

Localized information is necessary for scene categorization, including the Natural/Man-made distinction

Lester C. Loschky

Department of Psychology, Kansas State University,
Manhattan, KS, USA



Adam M. Larson

Department of Psychology, Kansas State University,
Manhattan, KS, USA



What information do people use to categorize scenes? Computational scene classification models have proposed that unlocalized amplitude information, the distribution of spatial frequencies and orientations, is useful for categorizing scenes. Previous research has provided conflicting results regarding this claim. Our previous research (Loschky et al., 2007) has shown that randomly localizing amplitude information (i.e., randomizing phase) greatly disrupts scene categorization at the basic level. Conversely, studies suggesting the usefulness of unlocalized amplitude information have used binary distinctions, e.g., Natural/Man-made. We hypothesized that unlocalized amplitude information contributes more to the Natural/Man-made distinction than basic level distinctions. Using an established set of images and categories, we varied phase randomization and measured participants' ability to distinguish Natural versus Man-made scenes or scenes at the basic level. Results showed that eliminating localized information by phase randomization disrupted scene classification even for the Natural/Man-made distinction, demonstrating that amplitude localization is necessary for scene categorization.

Keywords: perception, natural scenes, scene gist, recognition, phase, phase-randomization, amplitude, localization, categorization, man-made, natural, basic-level, superordinate

Citation: 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):4, 1–9, <http://journalofvision.org/8/1/4/>, doi:10.1167/8.1.4.

Introduction

While channel surfing, you can rapidly recognize a series of unrelated images such as a tennis match, a person cooking in the kitchen, a lifeguard rescuing someone at the beach, or a street battle. A key question is, what allows such rapid recognition of the overall *gist* of each scene? A recent, counter-intuitive, and thus intriguing claim has been that viewers use the unlocalized amplitude information of a scene, i.e., the distribution of spatial frequencies and orientations in an image without regard to their location, in order to categorize it (e.g., Gorkani & Picard, 1994; Guerin-Dugue & Oliva, 2000; Guyader, Chauvin, Peyrin, Héroult, & Marendaz, 2004; Héroult, Oliva, & Guerin-Dugue, 1997; Oliva & Torralba, 2001). Evidence, both consistent and inconsistent with this claim has been put forward (Guyader et al., 2004; Kaping, Tzvetanov, & Treue, 2007; Loschky et al., 2007). We hypothesized that a possible explanation of these discrepancies lies in the level of categorization measured, specifically that unlocalized amplitude information is not useful for discriminating basic level categories, such as “Highway” or “Street,” but it could be more useful in discriminating “Natural” versus

“Man-made” scenes. The current study directly tests this hypothesis.

Viewers can reliably apply a one-word category label to a scene after a masked image presentation as brief as 40–60 ms (Bacon-Macé, Macé, Fabre-Thorpe, & Thorpe, 2005; Fei-Fei, Iyer, Koch, & Perona, 2007; Loschky et al., 2007), with this process frequently called *scene gist recognition* (Oliva, 2005). Scene gist helps direct our attention in scenes (Eckstein, Drescher, & Shimozaki, 2006; Gordon, 2004; Torralba, Oliva, Castelano, & Henderson, 2006), may influence object recognition in scenes (Boyce & Pollatsek, 1992; Davenport & Potter, 2004; but see Hollingworth & Henderson, 1998) and affects later memory for scenes (Brewer & Treyns, 1981; Pezdek, Whetstone, Reynolds, Askari, & Dougherty, 1989). Thus, scene gist is important for driver safety (Shinoda, Hayhoe, & Shrivastava, 2001), eyewitness testimony (Greenberg, Westcott, & Bailey, 1998), and artificial vision (Torralba, 2003; Vailaya, Jain, & Zhang, 1998).

Several computational models have explored the idea that a scene can be categorized using only (or primarily) unlocalized amplitude information (e.g., Gorkani & Picard, 1994; Guerin-Dugue & Oliva, 2000; Guyader et al., 2004; Héroult et al., 1997; Oliva & Torralba, 2001).

A biological rationale for such models is the well-known sensitivity of complex cells in V1 to orientation and spatial frequency (De Valois & De Valois, 1988; Hubel & Wiesel, 1968). This has led to the well-known claim that scene gist recognition does not require prior recognition of the component objects in it (Oliva & Torralba, 2001). Less commented upon is the *implication* that scene categorization may be similarly independent of scene layout, i.e., knowing that a coastline image has strong horizontal and diagonal information in the low spatial frequency range would be more important than knowing that the main diagonal (waterline) descends from the main horizontal (horizon). For example, the Spatial Envelope model (Oliva & Torralba, 2001) correctly categorized images into eight basic level scene categories with 86% accuracy (vs. 12.5% chance) using only unlocalized amplitude information. Localizing that information only increased accuracy to 92%, suggesting that most of the useful information was unlocalized.

