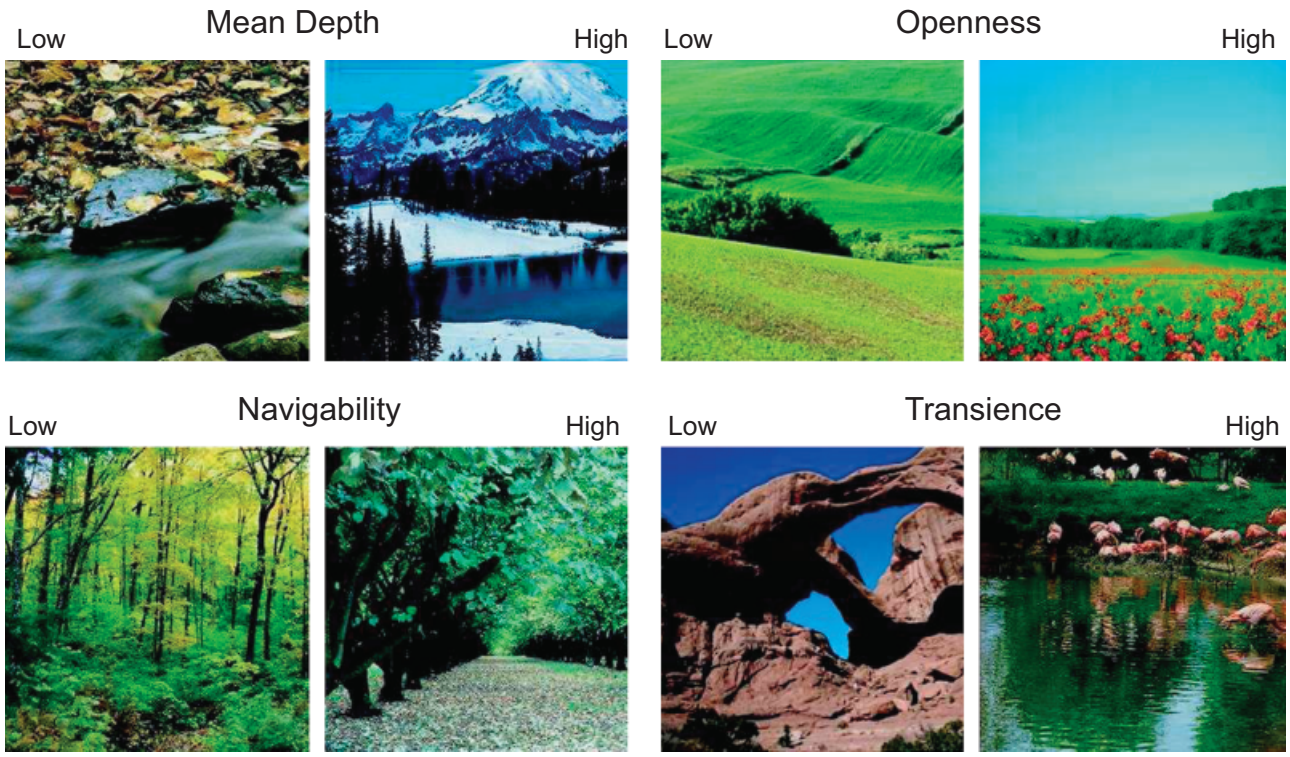


**Fig. 1.** Example images from the low and high poles of four global properties examined in this experiment. These images also illustrate the basic-level categories used in the experiment (river, mountain, field, and field, from left to right in the top row; forest, forest, desert, and lake, from left to right in the bottom row).

mask images minimizes interactions between the visual features of target images and masks, ensuring a more complete masking of image features.

Mask images (see Fig. 2) were synthesized images created from the test images, using a texture-synthesis algorithm designed by Portilla and Simoncelli (2000). We used the Matlab code provided on their Web site (http://www.cns.nyu.edu/~eero/texture/), enhancing it to include the color distribution of the model input image. The texture-synthesis algorithm uses a natural image as input, then extracts a collection of statistics from multiscale, multi-orientation filter outputs applied to the image, and finally coerces noise to have the same statistics. This method creates a nonmeaningful image that conserves marginal and first-order statistics, as well as higher-order statistics (cross-scale phase statistics, magnitude correlation and autocorrelation), while discarding object and spatial layout information. Power spectrum slopes for natural images and their masks were not significantly different ($p_{rep} = .76$).

