Categorization influences detection: A perceptual advantage for representative exemplars of natural scene categories

Eamon Caddigan Department of Psychology, University of Illinois, Champaign, IL, USA

Heeyoung Choo Beckman Institute, University of Illinois, Urbana, IL, USA

Li Fei-Fei Stanford University, Palo Alto, CA, USA

Diane M. Beck Department of Psychology and Beckman Institute, University of Illinois, Champaign, IL, USA

Traditional models of recognition and categorization proceed from registering low-level features, perceptually organizing that input, and linking it with stored representations. Recent evidence, however, suggests that this serial model may not be accurate, with object and category knowledge affecting rather than following early visual processing. Here, we show that the degree to which an image exemplifies its category influences how easily it is detected. Participants performed a two-alternative forced-choice task in which they indicated whether a briefly presented image was an intact or phase-scrambled scene photograph. Critically, the category of the scene is irrelevant to the detection task. We nonetheless found that participants "see" good images better, more accurately discriminating them from phase-scrambled images than bad scenes, and this advantage is apparent regardless of whether participants are asked to consider category during the experiment or not. We then demonstrate that good exemplars are more similar to same-category images than bad exemplars, influencing behavior in two ways: First, prototypical images are easier to detect, and second, intact good scenes are more likely than bad to have been primed by a previous trial.

Introduction

Human observers are surprisingly adept at categorizing briefly presented natural scenes. Not only are they able to extract scene category from very short presentation durations (e.g., <50 ms; Walther, Caddigan, Fei-Fei, & Beck, 2009), but category judgments themselves can be made in less than 350 ms (VanRullen & Thorpe, 2001a). Similarly, event-related potential data show that the brain differentiates categories of scenes in 150 ms (Thorpe, Fize, & Marlot, 1996) and under some conditions possibly even less time (VanRullen & Thorpe, 2001b). Finally, all of this can be done under conditions of limited attention (Li, VanRullen, Koch, & Perona, 2002; Rousselet, Fabre-Thorpe, & Thorpe, 2002).

Such fast and efficient categorization is surprising given traditional models of recognition. Most recognition models proceed from registering low-level features, perceptually segmenting and organizing that input, and culminating in recognition and categorization processes that link the visual input with stored learned representations (Bregman, 1981; Driver & Baylis, 1996; Marr, 1982; Nakayama, He, & Shimojo, 1995; Palmer & Rock, 1994a, 1994b; Rubin, 1958). In recent years, these serial models of vision, which dominated theoretical models of vision for almost half a century, have given way to predictive coding models that argue against a unidirectional view of vision (Bullier, 2001; Chen et al., 2007; Hochstein & Ahissar, 2002; Panichello et al., 2012; Rao & Ballard, 1999). Rather than simply building a representation with each successive step in the visual hierarchy, it is posited that later areas not only make a prediction based on the input from hierarchically earlier areas but send it back to the earlier area, thus generating an error signal that can be used to iteratively shape the signal in line with both the input and predictions. It follows from such models that stimuli that conform to expectations should emerge more quickly than those that do not.

Citation: Caddigan, E., Choo, H., Fei-Fei, L., & Beck, D. M. (2017). Categorization influences detection: A perceptual advantage for representative exemplars of natural scene categories. *Journal of Vision*, *17*(1):21, 1–11, doi:10.1167/17.1.21.

doi: 10.1167/17.1.21 Received November 7, 2015; published January 12, 2017 ISSN 1534-7362

In keeping with a model in which prediction influences even early visual processes, regions in a two-toned image that denote a meaningful object are more likely seen as a figure than their nonmeaningful counterparts (Peterson, Harvey, & Weidenbacher, 1991; Peterson & Gibson, 1994). More important with respect to the current work, Grill-Spector and Kanwisher (2005) have shown not only that participants are very fast at categorizing natural images but also that this categorization occurs in the same time frame as simply detecting the presence of a natural image. This result argues against a model of visual processing in which detection precedes categorization and instead suggests that observers can categorize images as soon as they can see that the image is meaningful. However, at best (see Bowers & Jones, 2008, and Mack, Gauthier, Sadr, & Palmeri, 2008, for alternatives), these data indicate only that detection and categorization co-occur. In the present study, we ask whether categorization not only co-occurs with detection but actually influences it. In particular, we ask whether images that are more easily categorized are actually detected more readily.