The usefulness of unlocalized amplitude information has received mixed support from human behavioral studies. Guyader et al. (2004) showed a small priming effect (15–18 ms) on scene categorization that was equivalent for normal scene primes and scene primes with normal amplitude spectrum but randomized phase (i.e., no preserved localization). In addition, Kaping et al. (2007) showed biasing of scene categorization by adaptation to noise roughly sharing the amplitude spectra of Natural versus Man-made scenes. However, scene categorization at the basic level is disrupted when phase is randomized without changing the amplitude spectrum information (Loschky et al., 2007; see also Wichmann, Braun, & Gegenfurtner, 2006 for animal detection in scenes). Converging evidence from visual gist masking studies supports the idea that amplitude localization is necessary for gist identification (Loschky et al., 2007).

An intuitively appealing explanation for these conflicting results is that unlocalized amplitude information, while not particularly useful for explicit basic level scene categorization such as “Mountain” or “Forest” (Tversky & Hemenway, 1983), may be more useful for the simpler discrimination between “Natural” versus “Man-made” scenes. In support of this hypothesis, Oliva and Torralba (Oliva & Torralba, 2001, Table 1, p. 148) found that the Natural/Man-made distinction corresponds to subjects’ first-pass scene image sorting behavior. Likewise, the Natural/Man-made distinction is the first stage of the Spatial Envelope model, which is based entirely on unlocalized amplitude information, since “the introduction of spatial information does not seem to improve the classification” (Oliva & Torralba, 2001, p. 158). This explanation is consistent with previous studies supporting the usefulness of unlocalized amplitude information for scene categorization, given that those either used the Natural/Man-made task (Kaping et al., 2007) or a task that could be treated as such by subjects (i.e., “Beach” vs. “City,” Guyader et al., 2004).

The current study tests the hypothesis, derived from the Spatial Envelope model (Oliva & Torralba, 2001), that unlocalized amplitude information is more useful for discriminating “Natural” versus “Man-made” scenes than for discriminating basic level scene categories. Its method is to compare observers’ scene category recognition for images with varying levels of phase randomization in two different tasks: 1) basic level scene categories, and 2) the Natural/Man-made distinction. The key prediction is that fully phase-randomized scenes should be more recognizable in the Natural/Man-made task than in the basic level task. If not, it would suggest that localization, via absolute phase, is necessary for normal explicit scene processing (Field, 1987, 1999; Simoncelli & Olshausen, 2001; Thomson & Foster, 1997), even for the purportedly primitive Natural/Man-made distinction. Our study follows the tradition of using demonstrations to test models of vision by generating images having characteristics predicted to be useful for image recognition and observing their effects. Such demonstrations have been used to informally test spatial vision models (Anstis, 1998; Piotrowski & Campbell, 1982; Tadmor & Tolhurst, 1993) and the Spatial Envelope model (Oliva & Torralba, 2006; Torralba & Oliva, 2002). The current study goes beyond a demonstration, however, to more formally test the above hypothesis using experimental methods.

Methods

Participants

160 Kansas State University students (120 females, Mean age = 18.91) participated for course credit. All subjects had at least 20/30 corrected to normal vision.

Stimuli

We randomly sampled 256 images from the Oliva and Torralba (2001) image set (<http://cvcl.mit.edu/database.htm>), 32 images from each of eight categories: (Natural): Coast, Forest, Mountain, Open Country; (Man-made): City Center, Highway, Street, Tall Building. All images were reduced to grayscale and put through five levels of phase randomization (0, .2, .4, .6, 1.0) using the RISE algorithm (Loschky et al., 2007; Sadr & Sinha, 2001, 2004), which included equalization of mean luminance and RMS contrast across all images and phase-randomization levels (see Figure 1). Images measured 256 × 256 pixels, and viewed at a distance of 53.34 cm from the monitor (using a chinrest) they subtended 10.12° × 10.12° visual angle. The 85 Hz Samsung SyncMaster 957 MBS monitors were calibrated for luminance and contrast.

Design and procedures

The study used a 2×5 between-subjects factorial design: 2 (task: Basic, Natural/Man-made) \times 5 (Phase randomization factor (hereafter called “RAND”): 0, 0.2, 0.4, 0.6, 1). 16 participants were randomly assigned to each of the 10 groups.