### Design and Procedure

Participants sat in a dark room about 40 cm away from a 21-in. CRT monitor (100-Hz refresh rate). Stimuli on the screen subtended 7° × 7° of visual angle. Each participant completed

**TABLE 1**

*Descriptions of the Global Properties, as Presented to Participants in the Experiment*

| Global property | Target description | Nontarget description |
|---|---|---|
| Concealment | The scene contains many accessible hiding spots, and there may be hidden objects in the scene. | If standing in the scene, one would be easily seen. |
| Mean depth | The scene takes up kilometers of space. | The scene takes up less than a few meters of space. |
| Naturalness | The scene is a natural environment. | The scene is a man-made, urban environment. |
| Navigability | The scene contains a very obvious path that is free of obstacles. | The scene contains many obstacles or difficult terrain. |
| Openness | The scene has a clear horizon line with few obstacles. | The scene is closed, with no discernible horizon line. |
| Temperature | The scene environment depicted is a hot place. | The scene environment depicted is a cold place. |
| Transience | One would see motion in a video made from this scene. | The scene is not changing, except for patterns of daylight. |

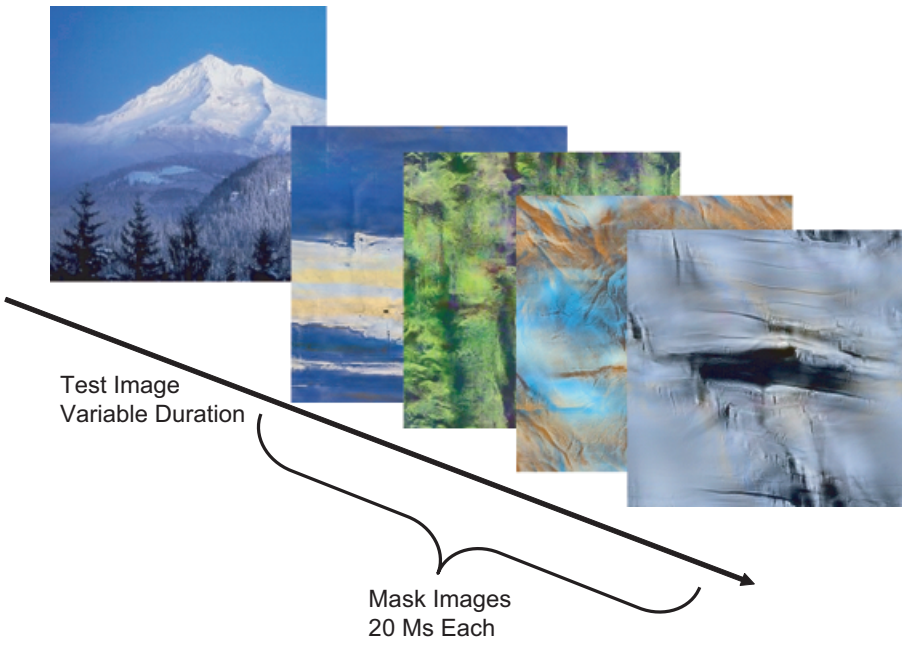


**Fig. 2.** Illustration of an experimental trial. The presentation duration of each test image was determined using a linear 3-up/1-down staircase, with the first trial of each block presented for 50 ms. Test images were dynamically masked using four colored textures.

seven 100-image blocks of basic-level categorization (one block for each target category) and seven 100-image blocks of global-property classification (one block for each property). The order of blocks was randomized and counterbalanced across participants. For each block, participants performed a yes/no forced-choice task; they were instructed to indicate as quickly and accurately as possible whether each briefly presented image was of the target category or global-property pole.

