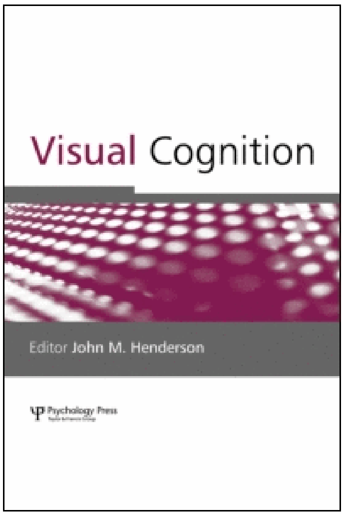

Visual Cognition
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713683696

The natural/man-made distinction is made before basic-level distinctions in scene gist processing

Lester C. Loschky [a];Adam M. Larson [a]
[a] Kansas State University, Manhattan, KS, USA

First published on: 23 July 2009

To cite this Article Loschky, Lester C. andLarson, Adam M.(2010) 'The natural/man-made distinction is made before basic-level distinctions in scene gist processing', Visual Cognition, 18: 4, 513 — 536, First published on: 23 July 2009 (iFirst)
To link to this Article: DOI: 10.1080/13506280902937606
URL: http://dx.doi.org/10.1080/13506280902937606

VISUAL COGNITION, 2010, 18 (4), 513–536

Psychology Press
Taylor & Francis Group

The natural/man-made distinction is made before basic-level distinctions in scene gist processing

Lester C. Loschky and Adam M. Larson
Kansas State University, Manhattan, KS, USA

What level of categorization occurs first in scene gist processing, basic level or the superordinate "natural" versus "man-made" distinction? The Spatial Envelope model of scene classification and human gist recognition (Oliva & Torralba, 2001) assumes that the superordinate distinction is made prior to basic-level distinctions. This assumption contradicts the claim that categorization occurs at the basic level before the superordinate level (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). The present study tests this assumption of the Spatial Envelope model by having viewers categorize briefly flashed and masked scenes after varying amounts of processing time. The results show that early levels of processing (SOA < 72 ms) (1) produced greater sensitivity to the superordinate distinction than basic-level distinctions, and (2) basic-level distinctions crossing the superordinate natural/man-made boundary are treated as a superordinate distinction. Both results support the assumption of the Spatial Envelope model, and challenge the idea of basic-level primacy.

Keywords: Basic level; Categorization; Man-made; Masking; Natural; Natural scenes; Recognition; Scene gist; Scene perception; Superordinate; Time course.

Recognizing the global meaning of a scene, its "gist", is possibly the earliest meaningful stage of scene perception (Oliva, 2005), which can occur with high accuracy after a masked scene presentation of as little as 40–60 ms (Bacon-Mace, Mace, Fabre-Thorpe, & Thorpe, 2005; Fei-Fei, Iyer, Koch, & Perona, 2007; Loschky et al., 2007). Furthermore, gist recognition affects many important later cognitive processes, such as attention within scenes (Eckstein, Drescher, & Shimozaki, 2006; Gordon, 2004; Loftus & Mackworth, 1978; Torralba, Oliva, Castelhano, & Henderson, 2006), possibly object recognition within scenes (Boyce & Pollatsek, 1992; Davenport & Potter, 2004; De Graef, De Troy, & D'Ydewalle, 1992; Hollingworth & Henderson, 1998; Palmer, 1975), and long-term memory for the contents of scenes (Brewer & Treyens,

Please address all correspondence to Lester C. Loschky, Visual Cognition Lab, Dept. of Psychology, 471 Bluemont Hall, 1100 Mid Campus Drive, Kansas State University, Manhattan, KS 66506, USA. E-mail: loschky@ksu.edu

http://www.psypress.com/viscog DOI: 10.1080/13506280902937606

1981; Pezdek, Whetstone, Reynolds, Askari, & Dougherty, 1989). Many studies have operationally defined scene gist recognition in terms of categorizing a scene with a single word or noun phrase, such as "natural", "man-made", "forest", "mountain", or "street" (e.g., Kaping, Tzvetanov, & Treue, 2007; Loschky et al., 2007; McCotter, Gosselin, Sowden, & Schyns, 2005; Oliva & Schyns, 2000; Rousselet, Joubert, & Fabre-Thorpe, 2005).

This conceptualization of gist has allowed for an easy interchange between human scene perception research and artificial vision research on scene image classification. In turn, this interchange has spurred the development of the currently dominant theory of scene gist recognition, the "Spatial Envelope" model (Oliva & Torralba, 2001, 2006; Oliva, Torralba, Castelhano, & Henderson, 2003; Torralba, 2003; Torralba et al., 2006). The Spatial Envelope model accounts for scene categorization at "both a superordinate level and a basic level" (Oliva & Torralba, 2001, p. 150). In their model, the superordinate level includes the categories "natural" and "man-made" and the level below that, the basic level, includes categories such as "forest", "mountain", or "street". The Spatial Envelope model categorizes scenes at each level based on a set of filter responses derived from a principal component analysis of the Fourier spectra of images in each category. These filters are labelled as the "naturalness", "openness", "roughness", "expansion", and "ruggedness" of a scene (Oliva & Torralba, 2001). More concretely, a scene most likely to produce a high value on the "naturalness" filter (and thus classified as "natural") would have low frequencies on the horizontal axis (e.g., the horizon) and roughly equal power at all spatial frequencies for diagonal orientations (e.g., foliage). Conversely, a scene most likely to score low on the "naturalness" filter (and thus classified as "man-made") would have more energy at middle and higher frequencies on the vertical and horizontal axes (e.g., carpentered structures such as buildings). Thus, the natural/man-made distinction would be made on the basis of very elementary features from the Fourier amplitude domain.

A critical assumption of the model is that classification of an image at the superordinate level as either "natural" or "man-made" occurs *prior to* classification of the image at the basic level (Oliva & Torralba, 2001, pp. 169–170). This assumption came from the fact that the natural/man-made distinction was the most commonly cited basis on which participants first segregated a group of 81 images into two groups (Oliva & Torralba, 2001). The natural/man-made distinction is also an important factor mediating the effects of colour diagnosticity in scene gist recognition (Oliva & Schyns, 2000).

This assumption of the Spatial Envelope model that the natural/man-made distinction is made prior to basic-level distinctions has not, to our knowledge, been discussed much in the scene gist recognition literature. This is surprising because it directly contradicts a fundamental principle of the theory of basic-level categories, namely that pictured items "are first seen or recognized as members of their basic category (with additional processing required to identify them as members of their superordinate ... category)" (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976, p. 412). Rosch et al. (1976) supported this claim with evidence from experiments measuring priming of picture detection, and precued picture verification, and numerous explanations have been put forward for these widely replicated findings (for review, see Gosselin & Schyns, 2001). Rosch et al. argued that, in perception, items are recognized at the basic level first because they share many features with same-category items, but few features with different-category items, thus making it possible to have an isomorphic representation (e.g., a prototypical visual image) of an entire basic-level category of physical stimuli. In contrast, superordinate items are so diverse that they share few features, and thus cannot be visualized, but instead must be inferred based on knowledge of class membership. Jolicoeur, Gluck, and Kosslyn (1984) showed that the entry point at which objects are recognized may not always be the basic level, but sometimes the subordinate level. However, they showed that the superordinate level was recognized later, which they argued was accomplished through its semantic associations with the basic-level category.