Participants were familiarized with the category labels by presenting a separate sample set of 80 labeled images, and the task, by performing 32 practice trials. The actual experiment had 256 trials. Figure 2 shows the sequence

of trial events. On each trial, participants looked at a fixation cross prompting them to push a key to display an image. The target image appeared for 24 ms, followed by a 750 ms gray screen (matched to the image set mean luminance), followed by a category cue, which remained until the participant made a “YES” or “NO” button press. Participants were encouraged to respond as quickly and accurately as possible. Each of the 256 images appeared once, and each of the cue categories ((8) basic level, or (2) Natural/Man-made) was used equally often, with equal cue validity for all

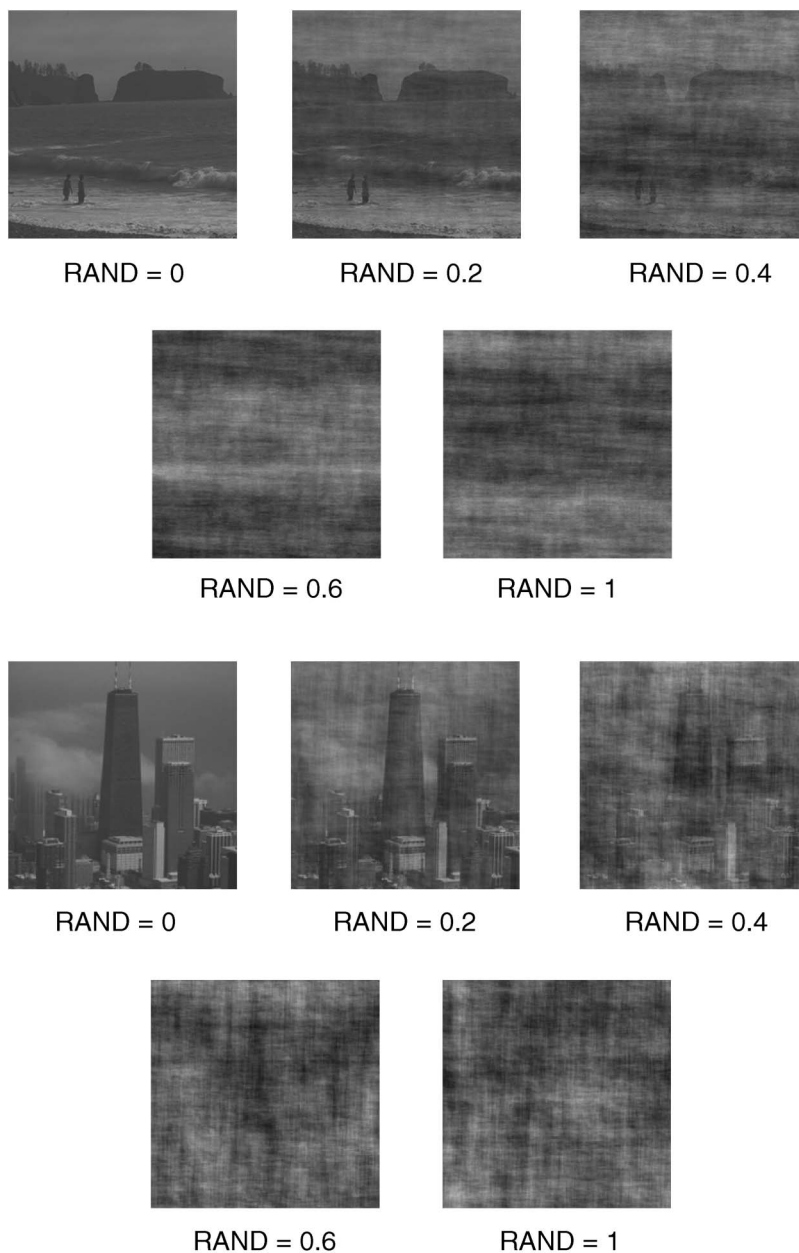


Figure 1. Two example scenes, a “Coast” (top) and “Tall Building” (bottom), each having five levels of phase randomization (range: 0–1). “RAND = 0” represents a phase randomization factor of 0 (an unaltered image); “RAND = 1” represents a phase randomization factor of 1 (completely randomized).

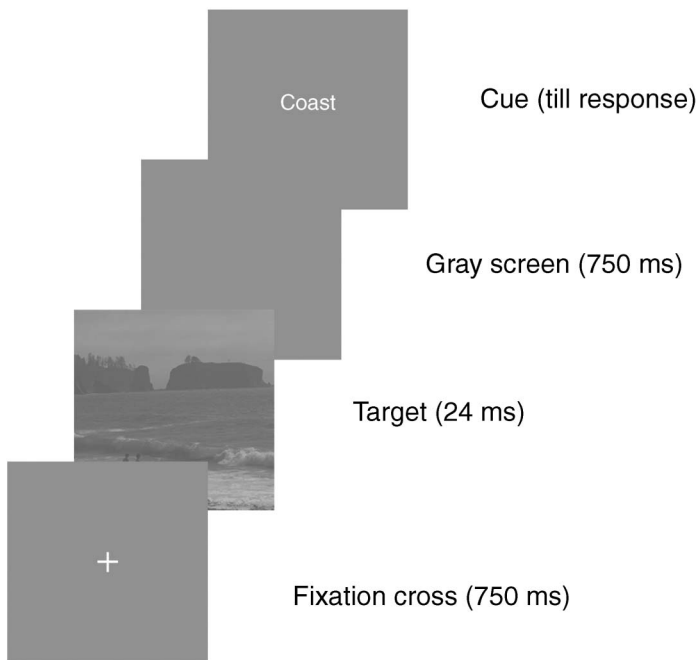


Figure 2. Schematic of the events in a trial.

categories. The basic level task average cue validity was 0.5, but due to an error in the Natural/Man-made task the average validity was 0.56. This discrepancy is addressed in the [Results](#) section.