During each block, a linear 3-up/1-down staircase was employed. The first image in each block was shown for 50 ms; subsequent presentation times were determined by the accuracy of the observer's previous response, increasing by 10 ms (to a ceiling of 200 ms) if the response was incorrect and decreasing by 30 ms (to a floor of 10 ms) if the response was correct. With this procedure, performance converges at 75% correct (Kaernbach, 1990).

At the beginning of each experimental block, an instruction page appeared on the screen, describing the task (i.e., the basic-level category or property pole to be detected; see Table 1) and giving a pictorial example of a target and a nontarget. Figure 2 is a pictorial representation of a trial. Each trial commenced with a fixation point for 250 ms, followed by a test image. As noted, the presentation time of the test image varied (10–200 ms, as determined by the staircase method). The test image was immediately followed by a sequence of four randomly drawn mask images, presented for 20 ms each, for a total of 80 ms. After the mask sequence, participants were to indicate as quickly and accurately as possible whether or not the test image was the target. Visual feedback was provided for incorrectly classified images (the word "Error" was displayed for 300 ms following the response). Participants were first given a practice block of 20 trials to get used to the staircase procedure. The task for the practice block was to categorize scenes as indoor or outdoor, a classification not used in the main experiment. This experiment was run using Matlab and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997).

## RESULTS

For all blocks, the image presentation threshold was the presentation duration required for a participant to achieve 75% accuracy on the task. For some participants, not all blocks yielded a stable threshold. Because of the adaptive nature of the staircasing algorithm, very poor performance at the beginning of a block could lead to a considerable number of trials at the 200-ms image duration (the ceiling duration), resulting in artificially high final threshold calculations. For all analyses reported, we excluded data from blocks in which more than 10% of trials were spent at the maximum duration of 200 ms. Altogether these trials constituted only 5% of the data, and were evenly distributed between global-property and basic-level classification, $t(13) < 1$. We discuss here two processing benchmarks: (a) the upper bound of the exposure duration necessary for a given categorization block (the maximum image duration for each participant during that block) and (b) the duration at which participants achieved 75%-correct performance (used to compare the time needed for equivalent performance across blocks).

To ensure that the tasks were equally difficult, we compared the maximum image exposures needed. As image presentation times were controlled adaptively in the staircase procedure, the longest presentation time for a participant corresponds to the duration at which that participant made no classification errors (recall that errors resulted in increased subsequent presentation times). If the global-property and category tasks were of comparable difficulty, we would expect them to have similar maximum durations. Indeed, the mean maximum duration was 102 ms for the global-property tasks and 97 ms for the category tasks, $t(19) < 1$ (see Table 2). This result indicates that the two types of tasks were of similar difficulty.

A classic method for estimating thresholds from up-down staircases such as ours is to take the mode stimulus value for a participant (Cornsweet, 1962; Levitt, 1971). The logic in this experiment is simple: Because presentation duration was reduced by 30 ms following a correct response and increased by 10 ms following an incorrect response, participants converged on 75%-correct performance over the course of a block (Kaernbach, 1990), viewing more trials around the perceptual threshold than above or below it.

As estimation with the mode is a rather coarse method, we also estimated thresholds by fitting a Weibull function to the accuracy data for each participant for each block (using the maximum likelihood procedure) and solving for the threshold. This function typically provides very good fits to psychometric data (Klein, 2001). Figure 3 shows the Weibull fits and histograms of presentation times for 1 participant for a global-property block and a category block.

**TABLE 2**

***Presentation-Time Thresholds and Maximum Image Exposures***

| Block | Threshold for 75%-correct performance (ms) | Maximum image exposure (ms) |
|---|---|---|
| Global-property classification | | |
| Concealment | 35 (2.7) | 97 (7.9) |
| Mean depth | 26 (2.8) | 75 (4.9) |
| Naturalness | 19 (1.9) | 63 (4.9) |
| Navigability | 36 (4.5) | 120 (9.2) |
| Openness | 47 (4.6) | 119 (9.5) |
| Temperature | 29 (2.4) | 119 (9.5) |
| Transience | 45 (4.0) | 123 (8.8) |
| Mean | 34 (10) | 102 (24) |
| Basic-level categorization | | |
| Desert | 47 (4.7) | 93 (7.2) |
| Field | 55 (4.6) | 95 (7.3) |
| Forest | 30 (3.4) | 78 (6.6) |
| Lake | 51 (3.7) | 100 (7.1) |
| Mountain | 46 (3.3) | 95 (6.2) |
| Ocean | 55 (3.9) | 105 (6.5) |
| River | 67 (5.1) | 113 (6.1) |
| Mean | 50 (11) | 97 (11) |