Interestingly, however, recent research has begun to challenge the claim of basic-level primacy, and instead has argued that the superordinate level is processed prior to the basic level, following a coarse-to-fine processing order (Large, Kiss, & McMullen, 2004; Rogers & Patterson, 2007). Large et al. (2004), used an event-related potential (ERP) paradigm to track the time course of categorization of pictured objects, and found that ERPs for superordinate categories showed an earlier latency than for basic-level categories both early (100–200 ms) and somewhat later (320–420 ms) in processing. Rogers and Patterson (2007) showed that (1) in semantic dementia, categories are lost at the basic level before the superordinate level, and (2) accuracy for superordinate-level tasks is greater than for basic-level tasks when viewers are forced to respond faster than normal (i.e., when processing time is limited). Thus, these results suggest that, consistent with the assumption of the Spatial Envelope Model, a coarse superordinate-level distinction could potentially be made prior to finer basic-level distinctions early in processing.

Nevertheless, it might be argued that since these findings are all from studies investigating categorization of objects, not scenes, we have no way of judging whether the principle of basic-level primacy does or does not hold for scene gist categorization. However, using Rosch et al.'s (1976) methods and logic, Tversky and Hemenway (1983) found clear evidence of a basic-level of scene categorization. Thus, a natural assumption would be that basic-level primacy should extend to recognizing scenes as well. Even so, from the viewpoint of the Spatial Envelope model, Tversky and Hemenway's study left important open questions, because (1) at the superordinate level, they did not investigate the natural/man-made distinction, and (2) they did not investigate the time course of different levels of categorization. These two unresolved issues therefore leave open the question of whether or not the assumption of natural/man-made primacy in scene gist recognition is valid.

The first unresolved issue was addressed by a recent study by Loschky and Larson (2008), which compared the superordinate-level natural/man-made distinction with basic-level distinctions, using Oliva and Torralba's (2001) taxonomy, and a random sample of their image set. The study showed that viewers were slightly, but statistically, more accurate at making the superordinate natural/man-made distinction than making basic-level categorizations. However, Loschky and Larson did not investigate the time course of scene categorization across levels of the taxonomy, and thus cannot speak to the issue of which level of categorization occurs first.

Both unresolved issues have been addressed, but not resolved, by recent studies. Joubert, Rousselet, Fize, and Fabre-Thorpe (2007) compared the results of their study, in which they used a go/no-go natural/man-made scene categorization task, with previous results from their laboratory (Rousselet et al., 2005), which had used a basic-level scene categorization task. They found a 47 ms difference in mean reaction times between the two studies, with the natural/man-made task being performed faster than the basic-level task, indirectly supporting the assumption of the Spatial Envelope model that the natural/man-made distinction occurs earlier in processing. Fei-Fei et al. (2007) investigated the time course of scene categorization using visual masking, and compared accuracy between the superordinate natural/man-made distinction and basic-level distinctions. In contrast to Joubert et al., the authors found that "in general, superordinate-level scene categories (e.g., … man made indoor, natural outdoor) seem to require the same amount of information in recognition as the basic-level scenes (e.g., field, beach, skyline, urban centers)" (p. 20). Thus, while Fei-Fei et al.'s results did not support the assumption that basic-level categorization occurs first, neither did they support the competing assumption of the Spatial Envelope model that the natural/man-made distinction occurs prior to those at the basic level. This lack of differentiation might have been due to the nature of the dependent variables measured, or the range of processing times investigated. Because the Fei-Fei et al. study was designed as an open-ended exploration of scene gist processing, participants were asked to write down whatever they saw on each trial, and their protocols were later coded into a response taxonomy. Because such measures are affected by memory limitations and other limitations of self-report data, their study might not have been sensitive enough to pick up processing differences between the natural/man-made distinction and those at the basic level. In addition, the shortest stimulus–onset asynchrony (SOA) included in their study was 27 ms, at which point many of the responses were not significantly above chance. Nevertheless, recent studies have shown above-chance gist recognition at SOAs as low as 10–12 ms (Bacon-Mace et al., 2005; Loschky et al., 2007). Thus, if processing differences between the superordinate natural/man-made distinction and basic-level distinctions occur at even earlier levels of processing, they would have been missed.

The current study was therefore designed to resolve both of these issues, in order to test the important assumption of the Spatial Envelope model, that the superordinate natural/man-made distinction is made prior to distinctions at the basic level. In order to keep the conditions as close as possible to those assumed by the Spatial Envelope model, we used the Oliva and Torralba (2001) taxonomy and a sample of their image set. In order to investigate the time course of processing, we used visual masking, including very short SOAs (12 ms) to tap very early processes. This is based on the assumption that variation of SOAs can be used to vary processing time, which is supported by over 100 years of visual masking research (for a review, see Breitmeyer & Ogmen, 2006), including recent studies combining behavioural and neurophysiological methods, such as event-related potentials and magnetoencephalography (Bacon-Mace et al., 2005; Rieger, Braun, Bulthoff, & Gegenfurtner, 2005). Consistent with many previous scene gist recognition studies (e.g., Kaping et al., 2007; Loschky et al., 2007; McCotter et al., 2005; Oliva & Schyns, 2000; Rousselet et al., 2005), the dependent measure will be category verification accuracy.

The two alternative assumptions about the earliest level of scene categorization lead to two alternative hypotheses. One hypothesis is that basic-level distinctions are recognized first, as predicted by Rosch et al. (1976) and others (for review, see Gosselin & Schyns, 2001). This hypothesis is slightly complicated by the fact that our previous findings (Loschky & Larson, 2008) suggest that when scenes are not masked, thus allowing unlimited processing time (within a single fixation), the superordinate natural/man-made distinction may be slightly more accurate than basic-level distinctions. However, if, as Rosch et al. argue, superordinate-level categorizations are only inferred from class membership at the basic level, the small advantage we previously found for the natural/man-made distinction without visual masking should have been due to relatively late processing. Thus, with short processing times we would expect a clear advantage for basic-level categorization, but with relatively long processing times we could expect a crossover to a small advantage for the superordinate natural/man-made distinction.

The competing hypothesis is that the superordinate natural/man-made distinction occurs before basic-level distinctions, as assumed by the Spatial

Envelope model (Oliva & Torralba, 2001). Thus, with short processing times we would expect a clear advantage for the natural/man-made distinction over basic-level distinctions. However, as processing time increases, we would expect that this advantage should gradually decrease to the relatively small advantage found in our previous study (Loschky & Larson, 2008). This would suggest that, at intermediate processing times, the information needed to make basic-level distinctions begins to catch up with that needed for the superordinate natural/man-made distinction.

EXPERIMENT 1

Method

Participants. The experiment included 80 undergraduates from Kansas State University (61 females, mean age = 19.68). All participants had corrected-to-normal vision of 20/30 or better, and received research credit for their participation. After collecting data from these participants, data from those whose total mean accuracy was two standard deviations below the mean for their respective task condition (basic, or superordinate natural/man-made) were discarded. This resulted in discarding data from two participants (one in each task condition). Two new participants' data was collected to replace them.

Stimuli. The experiment used 384 images from the Oliva and Torralba (2001) database (http://cvcl.mit.edu/database.htm). 192 of these images represented the superordinate "natural" category, and 192 images represented the "man-made" category. The "natural" and "man-made" categories were made up of four basic-level categories (48 images each): coast, forest, mountain, and open country for the "natural" condition, and city centre, street, tall building, and highway for the "man-made" condition. All images were reduced to monochrome to avoid producing spurious colours when phase-randomizing the masking stimuli. When the 256×256 pixel images were viewed at a distance of 53.34 cm using a chinrest, they subtended $10.12^{\circ} \times 10.12^{\circ}$ of visual angle. Mean luminance and RMS contrast were equalized for the entire set (as in Loschky & Larson, 2008; see Loschky et al., 2007, Appendix 2 for details). The 85 Hz Samsung SyncMaster 957 MBS (17-inch) monitors were calibrated for luminance and contrast.