Results

To distinguish sensitivity from response bias (potentially caused by the slight difference in cue validity across tasks), results are reported in terms of the non-parametric signal detection measures A' and B'' (Grier, 1971; Macmillan & Creelman, 2005). [Figure 3](#) shows that sensitivity greatly decreased with increasing phase randomization regardless of task ($F(4, 150) = 251.40, p < .001$). However, sensitivity also showed a cross-over interaction with task ($F(4, 150) = 7.58, p < .001$), with higher sensitivity in the Natural/Man-made task than the Basic level task for phase randomization levels at or below 0.4, and a reversal of this pattern for greater levels of phase randomization. Consistent with Loschky et al. (2007), the threshold level of phase randomization for scene categorization seems to be 0.5 (between $RAND = 0.4$ and 0.6)—sensitivity to scene categories is at a minimum when the phase randomization factor exceeds this value. Consistent with Oliva and Torralba's (2001) suggested primacy of the Natural/Man-made distinction, in the unaltered image condition ($RAND = 0$), sensitivity was greater for the Natural/Man-made task ($M = .97, SD = .02$) than for Basic level task ($M = .93, SD = .03$),

$t(30) = 4.37, p < .001$. However, inconsistent with our primary hypothesis, in the completely phase-randomized condition ($RAND = 1$), sensitivity was lower in the Natural/Man-made task ($M = .53, SD = .09$) than in the Basic level task ($M = .62, SD = .05$), $t(30) = 3.34, p = .003$ (equal variances not assumed), suggesting that unlocalized amplitude is insufficient for scene categorization, even at the primitive level of the Natural/Man-made distinction.

Bias differed between the two task conditions, with a “YES” bias in the Natural/Man-made task ($M = -.05, SD = .02$), and a “NO” bias in the basic level task ($M = .05, SD = .02$), $F(1, 150) = 14.59, p < .001$. This bias difference was only found in the most recognizable conditions ($RAND = 0-0.2$), probably because in the unrecognizable conditions ($RAND = 0.6-1$) participants were unaware of the difference in the base rate for cue validity (since their lack of sensitivity in these conditions means that they could not distinguish valid and invalid cues).

More importantly, [Figure 4](#) shows a different bias found in the Natural/Man-made task at phase randomization levels below the gist recognition threshold ($RAND = 0.6-1$). In the completely phase randomized condition ($RAND = 1$), there was a strong bias to perceive images ‘as being “Natural,”’ whereas no such bias was found for the normal unaltered images ($RAND = 0$), $\chi^2(1, N = 8189) = 435.39, p < .001$. For this analysis, responses were recoded such that “YES” to “Natural” or “NO” to “Man-made” were coded as “Natural” (and vice versa for “Man-made”). Thus, for example, in [Figure 1](#), the $RAND = 1$ Tall Building image had a .94 (= 15/16 participants)

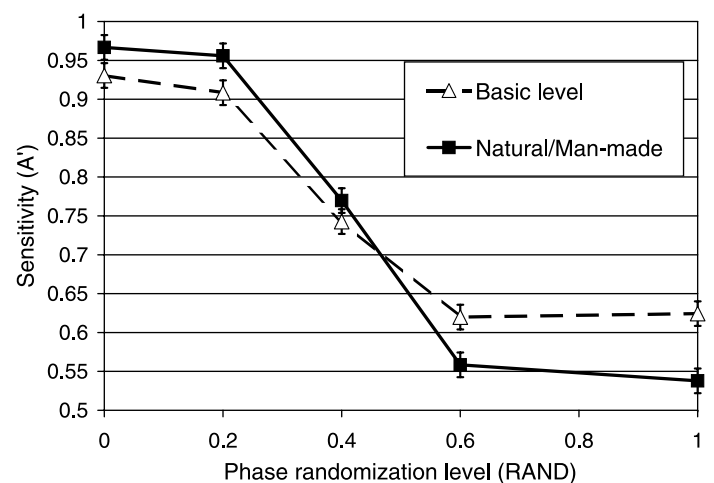


Figure 3. Sensitivity (A') to scene category as a function of phase randomization factor ($RAND = 0-1$) and categorization task (Basic level vs. Natural/Man-made). “ $RAND = 0$ ” represents a phase randomization factor of 0 (an unaltered image); “ $RAND = 1$ ” represents a phase randomization factor of 1 (completely randomized). Error bars represent 1 standard error of the mean.