Torralbo and colleagues (2013) found that briefly presented natural scene exemplars that were rated as more representative of their category ("good" images) were later categorized by a separate group of participants more quickly and accurately than exemplars that were rated as less representative ("bad" images). Such a result is in line with numerous typicality results (Rosch, Simpson, & Miller, 1976); participants are more quickly able to categorize good examples simply because they more readily evoke the concept of their category. Such results were not taken to mean that perception is better or worse for good and bad examplars but rather that once perceived, some images made better contact with the conceptual category. Here, we take this effect a step further and ask whether good exemplars are actually "seen" better than bad exemplars. Specifically, we ask whether human observers are better able to discriminate briefly presented intact scenes from phase-scrambled versions when the images are good exemplars rather than bad exemplars of a scene category. Critically, scene category is not relevant to the detection task; participants are not asked what the image is but simply whether it is a coherent image of any sort (as opposed to noise). If prediction is part of the perceptual process, however, then not only will participants know "what" the image is at the same time that they know whether it is intact but they should also be able to make the intact judgment more readily when the image is representative of its category.

Images were presented either in their original, intact state or with their power-spectrum amplitude intact but their phases randomized ("phase-scrambled"; Sadr & Sinha, 2004). Phase scrambling maintains an image's amplitude spectrum but disrupts its structure; disruption of local information is known to impair categorization performance (Loschky et al., 2007; Vogel, Schwaninger, Wallraven, & Bulthoff, 2007) but produces images that are similar to scenes in their first-order image statistics. Participants responded to each image by simply indicating whether it was intact or phase scrambled. We used 100% phase scrambling because our purpose was to create stimuli in which no discernible structure was present, ensuring that participants were simply judging whether or not an intact image was present. It is important to note that scene category is completely irrelevant to our detection task. We were interested in whether participants' sensitivity (d') to the intact/phase-scrambled distinction differed for good and bad exemplars. Because participants are not judging what is out there, just whether there is an intact image or not, a difference in sensitivity to good versus bad exemplars would indicate that good exemplars are actually perceived better.

We also asked whether such an effect may be dependent on whether or not observers perceived scene category as an important part of the experiment. Thus, in three experiments, we manipulated whether category was relevant to the participant by having participants perform an additional rating task after each intact/scrambled judgment; one group of participants made a category-related judgment and were informed of the categories used in the experiment (Experiment 1), a second group simply indicated whether the images were seen clearly (Experiment 2) and no reference to category was mentioned, and a third group performed only the intact/scrambled judgment.

Experiment 1

Does category representativeness influence scene detection? Participants performed a two-alternative forced-choice discrimination task, indicating whether a briefly presented image was intact or phase scrambled (Sadr & Sinha, 2004). Viewers' ability to successfully discriminate an unaltered photograph from a phase-scrambled image would imply the detection of coherent local structure in that image (Loschky et al., 2007), whereas a failure to do so would indicate that an image of a natural scene was not perceived as such. In other words, sensitivity to the intact/scrambled distinction provides a measure of whether a coherent image was detected or not.

In an attempt to make the category of the images relevant to the participants, we instructed them to retain the list of categories used in the experiment and then, after each intact/scrambled discrimination re-

sponse, rate the same image in relation to its category; in particular, they were to indicate how well the preceding image exemplified its category on a five-point scale.