Two visual masks were used, based on a pair of completely phase-randomized scenes (see Loschky et al., 2007, Appendix 2 for details). Loschky and Larson (2008) found that phase-randomizing scenes produces a "natural bias" such that viewers are more likely to perceive scenes as natural.

To avoid biasing the current results in cases in which integrative masking perceptually mixed information from the target and mask, we used masking images that our previous results had shown to have no such bias (i.e., the response rates for "natural" and "man-made" were both exactly 50% for both images). Additionally, both images had been equally poorly recognized (50% accuracy, i.e., chance) in both categorization tasks (basic-level and superordinate-evel natural/man-made).

Procedures. The study used a 2 × 6 mixed factorial design. Participants were randomly assigned to the between-participants factor of categorization task, i.e., categorizing scene images using either basic-level categories or the superordinate natural/man-made distinction. The within-participants factor was SOA, which had values of 12, 24, 36, 48, or 72 ms, and a no-mask condition representing unlimited visual processing time within a single fixation. A 12 ms SOA was assumed to be short enough to capture very early gist processing, and the 72 ms SOA and no-mask conditions were predicted to be sufficiently long to capture late gist recognition processes (Bacon-Mace et al., 2005; Fei-Fei et al., 2007; Loschky et al., 2007; Rieger et al., 2005).

Participants were familiarized with the categories used in the study by having them view 80 labelled scene images (10 scene images from each of eight basic-level categories, which were preceded by the appropriate labels at the basic level or at the superordinate-level "natural" or "man-made") for 1 s each. Then, participants were given practice with the experimental task, including masking at the same SOAs as the experiment, by having them do 32 practice trials without feedback (using four images from each basic-level category, postcued either at the basic level or as the superordinate-level "natural" or "man-made"). The practice trials included twice as many trials in the 72 ms SOA and no-mask conditions (eight trials each) as in the 12, 24, and 36 ms SOA conditions (four trials each). This was done to make the task somewhat easier for participants while still giving them experience with the full range of SOAs in the actual experiment, with the order of SOAs randomized across practice trials. All of the images used for familiarization and practice trials came from a separate sample of images from the Oliva and Torralba (2001) image set, and were not used in the actual experiment.

A schematic of a trial is shown in Figure 1. Participants were instructed to fixate the centre of the screen and press a button to initiate the trial. Following a 750 ms delay, the target image was presented for 12 ms. In masking trials, there was an interstimulus interval (ISI) of 0–60 ms between target and mask (i.e., an SOA of 12–72 ms) during which the screen was blank, followed by a phase-randomized mask presented for 96 ms. On no-mask trials, the ISI and mask were absent. Then, on all trials, there was a 750 ms delay (during which the screen was blank), followed by the cue word

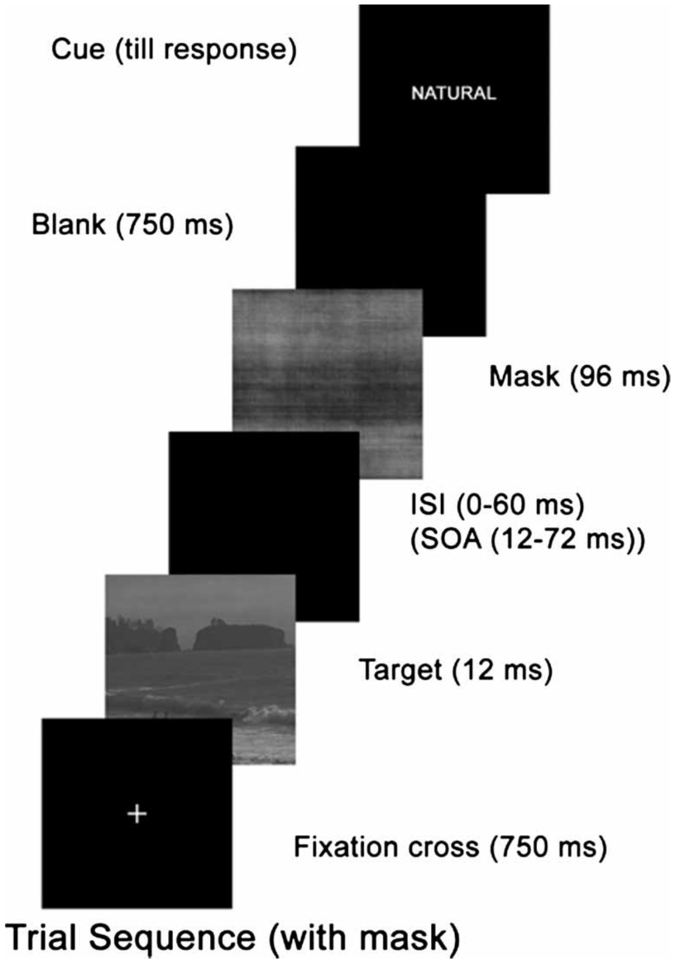

Figure 1. Trial schematic (with mask). The no-mask condition was the same except that it did not have an ISI or mask.

until the participant responded. Depending on the task condition, the cue was either (1) a basic-level category label, e.g., "mountain" or "street", or (2) the superordinate-category labels "natural" or "man-made". If the cue matched the scene image presented, i.e., it was valid, participants were instructed to press the "Yes" button; otherwise they were to press the "No" button. For each participant, cues were randomly paired with images, under the constraints that all cues were used equally often, all cue categories were 50% valid, and all image categories were cued with 50% validity. The 384 scene images were each presented once, and in a different random order for every participant.

Results and discussion

We analysed the data using nonparametric signal detection measures of sensitivity, A' (Grier, 1971), and bias, B''_D (Donaldson, 1992). Figure 2 clearly shows that viewers were more sensitive to the natural/man-made distinction at early levels of processing than they were to basic-level distinctions. Besides showing the normal monotonic relationship between sensitivity (A′) and processing time, F(Huynh-Feldt)(2.885, 225.037) = 219.83, $p < .001$, Figure 2 shows greater sensitivity in the natural/man-made task ($M = 0.88$, $SE = 0.009$) than the basic-level task ($M = 0.83$, $SE = 0.009$), $F(1, 78) = 15.78$, $p < .001$, though this advantage diminished as processing time increased, F(Huynh-Feldt) (2.885, 225.037) = 6.657, $p < .001$. Planned comparisons between the natural/man-made and basic-level tasks at SOAs less than 72 ms produced significant differences (SOA lower than 72 ms), ts > 2.863, ps < .005, whereas the difference at 72 ms was not significant, $t(78) = 1.41$ $p = .16$. The difference for the unmasked condition though not much larger, did, however, reach significance, $t(78) = 2.78$, $p = .007$. Overall, these results strongly support the hypothesis that, at early levels of processing, viewers are