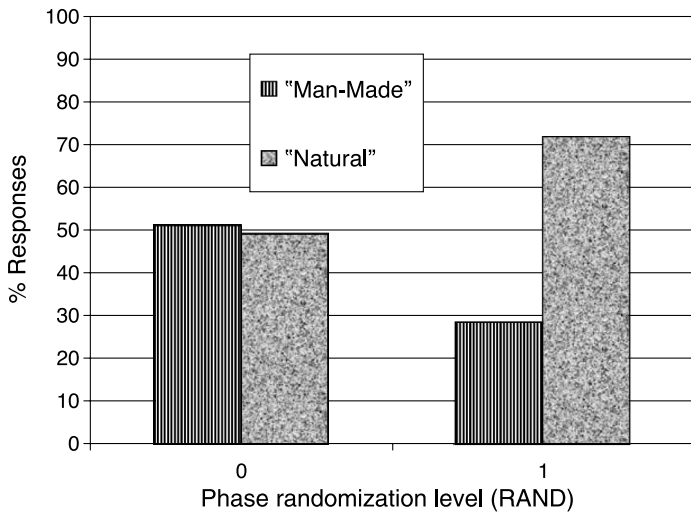


Figure 4. Percent of “Natural” versus “Man-made” responses as a function of phase randomization factor (RAND = 0 vs.1). “RAND = 0” represents a phase randomization factor of 0 (an unaltered image); “RAND = 1” represents a phase randomization factor of 1 (completely randomized).

“Natural” response rate. This “Natural” bias produced the drop in sensitivity in the Natural/Man-made task for images below the gist recognition threshold, because it increased false alarms to “Natural” cues and misses for “Man-made” cues. An explanation for the “Natural” bias is that below the gist recognition threshold, images lose sharp edges (due to a lack of phase alignment across spatial frequencies), and that this signals the “Natural” category. Conversely, then, the presence of sharp edges should signal the “Man-made” category.

We carried out a further analysis to determine which, if any, basic level distinctions viewers were able to make when phase-randomization has exceeded the 0.5 level. The analysis was therefore limited to the Basic level task, and included both the RAND = 0.6 and 1.0 conditions, which had identical accuracy levels (see Figure 3). To determine the Basic level distinctions viewers could make, we analyzed the percent correct for each image category

as a function of the cue category (see Table 1). The main diagonal of Table 1 represents the percent of hits (vs. misses) for each validly cued image category, while off-diagonal cells represent the percent of correct rejections (vs. false alarms) for invalid cues. The main diagonal shows clear evidence of the Natural Bias, in that viewers’ hit rates were all greater than 50% for the “Natural” categories and all less than 50% for the “Man-made” categories. The off-diagonal cells allow us to ask, for each image category, whether viewers were better at discerning that certain cues were invalid compared to others. Further inspection of Table 1 shows more evidence of the “Natural” bias. Specifically, correct rejection rates were high for pairs of “Natural” images with invalid “Man-made” cues (in the upper right quadrant), whereas correct rejections were lower (i.e., false alarms were higher) for pairs of “Man-made” images with invalid “Natural” cues (in the lower left quadrant). The upper left and lower right quadrants of Table 1 are made up of inverse pairings of image and cue categories. Inverse pairings, for example “Forest” images paired with “Street” cues (in the upper right quadrant, M = 75.4) versus “Street” images paired with “Forest” cues (in the lower left quadrant, M = 30.7), should have equal correct rejection rates if there is no bias. The differences clearly show a bias.

Evidence for the “Natural” bias shown in Table 1 becomes even clearer when we set a threshold of 75% correct (midway between 50% chance and 100% perfect) for further analysis, as shown in Table 2. Each row of Table 2 shows, from left to right, a combination of image and cue category that produced a correct rejection rate of ≥75%, followed by the inverse pairing of image and cue categories and corresponding correct rejection rate. A clear pattern is that all (but one) of the image and cue pairings that produced correct rejection rates ≥75% involved an implicit Natural/Man-made distinction. (We discuss the one exception below.) Consistent with the Spatial Envelope model (Oliva & Torralba, 2001), this suggests that the Natural/Man-made distinction underlies basic level distinctions. However, this process was highly influenced by the “Natural” bias. Specifically, all of the pairings that produced high correct rejection rates were for

Image Category	Cue Category							
	Coast	Forest	Mountain	Open Country	City Center	Highway	Street	Tall Building
Coast	68.6	53.7	63.2	50.8	81.1	62.3	60.3	82.8
Forest	79.3	74.7	50.6	51.4	71.6	82.1	75.4	83.6
Mountain	47.7	37.5	57.4	41.5	74.0	66.2	77.0	79.7
Open Country	36.4	41.8	53.1	61.7	76.5	56.8	70.7	80.3
City Center	64.5	25.4	55.6	59.7	41.4	72.1	63.9	65.8
Highway	36.1	53.7	55.9	43.0	75.7	37.6	71.1	88.2
Street	71.4	30.7	39.4	64.3	63.1	72.5	28.9	61.2
Tall Building	72.0	37.1	56.9	71.8	47.9	64.1	62.5	49.4

Table 1. Percent correct for image and basic level cue category pairings. All images had >0.5 phase randomization.