**Note.** For the specific blocks, standard errors of the means are given in parentheses; for the means, standard deviations are given in parentheses. The reported thresholds are the means of modal presentation durations and thresholds estimated from Weibull fits.

The thresholds reported in Table 2 are the averages of the estimates obtained using the two methods. The presentation-time thresholds for all 14 blocks were remarkably short: All were well under 100 ms; values ranged from 19 ms (naturalness) to 67 ms (river).

We compared the threshold values for the global-property blocks with the threshold values for the categorization blocks and found that the mean threshold (based on the average of the Weibull fit and mode value) was significantly lower for global-property classification (34 ms) than for basic-level categorization (50 ms), $t(19) = -7.94$, $p_{rep} = .99$; the difference was also significant when we used the Weibull fits only, $t(19) = 7.38$, $p_{rep} > .99$, and when we used the modes only, $t(19) = 3.51$, $p_{rep} = .98$. Note that to compare performance for different features, it is necessary to ensure that there were equivalently difficult distractor images. On the one hand, a distractor that differed from the target by only 1 pixel would produce extremely large presentation-time thresholds (if observers could perform the task at all). On the other hand, distinguishing targets from white-noise distractors should result in ceiling performance. In our tasks, distractors were always prototypically different from targets. That is, in the global-property blocks, the distractors represented the opposite pole of the queried property, and both targets and distractors came from several basic-level categories. In the categorization blocks, the distractors were prototypes of a variety of nontarget categories and were chosen so as to show the greatest variety of category prototypes. In this way, targets and distractors were chosen, to the extent possible, to vary only in the attribute being tested. Recall that ceiling performance was reached at similar presentation durations in the global-property and category tasks, which indicates that although performance had an early advantage in global-property blocks, this advantage was not due to the blocks of basic-level categorization being harder than the global-property blocks.

Figure 4a shows the distributions of participants' presentation-duration thresholds for both types of tasks, using a Gaussian fit. The distributions of participants' thresholds in categorization blocks were rather homogeneous in terms of both means and variances (see Fig. 4b). In contrast, the distributions of thresholds in global-property blocks (Fig. 4c) were more heterogeneous, with some coming very early and others more closely resembling the categorization thresholds. We calculated 95% confidence intervals around the means and found that the presentation-duration threshold was significantly shorter for *forest* than for other basic-level categories, and was significantly longer for *openness* and *transience* than for other global properties.

## DISCUSSION

A large amount of meaningful information can be gleaned from a single glance at a scene (Bacon-Mace et al., 2005; Biederman,