Method

Participants

Eighteen participants from the University of Illinois took part in these experiments for course credit in an introductory psychology course. All had normal or corrected-to-normal vision and gave written informed consent according to the procedures of the University of Illinois Institutional Review Board.

Stimuli

Full-color natural images were drawn from a set of 4,025 images of beaches, city streets, forests, highways, mountains, and offices. These six categories were selected in an attempt to capture a representative sample of natural and man-made environments. Each image was rated to indicate how representative it was of its category by workers via the Internet (Torralbo et al., 2013). Briefly, workers using Amazon Mechanical Turk rated each image on a scale from 1 (*poor*) to 5 (*good*) or indicated that it did not belong to the specified category (see Torralbo et al., 2013, for more details). If more than 25% of the workers indicated the image was not from the category, it was removed from the data set. For each of the six categories, we selected 40 "good," 40 "medium," and 40 "bad" images based on their mean ratings (mean scores were 4.70, 3.99, and 2.88, respectively). Scenes were phase scrambled by combining in the Fourier domain the amplitude of an intact scene with the phase from a random noise image and taking the inverse fast Fourier transform of this hybrid image. Examples of intact and scrambled good and bad exemplars are shown in Figure 1. Perceptual masks were colored images of white noise at multiple spatial frequencies with naturalistic textures overlaid used in previous studies of rapid scene categorization (Torralbo et al., 2013; Walther et al., 2009). All images were presented at a resolution of 800 $\times$ 600 pixels and subtended approximately 30° of visual angle.

Procedure

Before beginning the task, participants were presented with a list of the categories used in the experiment and asked to use these categories when rating how well the images exemplify their category. After being instructed on the task, participants performed 25 blocks of 30 trials each. The first nine blocks were used for staircasing stimulus onset asynchrony (SOA) and consisted of "medium" category exemplars drawn randomly with replacement. The SOA between target image and mask was staircased to 70% accuracy individually for each participant using the QUEST algorithm (Watson & Pelli, 1983). The SOA in the staircasing phase of the experiment began at 500 ms and was adjusted over the course of 270 trials to produce an average accuracy of approximately 70%. There was no interstimulus interval between image and mask; thus, adjusting the SOA amounted to adjusting the duration of the target image. SOA for the remaining 16 testing blocks was fixed at the mean of the probability density function obtained during staircasing (34–78 ms; mean across the experiments = 49 ms). The testing blocks consisted of "good" and "bad" category exemplars drawn randomly without replacement from each of the six categories. Each trial proceeded as follows: A fixation cross was presented at the center of the screen for 500 ms, followed by the presentation of the target intact or phase-scrambled image (with a fixation cross superimposed) at the SOA determined during staircasing. Immediately following the image, a perceptual mask was presented for 500 ms. Participants then indicated whether the image was intact or phase scrambled (each condition accounting for 50% of the trials) by pressing one of two keys on a computer keyboard. Trials timed out if participants failed to respond within 1600 ms after the offset of the mask; these trials were excluded from further analysis (less than 1% of trials were removed from all experiments, and no difference was observed between conditions). No feedback was given.

During the testing phase of the experiment, participants performed an additional task after making each intact/scrambled response. Participants were asked to rate how well each image exemplified its category by pressing a number between 1 (*poor example*) and 5 (*very good example*). Instructions given at the beginning of the experiment described this task, and participants were told to covertly categorize each image to make the judgment.