Image and Cue Category Pairs with $\geq 75\%$ Correct Rejection			Inversely Ordered Image and Cue Category Pairs		
Image	Cue	%CR	Image	Cue	%CR
Coast	City Center	81.1	City Center	Coast	64.5
Coast	Tall Building	82.8	Tall Building	Coast	72.0
Forest	Highway	82.1	Highway	Forest	53.7
Forest	Street	75.4	Street	Forest	30.7
Forest	Tall Building	83.6	Tall Building	Forest	37.1
Mountain	Street	77.0	Street	Mountain	39.4
Mountain	Tall Building	79.7	Tall Building	Mountain	56.9
Open Country	City Center	76.5	City Center	Open Country	59.7
Open Country	Tall Building	80.3	Tall Building	Open Country	71.8
Average		79.83	Average		53.98
Forest	Coast	79.3	Coast	Forest	53.7

Table 2. Image and basic level cue category pairings producing high correct rejection rates ($\geq 75\%$), and their corresponding inversely ordered pairings. All images had >0.5 phase randomization. “% CR” = percentage of correct rejection (vs. False alarms).

“Natural” image categories paired with “Man-made” cue categories (left half of Table 2, M correct rejection = 79.83%). There were much lower correct rejection rates for the corresponding inverse pairings of “Man-made” image categories and “Natural” cue categories (right half of Table 2, M correct rejection = 53.98%). This asymmetry suggests the Natural Bias because viewers readily rejected a “Man-made” cue category as describing a phase-randomized “Natural” image, while they also readily false alarmed to a “Natural” cue category describing a phase-randomized “Man-made” image.

The only image and cue category pairing with a high correct rejection rate that did not involve an implicit Natural/Man-made distinction was that of “Forest” images and “Coast” cues (Table 2, bottom row). Interestingly, the correct rejection rate for this pairing was virtually identical to the average for “Natural” images paired with “Man-made” cues. Likewise, the much lower correct rejection rate for “Coast” images paired with “Forest” cues was virtually identical to the average for “Man-made” images paired with “Natural” cues. Thus, paradoxically, the “Forest” and “Coast” contrast seems to epitomize the Natural Bias.

We note that there were two combinations of image and cue category that produced relatively symmetrical and high correct rejection rates: “Coast” images with “Tall Building” cues (or vice versa) and “Open Country” images with “Tall Building” cues (or vice versa). For both combinations, the distinction would seem to depend on a strong contrast between the dominant orientations associated with the two categories (horizontal vs. vertical), which may underlie an implicit Natural/Man-made distinction. The highly discriminable “Coast” versus “Tall Building” pairing is essentially the same as the “Beach” versus “City” pairing Guyader et al. (2004) used to show priming of scene gist by unlocalized amplitude informa-

tion, which suggests a resolution to the apparently conflicting results between studies of the role of unlocalized amplitude information in scene categorization. Specifically, while the discriminability of these pairs of categories is consistent with claims that unlocalized amplitude information is useful for scene categorization (Gorkani & Picard, 1994; Guerin-Dugue & Oliva, 2000; Guyader et al., 2004; Kaping et al., 2007; Oliva & Torralba, 2001; Oliva, Torralba, Guerin-Dugue, & Hurlault, 1999), the limited scope of the utility of this information is also very clear (i.e., only 2 of the 28 unique pairings of 8 image and 8 cue categories).

Discussion and conclusions

The current results are consistent with those of Loschky et al. (2007), as well as other studies that have shown the necessity of localization (via absolute phase information) for object recognition or animal detection (Sadr & Sinha, 2004; Wichmann et al., 2006). Conversely, the results fail to support the claimed usefulness of unlocalized amplitude information in explicit scene categorization, even for the most primitive Natural/Man-made scene distinction.