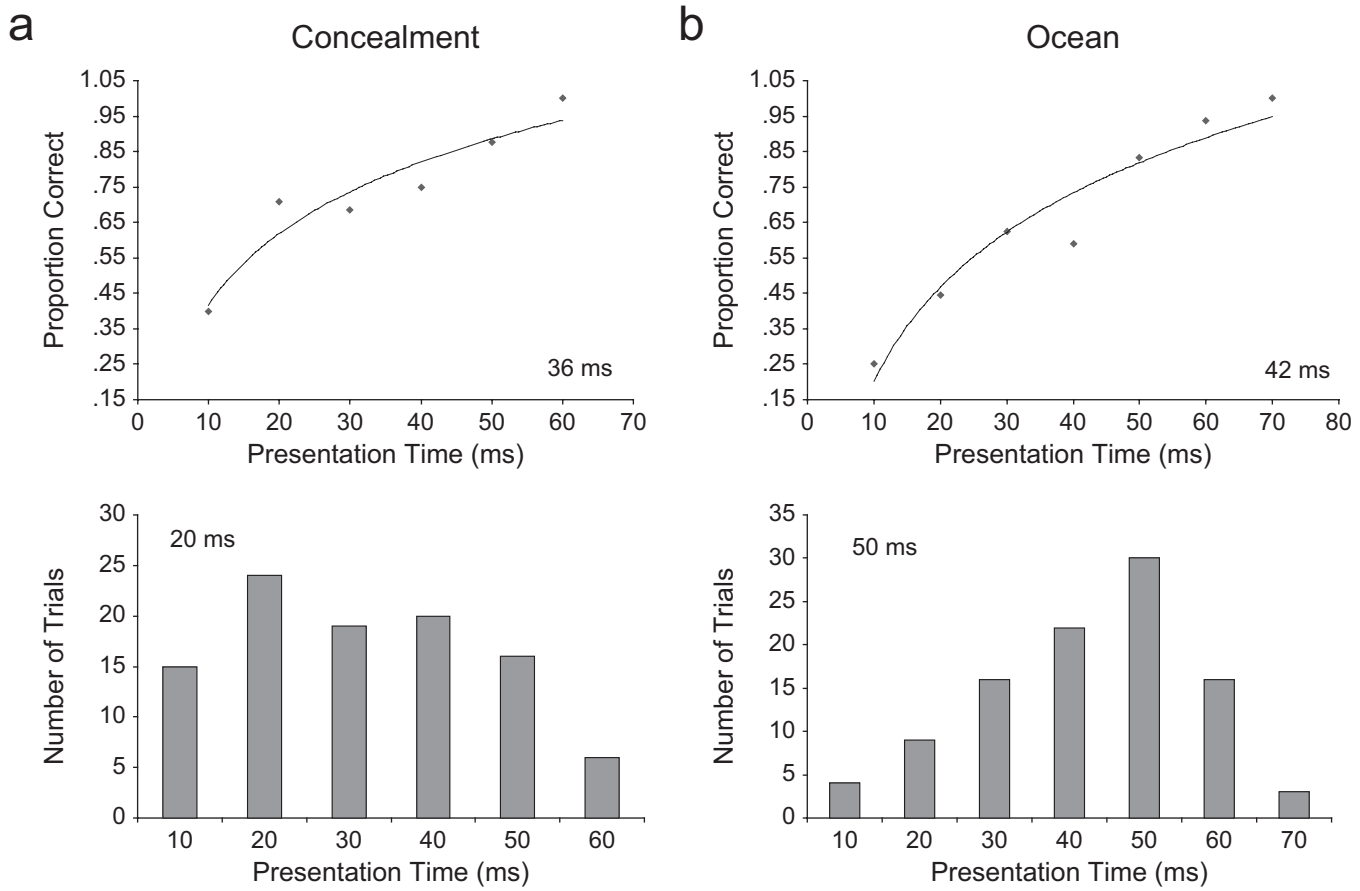


**Fig. 3.** Example of threshold computation for (a) a block of global-property classification (concealment) and (b) a block of basic-level categorization (ocean). Each graph shows data from the same participant. The top row shows accuracy as a function of presentation duration. The curves are Weibull fits, and thresholds for 75%-correct performance are indicated. The bottom row shows histograms of presentation times; the mode in each distribution was another measure of the threshold. The thresholds reported in Table 2 are the averages of the estimates obtained using the two methods.

Rabinowitz, Glass, & Stacy, 1974; Castelhano & Henderson, 2008; Fei-Fei et al., 2007; Grill-Spector & Kanwisher, 2005; Joubert et al., 2007; Maljkovic & Martini, 2005; Oliva & Schyns, 2000; Potter & Levy, 1969; Schyns & Oliva, 1994; Thorpe et al., 1996; Renninger & Malik, 2004; and many other studies), but our study is the first to establish perceptual benchmarks comparing the types of meaningful information that can be perceived during very early perceptual processing.