Results and discussion

Overall, participants were 85% accurate on the intact/scrambled distinction and needed only an average image duration of 47 ms to achieve that accuracy. Results are summarized in Table 1. Sensitivity for intact images was measured by calculating d′ for participants' intact/scrambled responses, with images correctly identified as intact classified as hits and those scrambled images falsely labeled as intact classified as false alarms. In keeping with our predictions, we observed a significant difference between d′

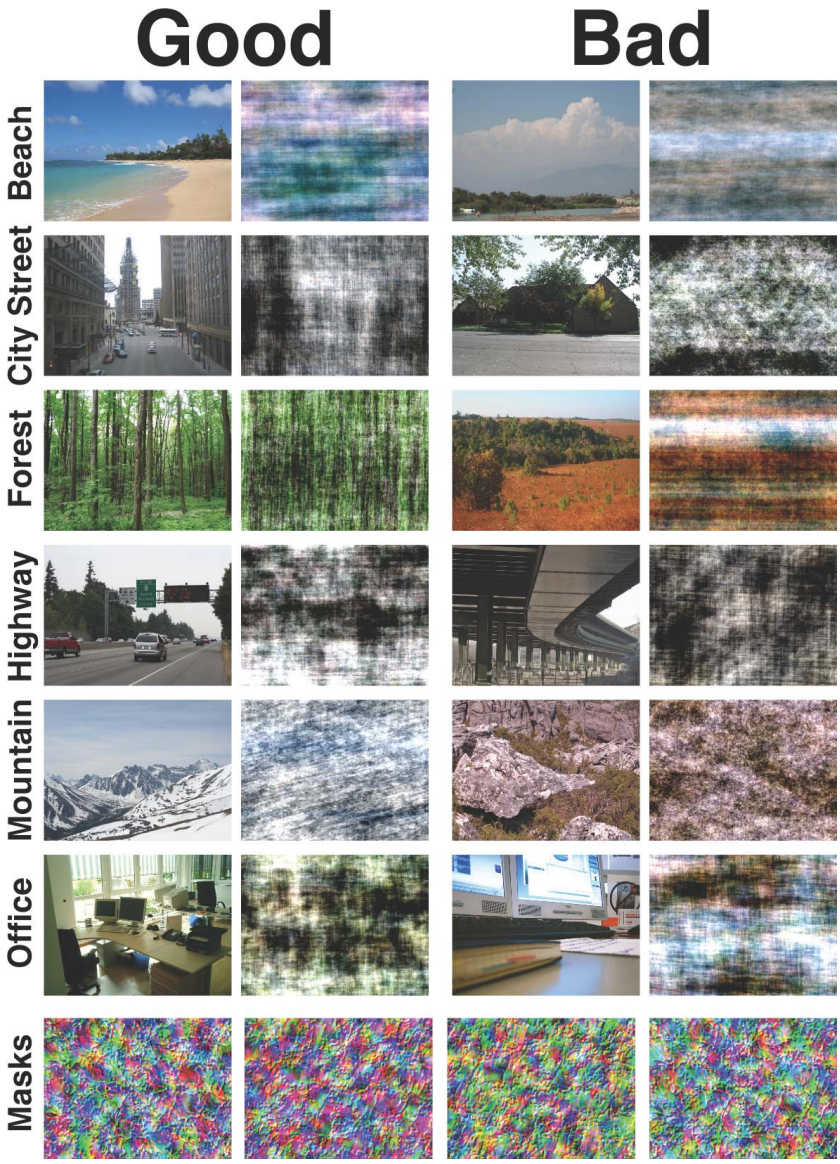

Figure 1. Examples of the stimuli used in the experiments. Intact and phase-scrambled versions of good and bad exemplars from the six image categories used in the experiment are shown, along with examples of the perceptual masks. Participants were asked to indicate whether a briefly presented scene was intact or scrambled, irrespective of its category or representativeness.