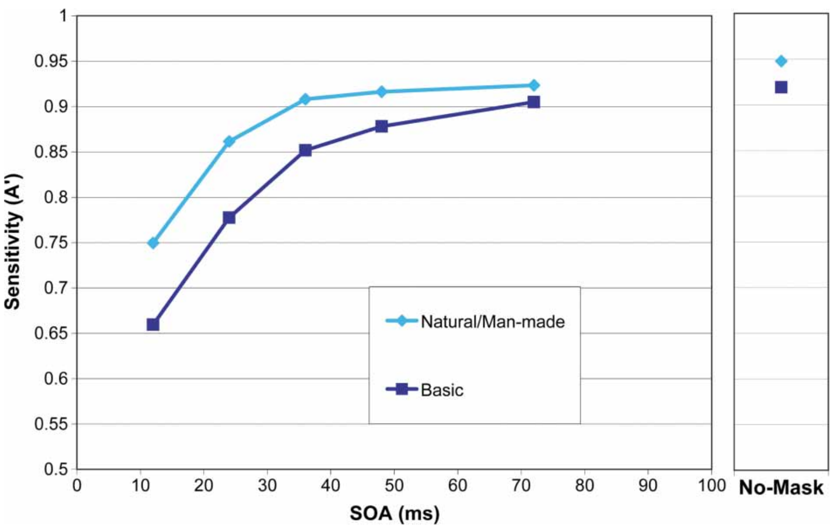

Figure 2. Comparison of sensitivity between basic and natural/man-made categories at each SOA. To view this figure in colour, please see the online issue of the Journal.

more sensitive to cues to the natural/man-made distinction than to basic-level scene distinctions.

We also analysed our data in terms of differences in response bias. Figure 3 shows that viewers were more biased in making basic-level judgements than natural/man-made judgements when they had little processing time. Viewers making judgements about basic-level distinctions were much more likely to say "No" ($M = 0.21$, $SE = 0.040$) than when making judgements about the natural/man-made distinction ($M = -0.02$, $SE = 0.040$), $F(1, 78) = 16.58$, $p < .001$, though this difference greatly diminished as processing time increased, F(Huynh-Feldt)(4.501, 351.053) = 9.148, $p < .001$. Planned comparisons showed that the differences were significant for all masking SOAs, ts > 2.097, ps < .04, while the difference for the unmasked condition was not, $t(78) = -1.71$, $p = .091$. Importantly, in the natural/man-made task, the bias never differed significantly from 0 (neutral), though it approached significance in the unmasked condition, ts < 1.881, ps > .067. Interestingly, a comparison of Figures 2 and 3 shows that, in the basic-level condition, there was a significant negative correlation between sensitivity and bias, $r = -.289$, $p < .001$, discussed later. However, in the natural/man-made condition, the correlation was much smaller and nonsignificant, $r = .083$, $p = .201$.

Taken together, these response bias results clearly show that, at early levels of processing, when sensitivity was very low, viewers set a conservative criterion for making basic-level distinctions. However, as processing time and sensitivity increased, viewers' criteria for making those distinctions became

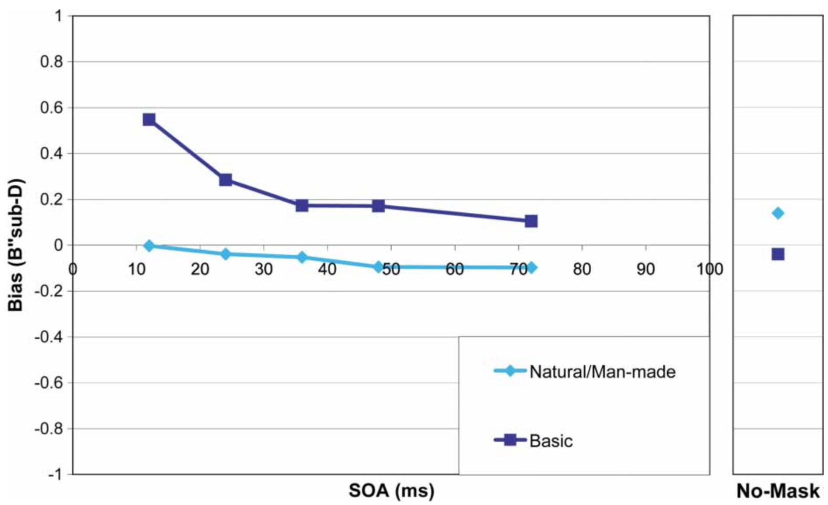

Figure 3. Comparison of response bias between basic and natural/man-made at each SOA. To view this figure in colour, please see the online issue of the Journal.

less biased. Very similar results for object detection and categorization have recently been reported by Kim and Chong (2008). Though a correlation between sensitivity and bias is contrary to a basic tenet of detection theory (Macmillan & Creelman, 2005), negative correlations of an even greater magnitude between sensitivity and bias measures (d' and c) have previously been reported (See, Warm, Dember, & Howe, 1997). Furthermore, the negative correlation accounted for only 8% of the variance in the bias measure. Such a change of bias with level of sensitivity was absent when viewers made natural/man-made distinctions. One explanation for these results is that viewers felt they needed more information in order to make the finer distinctions of the basic level than for the natural/man-made distinction. More concretely, given the same very limited amount of information, viewers felt less comfortable saying "Yes" to a cue such as "mountain" than to a cue such as "natural".

A more sophisticated explanation for the "No" bias, which was found only in the basic-level task, is in terms of the relative variability of the signal and noise distributions. In the basic-level task, the signal distribution would be made up of one basic-level category, the target, whereas the noise distribution would be made up of seven basic-level categories, the distractors. On the other hand, in the natural/man-made task, the signal distribution would be made up of four basic-level categories, as would the noise distribution. Thus, although the variability of signal and noise distributions would be expected to be equal in the natural/man-made task, it would be expected to be unequal in the basic-level task—more specifically with greater variability in the noise distribution than the signal distribution. In fact, according to signal detection theory, it is precisely this imbalance in distributional variability (noise > signal) that is expected to produce a "No" bias (Swets, 1986). If so, the "No" bias in the basic-level task should disappear if the variability of the signal and noise distributions is equalized.

The only way to equalize signal and noise distribution variability across both the natural/man-made task and the basic-level task would be to have an equal number of category options in the signal and noise distributions of each task. Since the natural/man-made task has only two categories, the same would need to be true of the basic level. Thus, in both tasks there would be only two categories.

These considerations also suggest a possible explanation for the sensitivity differences found in the current experiment. Specifically, although the correlation between sensitivity and bias was small, it might explain the differences in sensitivity between the two categorization tasks. If so, then reducing the number of basic-level categories to two, which should eliminate the differences in bias, should also eliminate the differences in sensitivity between the tasks.

EXPERIMENT 2

The scene categorization results of Experiment 1 are potentially explainable in terms of the different numbers of categories used in the two tasks (eight basic level vs. two natural/man-made) in the experiment. Thus, the current experiment tests these explanations by presenting only two scene categories at the basic and superordinate levels. If the greater sensitivity in the superordinate-level natural/man-made task than the basic-level task at short SOAs in Experiment 1 was due to the unequal number of categories, then that difference should disappear when the number of categories is equal.

A further consideration when one is comparing categorization across taxonomic levels, as in the current study, is how one selects distractors for trials in which the correct response to the category label is "No". Two options have been utilized in the literature. The first option uses distractors from every category besides the target scene image (e.g., the distractor for the target "open country" could be "street"). This strategy would use each category of distractor, regardless of its superordinate "parent" category. The second strategy uses distractors that are only members of the same superordinate "parent" category of the target scene image (e.g., the distractor for the target "open country" could be "coast" but could not be "street"). These strategies will be described as the "different parent" and "same parent" strategies, respectively. Large et al. (2004) compared these strategies with distractors that were from the same or different superordinate categories. When distractors were chosen from any nontarget category, if the relevant features include both those at the target taxonomic level and those at the next higher level, then no differences in reaction times were observed between the basic- and superordinate-level tasks. However, when distractors are chosen from the same superordinate category, in which the relevant features are limited to only those from the target level of the taxonomy, performance for the superordinate-level task was faster than for the basic-level task, though the stimuli were the same in both. The current experiment tests several hypotheses concerning sensitivity and bias between taxonomic levels utilizing these two strategies for the basic-level categorization task.