What meaningful perceptual and conceptual information can be understood from extraordinarily brief glances at a novel scene? We have provided insight into this question by comparing the shortest image exposures required for participants to achieve equivalent performance (75% correct) on a number of tasks involving classification of naturalistic scenes. We found that this threshold ranged from 19 ms to 67 ms, and that performance reached asymptote at about 100 ms of image exposure (range: 63–123 ms; see Table 2). Remarkably, the threshold presentation duration was, on average, shorter for perception of a scene's global properties than for perception of the scene's basic-level category. These results are related to other work in ultrarapid scene perception (Joubert et al., 2007; Rousselet et al., 2005), which has demonstrated that participants can classify a scene as natural versus man-made more quickly than they can make a semantic classification of the scene (e.g., mountain, urban). Indeed, we also found that the image exposure thresholds were shortest for classifying images according to whether or not they were natural (19 ms).

Our results are complementary to those of studies examining the accrual of image information over time (Fei-Fei et al., 2007; Intraub, 1981; Rayner, Smith, Malcolm, & Henderson, 2009; Tatler et al., 2003). For instance, Rayner et al. (2009) found that although participants rapidly understood the overall semantic topic of a scene, they required at least a 150-ms fixation to find an object within that scene (e.g., a broom in a warehouse image). Likewise, in the study by Fei-Fei et al. (2007), observers viewed briefly masked pictures depicting various events and scenery (e.g., a soccer game, a busy hair salon, a choir, a dog playing fetch) and then described in detail what they saw in the pictures. Global scene information, such as whether an outdoor or indoor scene was depicted, was perceived well above chance (50%) with less than 100 ms of exposure. Although free-report responses may be confounded with inference (observers may