	Hit rate	False alarm rate	Sensitivity (d′)	Bias	Response time	Ratings
Experiment 1 (with clarity rating)						
Bad	0.84 ± 0.02	0.14 ± 0.02	2.26 ± 0.19	0.04 ± 0.06	1031 ± 67	3.81 ± 0.18
Good	0.87 ± 0.02	0.15 ± 0.02	2.45 ± 0.21	−0.05 ± 0.06	1017 ± 76	4.16 ± 0.15
Experiment 2 (with clarity rating)						
Bad	0.86 ± 0.02	0.18 ± 0.04	2.26 ± 0.17	−0.04 ± 0.10	966 ± 82	3.57 ± 0.21
Good	0.90 ± 0.02	0.17 ± 0.03	2.51 ± 0.20	−0.14 ± 0.08	950 ± 84	3.79 ± 0.19
Experiment 3 (no rating task)						
Bad	0.74 ± 0.04	0.38 ± 0.05	1.13 ± 0.14	−0.18 ± 0.14	591 ± 43	N/A
Good	0.78 ± 0.04	0.38 ± 0.05	1.34 ± 0.19	−0.27 ± 0.14	587 ± 43	N/A

Table 1. Participant performance on good and bad images for Experiments 1, 2, and 3.

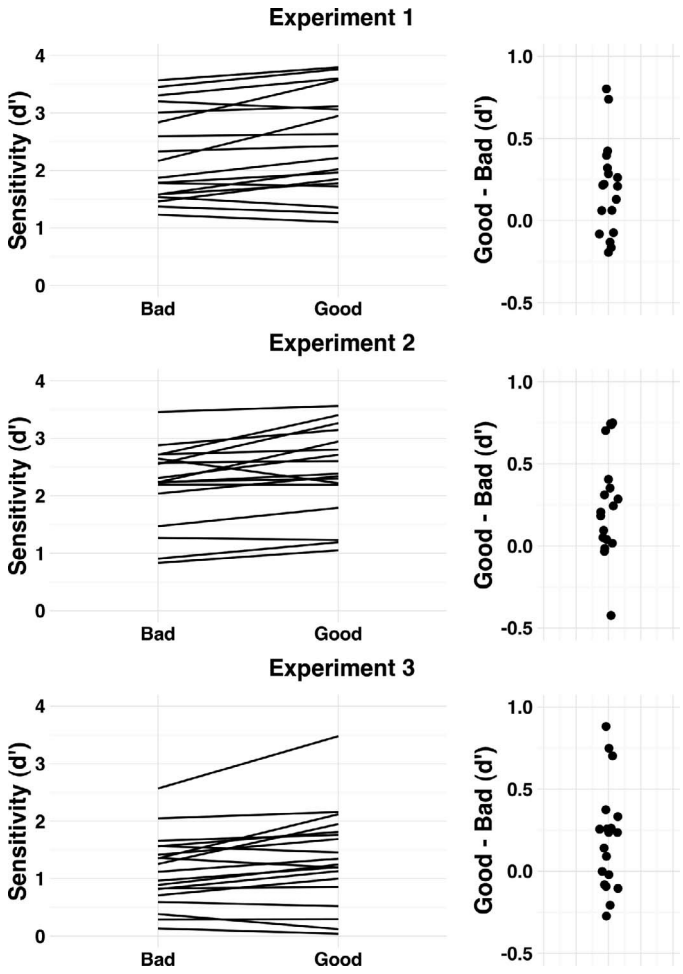

Figure 2. Intact/scrambled discriminations for Experiments 1–3. (Left) Sensitivity for intact versus scrambled image discrimination for "good" images (rated as high in representativeness) and "bad" images (those rated as low in representativeness) for all participants. (Right) The difference between sensitivity for good and bad images for each participant.

for good and bad images (2.45 vs. 2.26), $t(17) = 2.94$, $p = 9.2 \times 10^{-03}$; Cohen's $d = 0.23$, $d_z = 0.69$, such that participants were better able to discriminate intact from scrambled images if they were good category exemplars than if they were bad (see Figure 2). The difference in sensitivity was driven by higher hit rates (i.e., more correct responses to intact trials) for good versus bad images (87% vs. 84%), $t(17) = 2.64$, $p = 0.02$. False alarm rates (i.e., correct responses to scrambled trials: 15% vs. 14%), $t(17) = 0.60$, $p = 0.56$, and response times (1017 ms vs. 1031 ms), $t(17) = -0.55$, $p = 0.59$, showed no significant difference for good and bad images, nor was there a statistically significant difference in response bias (–0.05 vs. 0.04), $t(17) = -1.97$, $p = 0.07$.