First, consider the case of a two-category basic-level task, in which the target and distractor basic-level categories come from *different parent* superordinate-level categories (e.g., "coast" vs. "city centre"), which we will refer to as the basic (between natural/man-made) task. If the coarse superordinate-level natural/man-made scene distinction is processed before finer basic-level distinctions, then features used to make the superordinate-level natural/man-made distinction are the *only* features necessary to meet the demands of this binary category discrimination task—other features necessary to make finer basic-level distinctions are completely redundant.

Thus, in such a basic-level task, the viewer would only need to map those features used to make the superordinate-level natural/man-made distinction to the basic-level labels. Therefore, if the coarse superordinate-level natural/man-made scene distinction is processed before finer basic-level distinctions, and such coarse superordinate-level distinctions are *the only ones* needed for either task, then one would predict *no difference* in sensitivity between those two tasks at short SOAs.

Next, consider the case of a two-category basic-level task in which both target and distractor basic-level categories come from within the *same parent* superordinate category (e.g., natural), which we will refer to as the basic (within natural/man-made) task. In this case, features used to make the coarse superordinate-level natural/man-made distinction are no longer relevant to the task. Instead, to carry out the task, the viewer must rely on features necessary to make finer basic-level discriminations, because basic-level categories of the same superordinate "parent" category share course features, therefore more specific detailed features are needed to distinguish basic-level categories from the same superordinate parent. Thus, if the coarse superordinate-level natural/man-made scene distinction is processed before finer basic-level distinctions, one would predict greater sensitivity in a basic-level task that requires only the coarse superordinate-level natural/man-made distinction (i.e., basic-level categories having *different* parents) than a basic-level task that requires finer basic-level distinctions (i.e., basic-level categories having the *same parent*) at short SOAs.

Importantly, the previous two hypotheses directly contradict the well-researched basic-level supremacy, which we will refer to as the *basic-level primacy hypothesis*. If basic-level distinctions are processed first, when the tasks use only two scene categories, then basic-level tasks should produce greater sensitivity than superordinate-level natural/man-made tasks at short SOAs. This prediction would hold regardless of whether targets and distractors come from the same or different superordinate categories. Specifically, the superordinate category of a scene should only be processed *after* the basic level has been identified, if it is processed at all, thus rendering information from the superordinate category moot, and producing lower sensitivity in a natural/man-made task than in basic-level tasks at short SOAs.

Regarding bias, our predictions were simpler. We hypothesized that one could equalize bias across the two tasks (basic vs. superordinate-level natural/man-made) by equalizing the number of alternatives in each, specifically two. Having two alternatives in the superordinate-level natural/man-made task produced unbiased responding in Experiment 1. Thus, it was hypothesized that both tasks would be equally unbiased in the current experiment.

Method

Participants. A total of 56 undergraduates (21 females) participated in the study for credit in an introductory psychology course. Participants had a mean age of 19.36 ($SD = 1.41$) and all participants had a minimum visual acuity of 20/30.

Stimuli. The experimental setup and materials for Experiment 2 were the same as in Experiment 1, except that the experiment consisted of 128 trials. Stimuli included 32 randomly selected scene images from each of four scene categories (coast, open country, city centre, and street) used in Experiment 1.

Procedures. Participants were randomly assigned to one of three conditions, each of which presented pairs of scene image categories in two blocks of 64 trials. In the "basic within natural/man-made" condition, the two blocks of trials consisted of basic-level scene categories from within the same superordinate category, specifically, "coast" and "open country" images in one block (both "natural" categories) and "city centre" and "street" images in the other (both "man-made" categories), with order of blocks counterbalanced across subjects. In the "basic between natural/man-made" condition, the two blocks of trials consisted of basic-level scene categories that crossed the superordinate boundary, with "coast" and "city centr"e images in one block, and "open country" and "street" images in the other. In both "basic" conditions, the cue at the end of each trial was one of the pair of basic-level category labels. The "natural/man-made" condition was exactly the same as the "basic between natural/man-made" condition, except that the cue at the end of each trial was the superordinate-level "natural" or "man-made" label.

The study used a 3×4 mixed factorial design. Participants were randomly assigned to be in one of three conditions discussed previously (Ns = 19, 19, and 18, respectively). The within-subject factor was the SOA of 12, 24, and 48 ms, and a no-mask condition. The events within each trial were the same as in Experiment 1.

Results and discussion

From the 7168 total trials across all participants, 52 trials were eliminated without replacement from the analysis due to a program error that presented the mask at incorrect SOAs.

The nonparametric measures for sensitivity, A', and response bias, B''_{D}, were used. Here we focus on the critical results for evaluating our competing

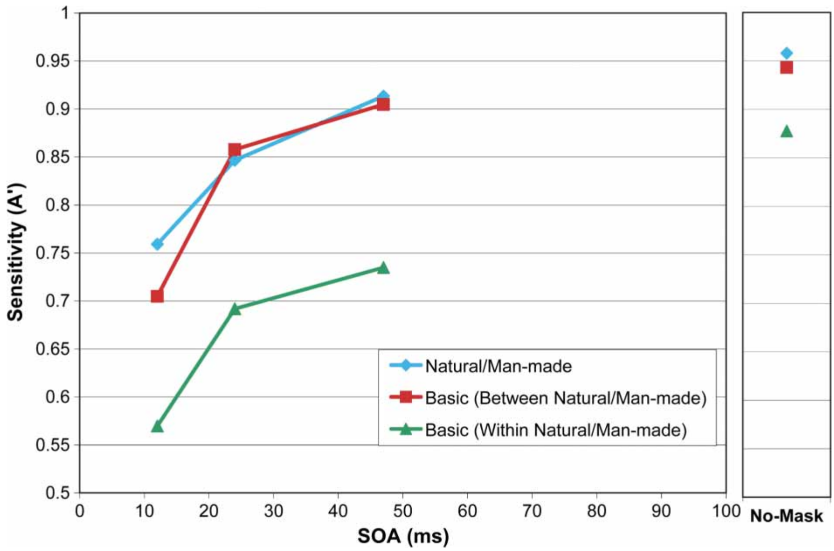

Figure 4. Comparison of sensitivity between the basic within natural/man-made, basic between natural/man-made, and natural/man-made categorization tasks at each SOA. To view this figure in colour, please see the online issue of the Journal.

hypotheses. Of key importance, Figure 4 shows that sensitivity was nearly identical in the "basic between natural/man-made" condition ($M = 0.85$, $SD = 0.14$) and the "natural/man-made" condition ($M = 0.87$, $SD = 0.10$), $t(35) = 0.806$, $p = .861$, *ns*, Sidak adjusted. A follow-up *t*-test also showed that the slight difference between these two conditions at the shortest SOA was not significant ($M = 0.70$, $SD = 0.19$ vs. $M = 0.76$, $SD = 0.09$), $t(35) = 1.11$, $p = .276$, *ns*. These results are consistent with the hypothesis that a basic-level task involving two categories from different parent superordinate categories (here, natural vs. man-made) only requires the viewer to make the natural/man-made distinction. In contrast, Figure 4 shows greater sensitivity in both of these conditions than in the "basic within natural/man-made" condition ($M = 0.72$, $SD = 0.16$), *t*s(36 and 35) > 5.372, *p*s $< .001$, Sidak adjusted). In addition, Figure 4 suggests that these differences in sensitivity were greater at earlier processing times (SOAs $<$ 50 ms) than later (no-mask), as evidenced by a marginally significant Condition $\times$ SOA interaction, F(Huynh-Feldt)(5.163, 136.831) $= 2.186$, $p = .057$. Together, these results are consistent with the hypothesis that viewers categorize scenes in terms of the coarse superordinate natural/man-made distinction before making finer grained basic-level distinctions.