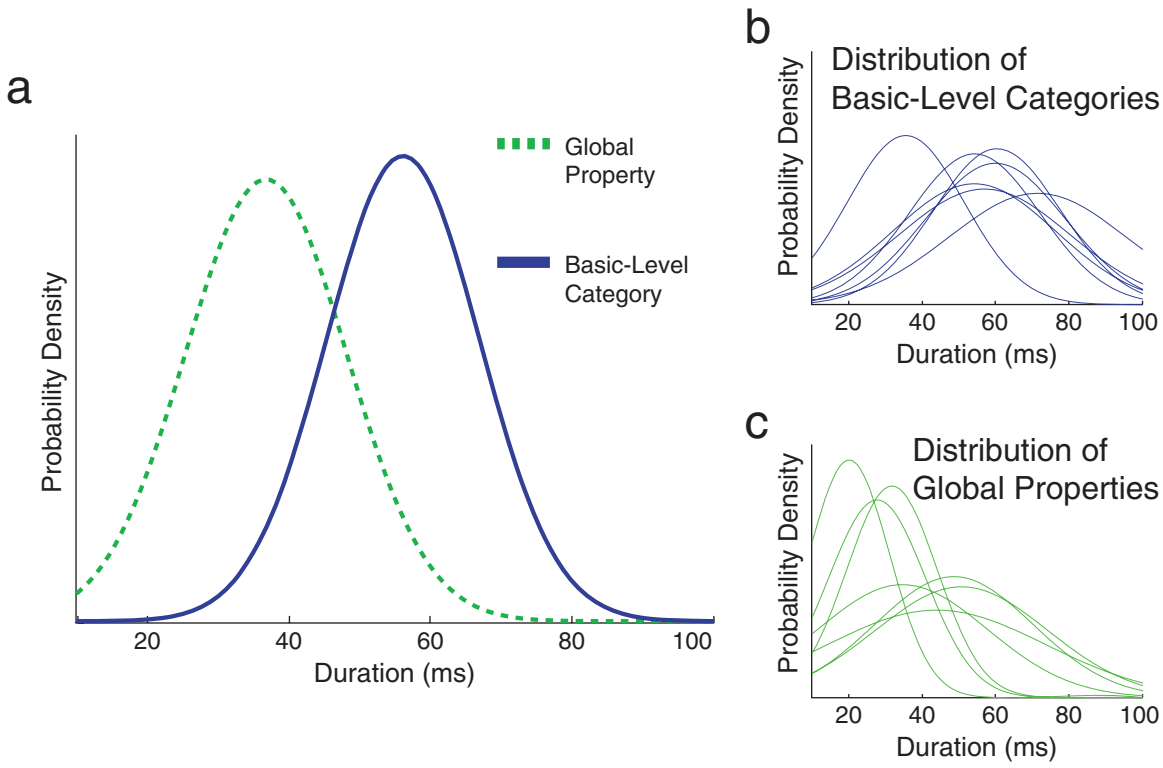


**Fig. 4.** Distributions of observers' presentation-duration thresholds for (a) global-property classification and basic-level categorization, (b) each block of basic-level categorization, and (c) each block of global-property classification. All shown curves are best-fitting Gaussians.

overestimate the information seen in a brief glance of an image on the basis of knowledge of the covariance among features and objects in the real world; see Brewer & Treyans, 1981), and may be biased toward reporting verbally describable information, this study is consistent with others (e.g., Biederman et al., 1974; Intraub, 1981; Potter, 1975; Schyns & Oliva, 1994; Tatler et al., 2003) finding that as exposure duration increases, observers are better able to fully perceive the details of an image.

Our findings agree with a global-to-local view of scene perception (Navon, 1977; Oliva & Torralba, 2001; see also, e.g., Joubert et al., 2007; Schyns & Oliva, 1994). We have shown that at the very early stages of visual analysis, certain global visual information can be more easily gleaned from an image than even its basic-level category. This result suggests the intriguing possibility that there exists a time during early visual processing when an observer will know, for example, that a scene is a natural landscape or a large space, but does not yet know that it is a mountain or a lake scene. Our result may be predicted by computational work showing that basic-level scene categories cluster along global-property dimensions describing the spatial layout of scenes (the *spatial-envelope* theory; Oliva & Torralba, 2001). Furthermore, for human observers rapidly categorizing scenes at the basic level, more false alarms are produced by distractors that share global properties with the target category than by distractors that do not share such properties with the target category (e.g., more false alarms to images showing a closed, rather than open, space when the target category is *forest*; Greene & Oliva, 2009). The current results lend credence to the possibility that rapid categorization of a scene may be achieved through the perception of a few robust global properties of the scene.

In the current study, the range of presentation-time thresholds was large (19–67 ms), but remained well below 100 ms. There was also a large range of thresholds within both types of classifications (19–47 ms for global-property classification and 30–67 ms for basic-level categorization). This suggests that there is substantial diversity in the diagnostic image information used by observers to perform these tasks, and that these pieces of information may be processed with different time courses.

It remains for future work to determine which image features observers use to reach these remarkable levels of performance. Recent studies in visual cognition suggest the intriguing possibility that the brain may be able to rapidly calculate robust statistical summaries of features and objects—such as the mean size of a set of shapes (Ariely, 2001; Chong & Treisman, 2005), the average orientation of a pattern (Parkes, Lund, Angelucci, Solomon, & Morgan, 2001), the center of mass of a set of objects (Alvarez & Oliva, 2008), or even the average emotion of a set of faces (Haberman & Whitney, 2007)—in an automatic fashion (Chong & Treisman, 2005) and outside the focus of attention (Alvarez & Oliva, 2008). This suggests that some tasks might be performed with shorter presentation times than others because

their diagnostic features are coded somewhat more efficiently. For instance, naturalness classification had the fastest threshold in our study and the fastest reaction time in Joubert et al. (2007), and has been shown to be correlated with low-level features, distributed homogeneously over the image (Torralba & Oliva, 2003). Likewise, Renninger and Malik (2004) demonstrated that texture statistics provided good predictions of human scene categorization at very short presentation times. By abstracting statistical homogeneities related to structural and functional properties of a scene, the human brain may be able to comprehend complex visual information in a very short time.

Although people feel as if they instantaneously perceive a full, rich, and meaningful world, this full understanding accumulates over time. By understanding the time course of visual processing, researchers can uncover bottlenecks in the accumulation of this information. Uncovering the benchmarks of visual processing at the feature level will be a significant step forward in understanding the algorithms of human visual processing.