We stress that to make this distinction, participants need detect only some coherent structure in the image to be able to rule out that it was phase scrambled. Mistakes occur because presentations are so brief that observers often experience just a flash. Importantly, however, they were less likely to experience this incoherent flash if the image was a good exemplar of its category.

The image ratings used to determine good and bad category exemplars were obtained with different participants and under different viewing conditions than those used in the current study. It is therefore possible that this distinction would be lost due to the short presentation times and perceptual masking present in this experiment. However, we again observed a higher average good/bad rating for intact good images than that for bad (4.16 vs. 3.81, $t(17) = 5.33$, $p = 5.6 \times 10^{-5}$) in the rating task performed in the present experiment. This effect was not driven by the fact that participants failed to see more of the bad exemplars as intact, as the difference remained significant when only trials with correct "intact" responses were considered (4.43 vs. 4.10, $t(17) = 5.00$, $p = 1.1 \times 10^{-5}$). No difference was observed for ratings made to scrambled images (2.34 vs. 2.37), $t(17) = 0.75$, $p = 0.46$. In other words, the good/bad distinction in judgments of representativeness was only apparent when participants correctly detected coherent structure in the image. Nonetheless, this distinction influenced the probability that they would see an image as containing coherent structure.

Experiment 2: Clarity

In Experiment 1, we explicitly evoked the scene categories with our instructions and secondary rating task. In Experiment 2, we asked whether such instructions and secondary category task were necessary for the good/bad effect. Specifically, we asked whether this advantage would persist when a different rating task was used that did not evoke the image category. To further lessen the likelihood of participants relying on category information, they received no indication that the images would be drawn from a specific set of categories. Instead of a secondary category judgment, participants were asked to rate the clarity of each image. Such a task will tell us if participants subjectively experience the good exemplars as clearer than the bad.

Method

The design of Experiment 2 was the same as Experiment 1, with two exceptions. Instead of being prompted to rate how well the image exemplified its

category, participants were prompted to provide a rating of how "clearly" they felt they saw the image on the preceding trial by pressing a number between 1 (*not clearly*) and 5 (*very clearly*) immediately after each intact/scrambled response. Moreover, participants were never given a list of the categories used in the experiment; they were simply told that they would be looking at pictures of scenes.

Results and discussion

Participants' overall accuracy on the intact/scrambled distinction was 85%, and they achieved this accuracy with an average image durations of 43 ms. We again observed a significantly greater d′ for good exemplars than bad exemplars (2.51 vs. 2.26), $t(17) = 3.50$, $p = 2.7 \times 10^{-03}$, Cohen's $d = 0.32$, $d_z = 0.83$, which was again driven by a higher hit rate (90% vs. 86%), $t(17) = 5.62$, $p = 3.06 \times 10^{-5}$, with no observed difference in false alarm rate (17% vs. 18%), $t(17) = -1.35$, $p = 0.20$, or response time (950 ms vs. 966 ms), $t(17) = -1.23$, $p = 0.23$, for good versus bad exemplars (see Figure 2). In this experiment, the higher hit rate accompanied by no difference in false alarm rate translated to a significant difference in bias (–0.14 vs. –0.04), $t(17) = -2.75$, $p = 0.01$.

These results indicate that the category rating judgment required in Experiment 1 was not necessary for the good exemplar advantage. Indeed, a mixed-design analysis of variance (ANOVA) with one within-subject factor (good vs. bad category exemplars) and one between-subject factor (Experiment 1 vs. Experiment 2) found a significant effect of good versus bad, $F(1, 34) = 20.88$, $p = 6.18 \times 10^{-5}$, but no main effect of experiment, $F(1, 34) = 0.01$, $p = 0.922$, and no interaction between experiment and exemplar quality, $F(1, 34) = 0.37$, $p = 0.55$, on participants' d′.