Figure 5 shows that response bias did not differ between the three conditions, $F(2, 53) = 1.251$, $p = .295$, *ns*, or as a function of processing time,

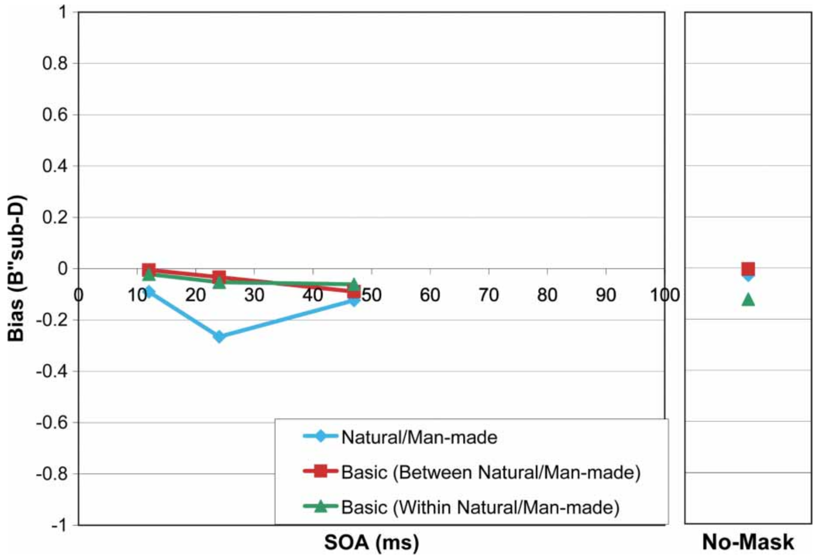

Figure 5. Comparison of response bias between the basic-within natural/man-made, basic-between natural/man-made, and natural/man-made categorization tasks at each SOA. To view this figure in colour, please see the online issue of the Journal.

F(Huynh-Feldt)(2.162, 114.573) = 0.707, $p = .506$, *ns*, nor as an interaction between the two factors, *F*(Huynh-Feldt)(4.324, 114.573) = 0.861, $p = .496$, *ns*. Indeed, bias did not differ significantly from zero at any SOA, indicating that participants were essentially unbiased in their responses. These results are consistent with the hypothesis that equalizing the number of category options to two would eliminate the "No" bias found at early processing times in the basic-level condition in Experiment 1.

It should be noted that if we combine the sensitivity data from both basic-level tasks (between natural/man-made and within natural/man-made), it is more representative of the type of basic-level tasks encountered in the real world. Doing so results in an average sensitivity level that is considerably lower than in the superordinate natural/man-made task. This provides converging evidence to that already shown by Experiment 1, but without the "No" bias found there, and strongly contradicts the basic-level primacy hypothesis.

Overall, the results of Experiment 2 were consistent with the assumption of the Spatial Envelope model (Oliva & Torralba, 2001) that the coarse superordinate-level natural/man-made distinction is made prior to finer basic-level scene category distinctions, and inconsistent with the opposing theory (Jolicoeur et al., 1984; Rosch et al., 1976) that basic-level distinctions are made first. We predicted that, in a binary basic-level categorization task in which the two categories (e.g., coast vs. city centre) come from *different parent* superordinate categories (natural vs. man-made), if the natural/man-made distinction is made first, there should be no difference in sensitivity between distinguishing natural versus man-made scenes and distinguishing the two basic-level scene categories, because the basic-level task could be achieved using *only* the features necessary for the superordinate-level natural/man-made distinction. Using the same logic, if the superordinate-level natural/man-made distinction is made first, we predicted that when participants instead made basic-level distinctions within the *same parent* superordinate category (either natural or man-made), which *cannot* depend on the natural/man-made distinction, their sensitivity would be lower, particularly at early levels of processing. The results of the current experiment supported both sets of predictions.

It is important to note that the use of our masking paradigm allows us to discuss the current results not only in terms of coarse versus fine feature discriminations, but also in terms of their time course of processing. In masking paradigms that vary the target-to-mask SOA, variations in accuracy with SOA are interpreted as following from variations in the amount of information integrated over time due to interruption of target processing by the mask (Kovacs, Vogels, & Orban, 1995; Loftus & Mclean, 1999). Thus, tasks which require less information integration (i.e., shorter SOAs) for a given level of accuracy, achieve it at earlier levels of processing than tasks which require more information integration (i.e., longer SOAs) for the same level of accuracy. This reasoning has been tested and refined by over 100 years of behavioural research on masking (for review, see Breitmeyer & Ogmen, 2006), with even stronger support being provided recently by masking studies using both single cell recording with macaques (Kovacs et al., 1995; Rolls, Tovée, & Panzeri, 1999) and brain imaging with humans (Bacon-Mace et al., 2005; Rieger et al., 2005).

The sensitivity results shown in Figure 4 can be explained in terms of a coarse-to-fine order of scene gist categorization, using the previously discussed time-course logic of masking effects together with our previous explanation in terms of coarse versus fine category discriminations. At the shortest SOA (12 ms), scene gist processing is stopped after only a small amount of visual information has been integrated. Because the features required to make the superordinate-level natural/man-made distinction are relatively simple, the natural/man-made task does not require much information to achieve a sensitivity level of .76, which is roughly half-way between chance (.5) and asymptotic performance (.95). Similarly, because the basic (between natural/man-made) task only requires perception of the features necessary for the natural/man-made distinction in order to distinguish the two scene categories, sensitivity in this task (.70) is essentially equivalent. By contrast, the basic (within natural/man-made) task cannot rely on features used for the natural/man-made distinction, but instead requires perception of more features needed to make true basic-level distinctions. Thus, at the shortest SOA, the amount of information that has been integrated results in a sensitivity level in this task that is not much above chance (.57). Indeed, in order to attain the same level of sensitivity as the tasks that depend only on the natural/man-made distinction, the basic (within natural/man-made) task requires an SOA of 24 ms—a doubling of processing time. In addition, these differences in sensitivity are greatly diminished when relatively unlimited processing time in a single fixation is allowed (the no-mask condition). Thus, the differences in processing the different category levels are primarily at the early stages of processing (i.e., SOAs $<$ 50 ms). The above results are therefore consistent with an explanation of scene gist categorization in terms of a coarse-to-fine processing order, which starts with the natural/man-made distinction and proceeds to basic-level distinctions.

The results of Experiment 2 also rule out the argument that the sensitivity advantage in Experiment 1 for the superordinate-level natural/man-made distinction was simply due to only needing to make a binary distinction. Instead, the current results show that binary basic-level scene distinctions vary in difficulty, depending on whether or not they cross the superordinate-level natural/man-made boundary, though with longer processing times these differences lessen. On the other hand, the results of the current experiment suggest that the "No" bias found in the basic-level task in Experiment 1 was likely due to having multiple scene categories, even though participants only had to respond to one category on any given trial. This conclusion is supported by the fact that the "No" bias was eliminated by reducing the number of basic-level categories being considered to only two.