***Acknowledgments***—The authors wish to thank Timothy Brady, Krista Ehinger, Molly Potter, Keith Rayner, and one anonymous reviewer for helpful comments and discussion. This research was supported by a National Science Foundation graduate research fellowship to M.R.G. and by a National Science Foundation CAREER award (0546262) and grant (0705677) to A.O.

## REFERENCES

Alvarez, G., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science*, *19*, 392–398.

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, *12*, 157–162.

Bacon-Mace, N., Mace, M.J.M., Fabre-Thorpe, M., & Thorpe, S.J. (2005). The time course of visual processing: Backward masking and natural scene categorization. *Vision Research*, *45*, 1459–1469.

Biederman, I., Rabinowitz, J.C., Glass, A.L., & Stacy, E.W. (1974). On the information extracted from a glance at a scene. *Journal of Experimental Psychology*, *103*, 597–600.

Brainard, D.H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 443–446.

Brewer, W., & Treyans, J. (1981). Role of schemata in memory for places. *Cognitive Psychology*, *13*, 207–230.

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

Chong, S., & Treisman, A. (2005). Statistical processing: Computing the average size in perceptual groups. *Vision Research*, *45*, 891–900.

Cornsweet, T. (1962). The staircase method in psychophysics. *American Journal of Psychology*, *75*, 485–491.

Evans, 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.

Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? *Journal of Vision*, *7*(1), Article 10. Retrieved January 16, 2009, from http://www.journalofvision.org/7/1/10/

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

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.

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology*, *17*, 751–753.

Henderson, J.M. (2003). Human gaze control in real-world scene perception. *Trends in Cognitive Sciences*, *7*, 498–504.

Intraub, H. (1981). Rapid conceptual identification of sequentially presented pictures. *Journal of Experimental Psychology: Human Perception and Performance*, *7*, 604–610.

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

Kaernbach, C. (1990). A single-interval adjustment-matrix procedure for unbiased adaptive testing. *Journal of the Acoustical Society of America*, *88*, 2645–2655.

Kaplan, S. (1992). Environmental preference in a knowledge-seeking, knowledge-using organism. In J.H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adaptive mind* (pp. 535–552). New York: Oxford University Press.

Klein, S. (2001). Measuring, estimating and understanding the psychometric function: A commentary. *Perception & Psychophysics*, *63*, 1421–1455.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, *49*, 467–477.

Maljkovic, V., & Martini, P. (2005). Short-term memory for scenes with affective content. *Journal of Vision*, *5*(3), Article 6. Retrieved January 16, 2009, from http://www.journalofvision.org/5/3/6/

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

Oliva, A., & Schyns, P. (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.

Parkes, L., Lund, J., Angelucci, A., Solomon, J., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, *4*, 739–744.

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

Portilla, J., & Simoncelli, E. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision*, *40*, 49–71.

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

Potter, M.C., & Levy, E.I. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology*, *81*, 10–15.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372–422.

Rayner, K., Smith, T.J., Malcolm, G.L., & Henderson, J.M. (2009). Eye movements and visual encoding during scene perception. *Psychological Science*, *20*, 6–10.

Renninger, L.W., & Malik, J. (2004). When is scene identification just texture recognition? *Vision Research*, *44*, 2301–2311.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Erlbaum.

Rousselet, G.A., Joubert, O.R., & Fabre-Thorpe, M. (2005). How long to get to the "gist" of real-world natural scenes? *Visual Cognition*, *12*, 852–877.

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.

Tatler, B., Gilchrist, I., & Risted, J. (2003). The time course of abstract visual representation. *Perception*, *32*, 579–593.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*, 520–522.

Torralba, A., & Oliva, O. (2003). Statistics of natural images categories. *Network: Computation in Neural Systems*, *14*, 391–412.

Van Rullen, R., & Thorpe, S. (2001). The time course of visual processing: From early perception to decision making. *Journal of Cognitive Neuroscience*, *13*, 454–461.

(Received 6/27/08; Revision accepted 9/13/08)