Moreover, a good/bad effect was also seen in participants' clarity ratings. Intact good images were rated as more "clear" than intact bad images (3.79 vs. 3.57), $t(17) = 5.74$, $p = 2.4 \times 10^{-5}$, whereas no difference was observed in the clarity ratings of scrambled good and bad exemplars (2.50 vs. 2.48), $t(17) = 0.82$, $p = 0.43$. In other words, not only did good images result in higher sensitivity to the intact versus scrambled distinction, but participants experienced them as being more clear, in keeping with our hypothesis that good exemplars are actually perceived more readily than bad exemplars. Taken together, these results imply that participants tend to see images that are good examples of a basic-level scene category more clearly than bad, regardless of whether they are asked to perform a covert categorization task.

Experiment 3: Intact/scrambled task only

To ensure that the effects found in Experiments 1 and 2 were not due to influences from the secondary rating tasks, in a final experiment we replicated the good exemplar advantage in an experiment with no intervening task. Participants made only intact/scrambled judgments. In addition, we performed a power analysis using the smaller of the effect sizes (Cohen's $d_z = 0.69$) from Experiments 1 and 2 to determine sample size; a sample size of 19 participants will give us 80% power to detect an effect of this size.

Method

The design and procedures for Experiment 3 were identical to that of Experiments 1 and 2, except that no intervening rating task was used. A total of 19 participants were run in this experiment and were paid $8 for their participation.

Results and discussion

Participants' overall accuracy on the intact/scrambled task was 68%, and they achieved this accuracy with an average image durations of 45 ms. We once again observed a significantly greater d′ for good exemplars than bad exemplars (1.34 vs. 1.13), $t(18) = 2.78$, $p = 0.012$, Cohen's $d = 0.28$, $d_z = 0.64$. As in the previous experiments, the sensitivity difference was due to higher hit rates (78% vs. 74%), $t(18) = 4.70$, $p = 1.80 \times 10^{-4}$, and not a difference in false alarm rates (38% vs. 38%), $t(18) = 0.05$, $p = 0.96$ (see Figure 2). The difference in hit but not false alarm rates again resulted in a difference in bias (–0.27 vs. –0.18), $t(18) = -2.47$, $p = 0.02$.

These results provide further evidence that the ratings tasks were not necessary to produce the effect. A mixed-design ANOVA on d′ with one within-subject factor (good vs. bad category exemplars) and one between-subject factor (Experiment 1, 2, or 3) confirmed this conclusion; we observed a significant effect of good versus bad, $F(1, 52) = 28.20$, $p = 2.3 \times 10^{-6}$, and a main effect of experiment, $F(2, 52) = 13.51$, $p = 1.9 \times 10^{-5}$, but importantly no interaction with experiment, $F(2, 52) = 0.21$, $p = 0.81$. Although participants' responses were faster and less accurate in this experiment than in Experiments 1 and 2, we still did not observe a difference in response times (587 ms vs. 591 ms), $t(18) = -0.81$, $p = 0.43$. The faster responses in Experiment 3 suggest that the increased error rates for this experiment compared with the first two are due to a

speed/accuracy tradeoff. It would seem that the presence of a secondary interleaved task not only slowed performance on the primary task, presumably because of task switching, but also allowed participants more time to consider their responses and thus increase their accuracy.

We note that the good/bad effect size ($d_z = 0.64$ to 0.89) was comparable across all three experiments. To place this is some context, these effects are considerably smaller than the effect of good versus bad exemplars on accuracy in an overt scene categorization task ($d_z = 3.86$; Torralbo et al., 2013). Instead, our good/bad effect is comparable to the priming effect of words on a picture detection task ($d_z = 0.61$; Lupyan & Ward, 2013).