Discussion and conclusion

The current study explicitly tested a critical assumption of the Spatial Envelope model that the superordinate-level natural/man-made distinction is made before basic-level distinctions, and provides clear supporting evidence for it. This finding is important for at least two reasons. First, the current results use human subject data to test a central assumption of an influential computational model of scene gist recognition, the Spatial Envelope model (Oliva & Torralba, 2001, 2006; Oliva et al., 2003; Torralba, 2003; Torralba et al., 2006), which has been essentially untested. Confirming this assumption strengthens support for the model, and suggests that it may be validly extended to other computational scene classification models as well. Second, the current results question the validity of the opposing default assumption, based on the seminal work of Rosch et al. (1976) combined with that of

Tversky and Hemenway (1983), that scene categorization starts at the basic level and works its way up to the superordinate level. Though further research may be needed to more rigorously test this assumption, the current results suggest that it may actually be an impediment to understanding the processes involved in scene gist recognition.

This discussion raises the important question of why our results contradict the default assumption that the basic level of scene categorization is recognized first. A possible explanation comes from the previously mentioned study by Rogers and Patterson (2007) that investigated object categorization at the basic and superordinate category levels. The authors similarly found that viewers begin by processing objects at the superordinate level before the basic level, but that processing of objects at the basic level gradually catches up with and overtakes that at the superordinate level as processing nears a high-confidence recognition threshold. In their study, when processing time was limited, by forcing participants to respond quickly, accuracy for the superordinate level was better than at the basic level. However, when participants were given their normal amount of time to respond, they were more accurate at the basic level than the superordinate. As noted earlier, using an ERP methodology, Large et al. (2004) have shown converging results.

Rogers and Patterson (2007) explained these results in terms of a parallel distributed processing (PDP) semantic space, in which the largest clusters of nodes are at the superordinate level and are widely separated from each other. Within those superordinate-level clusters, basic-level clusters are more tightly packed. When the model encounters a new stimulus, it begins at a neutral point in the semantic space, and begins moving in the direction of nodes that most closely match it. As the model state moves through the semantic space, it first enters clusters at the superordinate level, thus producing a superordinate advantage at the earliest stages of processing. However, once the model enters a basic-level cluster, generalization to similar nodes within that cluster is easier due to being more tightly packed. This produces a basic-level advantage at later stages of processing. Interestingly, this model's semantic space shares parallels with the multidimensional feature space of the Spatial Envelope model, in which each node is a particular image, and local clusters correspond to the basic level, with superordinate clusters (natural and man-made) lying in a higher dimension of the model, where they are assumedly more distant from one another.

Considering this discussion, one might argue that the current results are unsurprising, since pairs of natural versus man-made scenes look less similar to each other than do pairs of basic-level scenes, i.e., basic-level scenes have greater feature overlap (Gosselin & Schyns, 2001). However, such a critique ignores a fundamental argument for why the basic level should be processed earlier than the superordinate level. Specifically, superordinate categories tend to be uninformative of their expected features, i.e., they are so diverse that few features are reliably associated with them (Gosselin & Schyns, 2001; Rosch et al., 1976). For example, it is easier to imagine sets of features associated with a "forest" (leaves, branches, tree trunks) or "coast" (waves, horizon, clouds) than with a "natural" scene. Thus, although superordinate categories are more distinct than basic-level categories, because of less feature overlap, superordinate-level categories are less informative—i.e., knowing the category does little to help one predict its features. It has therefore been argued that because basic-level categories have a balance of both (moderate) distinctiveness and (high) informativeness, they should be processed earlier (Gosselin & Schyns, 2001; Murphy, 1991; Rosch et al., 1976).

In fact, only under specific conditions has categorization at the superordinate level shown greater accuracy or faster reaction times than that at the basic level. One such condition is the previously mentioned advantage for superordinate categories when viewers have limited processing time (Rogers & Patterson, 2007). Another condition under which superordinate categories have been shown to produce faster reaction times than basic-level categories is when they can be identified based on a small number of singly sufficient features (Gosselin & Schyns, 2001; Murphy, 1991). This suggests that a few singly sufficient features may be used to distinguish natural versus man-made scenes, especially when processing time is limited (i.e., early in processing). Indeed, the Spatial Envelope model argues that the superordinate-level natural/man-made distinction is made on the basis of primitive perceptual information, such as the global distribution of orientations within a scene. For example, oblique orientations may be indicative of the natural category, whereas vertical orientations may be indicative of the man-made category (Oliva & Torralba, 2001). Similarly, our previous research has suggested that the presence or absence of sharp edges in a scene image strongly influences whether it is judged to be "man-made" or "natural", respectively (Loschky & Larson, 2008). For objects, similarly low-level features appear to produce extremely rapid distinctions between the superordinate categories of "animal" versus "artefact" (Levin, Takarae, Miner, & Keil, 2001). This may explain why we have found a small advantage for the natural/man-made distinction over basic-level distinctions even with unlimited processing time (in a single fixation). Based on these points, one might argue that one cannot say in general which taxonomic level, superordinate or basic, is processed first, but rather only that categorical distinctions, whatever their taxonomic level, whose sufficient features are more efficiently acquired will be processed earlier, as suggested by Gosselin and Schyns (2001). Our results suggest that the ecologically important superordinate-level distinction between "natural" versus "man-made" scenes is one such case, and the results of Fei-Fei et al. (2007) suggest that the "indoor" versus "outdoor" distinction may be another. Further research will be needed to determine whether these cases are exceptions or the norm.

Another possible critique of the current study is that the results may reflect differences in difficulty, rather than priority of processing. Our reasoned use of visual masking lends credence to the claim that differences in sensitivity at different SOAs reflect different time courses of information integration and use (Bacon-Mace et al., 2005; Breitmeyer & Ogmen, 2006; Kovacs et al., 1995; Rieger et al., 2005; Rolls et al., 1999). Nevertheless, studies measuring the time course of brain processes at different levels of categorization (e.g., Large et al., 2004) may be required to fully address this issue.

Another possible explanation for the differences observed between the basic level and superordinate-level natural/man-made tasks could be that there were an unequal number of training trials for each specific scene category between the tasks. Specifically, although both the basic-level and superordinate-level natural/man-made tasks in Experiment 1 each had 80 familiarization trials and 32 practice trials, because the natural/man-made task had fewer categories, participants saw more examples of images identified at the superordinate level ("natural" or "man-made") than at the basic level. This inequality in the number of training trials per category may have contributed to the greater sensitivity in the superordinate-level natural/man-made task at the shortest SOAs. The fact that Experiment 2 reduced the number of basic-level categories by half, yet showed stronger evidence consistent with the superordinate-level natural/man-made distinction being made earlier than basic-level distinctions, weakens this argument. Thus, if differences in training affected the results, it seems likely that they were only a minor contributing factor.

It is important to note that although our results are inconsistent with aspects of the basic-level theory, this inconsistency may only hold for relatively early levels of conceptual processing, which heavily involve perception. The claim that the basic level is generally preferred in language production for describing objects and scenes (Rosch et al., 1976; Tversky & Hemenway, 1983) seems indisputable. In most communication contexts, it is clearly less informative to describe a scene as "natural" or "man-made" than to describe it with a basic-level term such as "forest" or "street". However, with regard to the mental processes involved in taking a retinal image of a scene and assigning a conceptual label to it, it may make more sense to proceed from broader categories to narrower ones.