How can we reconcile the current results with studies showing adaptation or priming effects on scene categorization based on generalized amplitude information (Guyader et al., 2004; Kaping et al., 2007)? In the current study, observers attempted to recognize the category of fully phase-randomized scenes but could not do so, except in the case of a very limited number of category contrasts (e.g., “Coast” vs. “Tall Building”). In Guyader et al. (2004), observers were primed by fully phase-randomized

scenes prior to recognizing unaltered scene images, and showed a 15–18 ms priming effect, when contrasting same-category priming with priming across the “Beach” versus “City” contrast. In sum, it seems that unlocalized amplitude information may play a facilitative role in scene categorization for a limited number of category contrasts, but, by itself, such information is far from sufficient to categorize most scenes.

Interestingly, the current study strongly suggests that the presence or absence of sharp edges serves as a discriminative feature in distinguishing Natural versus Man-made scenes. Wichmann et al. (2006) have shown that decrements in image categorization that accompany phase randomization are specifically due to the loss of “features such as local edges” (p. 1526) rather than simple reductions in contrast. The current study has shown that such a lack of sharp edges greatly increased the likelihood of categorizing scenes as “Natural” in images having more than 50% phase-randomization (i.e., the gist recognition threshold). An interesting question is whether this “Natural” bias is somehow related to the “Outside” bias found by Fei-Fei et al. (2007).

On the surface, the fact that there is a “Natural” bias could seem problematic for studies using phase-randomized scenes to investigate the role of unlocalized amplitude information on scene gist. The problem is that responses of subjects will be biased. Nevertheless, studies using phase-randomized scenes have been very useful in showing that unlocalized (or randomly localized) amplitude information is insufficient to recognize scene gist. Importantly, careful consideration suggests that it is this very *lack* of localized information that *produces* the “Natural” bias.

Consistent with the Spatial Envelope Model (Oliva & Torralba, 2001), the current study suggests that it is easier to make the Natural/Man-made distinction than to distinguish basic level scene categories, such as “Mountain,” “Forest,” or “Street” (Tversky & Hemenway, 1983). This conflicts with the standard finding that basic level categories are easier to process than superordinate categories, of which “Natural” and “Man-made” are clear examples (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976; Tversky & Hemenway, 1983). Gosselin and Schyns (2001) point out that the level at which categorizations are easiest depends on “the ease with which [a category] can be accessed from its defining features” (p. 738). Generally, intermediate level “basic” categories are accessed most easily because they have more redundant features than superordinate or subordinate level categories. However, Gosselin and Schyns (2001, p. 742) point out that superordinate categories can be accessed very easily if they depend on “a set of singly sufficient values” (e.g., Murphy, 1991, Experiment 5). This is consistent with Oliva and Torralba’s (2001) claim that the Natural/Man-made distinction relies on highly redundant, primitive features, such as edges at dominant orientations (horizontal vs. vertical). Nevertheless, the current results

strongly suggest that even such simple features must have sufficient localization to form configurations if they are to be useful in discriminating scene categories, including the primitive Natural/Man-made distinction.

Acknowledgments

This study contains some information that was presented at the Annual Meeting of the Vision Sciences Society (2007), with the abstract published in the *Journal of Vision*. This work was supported by funds from the Kansas State University Office of Research and Sponsored Programs, and by the NASA Kansas Space Grant Consortium. The authors wish to acknowledge the work of Elise Matz, Ben Bilyeu, and Laura Artman who helped carry out the experiments, and Scott Smerchek, who programmed them. The authors also wish to thank Daniel Simons for his helpful discussions of the paper.

Commercial relationships: none.

Corresponding author: Lester Loschky.

Email: loschky@ksu.edu.

Address: Department of Psychology, Kansas State University, 471 Bluemont Hall, Manhattan, Kansas 66506-5302.

References

- Anstis, S. M. (1998). Picturing peripheral acuity. *Perception, 27*, 817–825. [PubMed]
- 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. [PubMed]
- 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.
- Brewer, W. F., & Treyens, J. C. (1981). Role of schemata in memory for places. *Cognitive Psychology, 13*, 1207–1230.
- Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science, 15*, 559–564. [PubMed]
- De Valois, R. L., & De Valois, K. K. (1988). *Spatial vision*. New York, NY, US: Oxford University Press.
- Eckstein, M. P., Drescher, B. A., & Shimozaki, S. S. (2006). Attentional cues in real scenes, saccadic targeting, and Bayesian priors. *Psychological Science, 17*, 973–980. [PubMed]