It is also important to note that our results are consistent with Joubert et al.'s (2007) comparison across studies which indirectly supported the idea that the natural/man-made distinction is faster than those at the basic level. Our results are likewise consistent with Fei-Fei et al.'s (2007) finding that the earliest categorical distinction among scenes occurred at the superordinate level, though in their study this was the indoor/outdoor distinction. Again, these results are consistent with the general arguments made previously as to why the superordinate level may be processed earlier than the basic level, at least in certain cases. It will also be interesting to determine whether the indoor/outdoor distinction is even more primitive than the natural/man-made distinction, as suggested by Fei-Fei et al.'s (2007) results.

An open question is whether the categories that Oliva and Torralba (2001) labelled as "basic level" actually meet the criteria for "basic-levelness" put forward by Rosch et al. (1976). To our knowledge, this has never been verified by Oliva and Torralba in any of their studies. It is, of course, possible to test the validity of the assumption that the natural/man-made distinction is made prior to distinctions that Oliva and Torralba labelled as basic level, without testing the separate claim that those categories are indeed basic level according to Rosch et al. Furthermore, the claim that Oliva and Torralba's proposed basic-level categories are at a lower taxonomic level than the "natural" and "man-made" categories seems difficult to dispute. Still, the generality of the current findings regarding the (non)primacy of basic-level categories in scene gist perception could be buttressed by further studies using categories that have been validated as "basic-level" according to the criteria set forth by Rosch et al., for example by using those identified by Tversky and Hemenway (1983). It should be noted, however, that even Tversky and Hemenway's taxonomy, when applied to pictorial stimuli, was only incompletely validated.

A critical next step in advancing our understanding of scene gist processing will involve greater integration with the literature on categories and concepts. The current study takes a first step in that direction by identifying the order of processing of scenes across taxonomic levels. As suggested by the current study, further research investigating points of contact between the scene gist and concepts and categories research areas should enrich our understanding of both.

REFERENCES

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

Boyce, S., & Pollatsek, A. (1992). An exploration of the effects of scene context on object identification. In K. Rayner (Ed.), *Eye movements and visual cognition* (pp. 227–242). New York: Springer-Verlag.

Breitmeyer, B. G., & Ogmen, H. (2006). *Visual masking: Time slices through conscious and unconscious vision*. Oxford, UK: Clarendon Press.

Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology*, *13*(2), 1207–1230.

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

De Graef, P., De Troy, A., & D'Ydewalle, G. (1992). Local and global contextual constraints on the identification of objects in scenes. *Canadian Journal of Psychology*, *46*, 489–508.

Donaldson, W. (1992). Measuring recognition memory. *Journal of Experimental Psychology: General*, *121*(3), 275–277.

Eckstein, M. P., Drescher, B. A., & Shimozaki, S. S. (2006). Attentional cues in real scenes, saccadic targeting, and Bayesian priors. *Psychological Science*, *17*(11), 973–980.

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), 1–29.

Gordon, R. D. (2004). Attentional allocation during the perception of scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(4), 760–777.

Gosselin, F., & Schyns, P. G. (2001). Why do we slip to the basic level? Computational constraints and their implementation. *Psychological Review*, *108*(4), 735–758.

Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias—computing formulas. *Psychological Bulletin*, *75*(6), 424–429.

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

Jolicoeur, P., Gluck, M. A., & Kosslyn, S. M. (1984). Pictures and names: Making the connection. *Cognitive Psychology*, *16*(2), 243–275.

Joubert, O. R., Rousselet, G. A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: Fast categorization and object interference. *Vision Research*, *47*(26), 3286–3297.

Kaping, D., Tzvetanov, T., & Treue, S. (2007). Adaptation to statistical properties of visual scenes biases rapid categorization. *Visual Cognition*, *15*(1), 12–19.

Kim, J., & Chong, S. C. (2008). Do you know what it is as soon as you know it is there? *Journal of Vision*, *8*(6), 512–512.

Kovacs, G., Vogels, R., & Orban, G. A. (1995). Cortical correlate of pattern backward-masking. *Proceedings of the National Academy of Sciences of the USA*, *92*(12), 5587–5591.

Large, M. E., Kiss, I., & McMullen, P. A. (2004). Electrophysiological correlates of object categorization: Back to basics. *Cognitive Brain Research*, *20*(3), 415–426.

Levin, D. T., Takarae, Y., Miner, A. G., & Keil, F. (2001). Efficient visual search by category: Specifying the features that mark the difference between artifacts and animals in preattentive vision. *Perception and Psychophysics*, *63*(4), 676–697.

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

Loftus, G. R., & Mclean, J. E. (1999). A front end to a theory of picture recognition. *Psychonomic Bulletin and Review*, *6*(3), 394–411.

Loschky, L. C., & Larson, A. M. (2008). Localized information is necessary for scene categorization, including the natural/man-made distinction. *Journal of Vision*, *8*(1), 1–9.

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

Macmillan, N. A., & Creelman, D. C. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

McCotter, M., Gosselin, F., Sowden, P., & Schyns, P. (2005). The use of visual information in natural scenes. *Visual Cognition*, *12*(6), 938–953.

Murphy, G. L. (1991). Parts in object concepts: Experiments with artificial categories. *Memory and Cognition*, *19*(5), 423–438.

Oliva, A. (2005). Gist of a scene. In L. Itti, G. Rees, & J. K. Tsotsos (Eds.), *Neurobiology of attention* (pp. 251–256). Burlington, MA: Elsevier Academic Press.

Oliva, A., & Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. *Cognitive Psychology*, *41*(2), 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*(3), 145–175.

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

Oliva, A., Torralba, A., Castelhano, M. S., & Henderson, J. M. (2003). Top down control of visual attention in object detection. *IEEE Proceedings of the International Conference on Image Processing*, *1*, 253–256.

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

Pezdek, K., Whetstone, T., Reynolds, K., Askari, N., & Dougherty, T. (1989). Memory for real-world scenes: The role of consistency with schema expectation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(4), 587–595.

Rieger, J. W., Braun, C., Bulthoff, H. H., & Gegenfurtner, K. R. (2005). The dynamics of visual pattern masking in natural scene processing: A magnetoencephalography study. *Journal of Vision*, *5*(3), 275–286.

Rogers, T. T., & Patterson, K. (2007). Object categorization: Reversals and explanations of the basic-level advantage. *Journal of Experimental Psychology: General*, *136*(3), 451–469.

Rolls, E., Tovée, M. J., & Panzeri, S. (1999). The neurophysiology of backward visual masking: Information analysis. *Journal of Cognitive Neuroscience*, *11*(3), 300–311.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*(3), 382–439.

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*(6), 852–877.

See, J. E., Warm, J. S., Dember, W. N., & Howe, S. R. (1997). Vigilance and signal detection theory: An empirical evaluation of five measures of response bias. *Human Factors*, *39*(1), 14–29.

Swets, J. A. (1986). Form of empirical rocs in discrimination and diagnostic tasks: Implications for theory and measurement of performance. *Psychological Bulletin*, *99*(2), 181–198.

Torralba, A. (2003). Modeling global scene factors in attention. *Journal of the Optical Society of America: Optics, Image Science, and Vision*, *20A*(7), 1407–1418.

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, *113*(4), 766–786.

Tversky, B., & Hemenway, K. (1983). Categories of environmental scenes. *Cognitive Psychology*, *15*(1), 121–149.

Manuscript received November 2008
Manuscript accepted March 2009
First published online July 2009