- 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):10, 1–29, <http://journalofvision.org/7/1/10/>, doi:10.1167/7.1.10. [PubMed] [Article]
- Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A, Optics and Image Science*, 4, 2379–2394. [PubMed]
- Field, D. J. (1999). Wavelets, vision and the statistics of natural scenes. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, 357, 2527–2542.
- Gordon, R. D. (2004). Attentional allocation during the perception of scenes. *Journal of Experimental Psychology: Human Perception and Performance*, 30, 760–777. [PubMed]
- Gorkani, M. M., & Picard, R. W. (1994). Texture orientation for sorting photos “at a glance”. *Computer Vision and Image Processing*, 1, 459–464.
- Gosselin, F., & Schyns, P. G. (2001). Why do we SLIP to the basic level? Computational constraints and their implementation. *Psychological Review*, 108, 735–758. [PubMed]
- Greenberg, M. S., Westcott, D. R., & Bailey, S. E. (1998). When believing is seeing: The effect of scripts on eyewitness memory. *Law and Human Behavior*, 22, 685–694. [PubMed]
- Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin*, 75, 424–429. [PubMed]
- Guerin-Dugue, A., & Oliva, A. (2000). Classification of scene photographs from local orientations features. *Pattern Recognition Letters*, 21, 1135–1140.
- Guyader, N., Chauvin, A., Peyrin, C., Héroult, J., & Marendaz, C. (2004). Image phase or amplitude? Rapid scene categorization is an amplitude-based process. *Comptes Rendus Biologies*, 327, 313–318. [PubMed]
- Héroult, J., Oliva, A., & Guerin-Dugue, A. (1997). Scene categorisation by curvilinear component analysis of low frequency spectra. In M. Verleysen (Ed.), *Proceedings of the 5th European symposium on artificial neural networks* (pp. 91–96). Bruges, Belgium: D Facto, Brussels: Belgium.
- Hollingworth, A., & Henderson, J. M. (1998). Does consistent scene context facilitate object perception? *Journal of Experimental Psychology: General*, 127, 398–415. [PubMed]
- Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, 195, 215–243. [PubMed] [Article]
- Kaping, D., Tzvetanov, T., & Treue, S. (2007). Adaptation to statistical properties of visual scenes biases rapid categorization. *Visual Cognition*, 15, 12–19.
- Loschky, L. C., Sethi, A., Simons, D. J., Pydimarri, T. N., Ochs, D., & Corbeille, J. L. (2007). The importance of information localization in scene gist recognition. *Journal Of Experimental Psychology: Human Perception & Performance*, 33, 1431–1450. [PubMed]
- Macmillan, N. A., & Creelman, D. C. (2005). *Detection theory: A user’s guide* (2nd ed.).
- Murphy, G. L. (1991). Parts in object concepts: Experiments with artificial categories. *Memory & Cognition*, 19, 423–438. [PubMed]
- 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., & Torralba, A. (2001). Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42, 145–175.
- Oliva, A., & Torralba, A. (2006). Building the gist of a scene: The role of global image features in recognition. In (vol. 155, pp. 23–36).
- Oliva, A., Torralba, A., Guerin-Dugue, A., & Héroult, J. (1999). Global semantic classification using power spectrum templates. In *Proceedings of The Challenge of Image Retrieval, Electronic Workshops in Computing series* (pp. 1–12). Newcastle: Springer-Verlag.
- 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, 587–595.
- Piotrowski, L. N., & Campbell, F. W. (1982). A demonstration of the visual importance and flexibility of spatial-frequency amplitude and phase. *Perception*, 11, 337–346. [PubMed]
- Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8, 382–439.
- Sadr, J., & Sinha, P. (2001). Exploring object perception with random image structure evolution (No. Memo #2001-06): Massachusetts Institute of Technology, Artificial Intelligence Laboratory.
- Sadr, J., & Sinha, P. (2004). Object recognition and random image structure evolution. *Cognitive Science*, 28, 259–287.
- Shinoda, H., Hayhoe, M.-M., & Shrivastava, A. (2001). What controls attention in natural environments? *Vision Research*, 41, 3535–3545. [PubMed]

- Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience*, *24*, 1193–1216. [[PubMed](#)]
- Tadmor, Y., & Tolhurst, D. J. (1993). Both the phase and the amplitude spectrum may determine the appearance of natural images. *Vision Research*, *33*, 141–145. [[PubMed](#)]
- Thomson, M. G. A., & Foster, D. H. (1997). Role of second- and third-order statistics in the discriminability of natural images. *Journal of the Optical Society of America*, *14*, 2081–2090.
- Torralba, A. (2003). Modeling global scene factors in attention. *Journal of the Optical Society of America, A, Optics, Image Science and Vision*, *20*, 1407–1418. [[PubMed](#)]
- Torralba, A., & Oliva, A. (2002). Depth estimation from image structure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*, 1226–1238.
- Torralba, A., Oliva, A., 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*, 766–786. [[PubMed](#)]
- Tversky, B., & Hemenway, K. (1983). Categories of environmental scenes. *Cognitive Psychology*, *15*, 121–149.
- Vailaya, C., Jain, A., & Zhang, H. J. (1998). On image classification: City images vs. landscapes. *Pattern Recognition*, *31*, 1921–1935.
- Wichmann, F. A., Braun, D. I., & Gegenfurtner, K. (2006). Phase noise and the classification of natural images. *Vision Research*, *46*, 1520–1529. [[PubMed](#)]