Scene similarity

The previous experiments showed that the "representativeness" of an image predicts how well it will be detected in an intact versus scrambled judgment task. We considered two possible explanations of this effect. One, the human visual system may be tuned to "typical" environments so that they can be processed with greater efficiency. Such a mechanism would presumably be tuned over a long timescale, as observers gain more experience with the visual world. On the other hand, it is also possible that the mechanism responsible for this effect operates over shorter timescales; for example, good scenes may be more effectively primed by the targets from preceding trials, resulting in better performance for good than bad exemplars. Such a priming effect may result from something as simple as similarity priming, as we have previously shown that good exemplars are more similar to each other than are bad exemplars (Torralbo et al., 2013).

We considered both of these possibilities by examining the effects of typicality and priming on scene detection accuracy. Although these two factors suggest different mechanisms underlying the good-bad effect reported in Experiments 1–3, the evaluation of both will rely on a measurement of the similarity between pairs of scene images. We estimated similarity using the "spatial envelope" model of scene perception (Oliva & Torralba, 2001). Work on this model has shown that spectral information can be used both to describe spatial properties of scenes, such as "openness" and "naturalness," and also to categorize scenes at the basic level (Greene & Oliva, 2009). Images were first rescaled to 400×300 pixels, and spectral information was extracted by calculating the response to Gabor filters at three spatial frequencies and eight orientations over a fixed window size of 100×75 pixels, preserving local information in each subregion, which has been shown to be necessary for predicting human image similarity judgments (Schwaninger et al., 2006). The filter responses for each window were concatenated to obtain a feature vector for each image, and the Euclidean distance between pairs of vectors provided a measure of similarity. Using this measure, the typicality of each image was estimated by computing its average similarity to the remaining images in its category. Priming effects were tested by examining the similarity between the target image of each trial and its predecessor, irrespective of category. We also considered similarity and typicality values that were derived using the first 28 principle components of the spatial envelope features, which captured 90% of the variance in intact scene images; the sign and magnitude of the results described below were unchanged when restricting the analyses to these 28 components.

Typicality

We estimated the typicality of each image in the experiment by first extracting a feature vector using the spatial envelope model (Oliva & Torralba, 2001). The similarity between that image and each remaining image in its category was then calculated, and the mean value of these distances served as an objective measure of a scene's typicality. Typicality values were z-scored prior to subsequent analyses. Observers' accuracy on intact trials from all three experiments ($n = 55$) was modeled using a hierarchical logistic regression with a fixed effect of typicality and a random intercept for each participant. Results show that typicality is a reliable predictor of accuracy ($\beta = 1.60$, $Z = 9.49$, $p = 2.26 \times 10^{-21}$); participants were better at detecting more typical intact images. The model was then extended to include an additional fixed effect of representativeness ("good" vs. "bad"). Both factors were significant predictors of accuracy, (representativeness: $\beta = 0.25$, $Z = 5.89$, $p = 3.84 \times 10^{-9}$; typicality: $\beta = 1.37$, $Z = 7.88$, $p = 3.18 \times 10^{-15}$). These data suggest that although typicality influenced detection, there may be another separate effect of the good images. To verify this, we used a log-likelihood ratio test to compare these models to determine whether the inclusion of this representativeness factor resulted in a model that better explained the data. The model including both typicality and representativeness was found to provide a significantly better fit to the data, $\chi^2(1) = 34.63$, $p = 3.98 \times 10^{-9}$. In other words, the inclusion of representativeness allows for a better prediction of observers' accuracy, suggesting that typicality alone does not account for the good/bad effect.

Priming

Prototype models of scene perception (Rosch, 1975) predict that pairs of "good" exemplars from the same

category should be more similar to each other than same-category pairs of "bad" images. In keeping with this prediction, we have previously shown this to be true of the images used here (Torralbo et al., 2013). If an intertrial priming effect is influencing participants' perception of scenes, such that a scene is easier to detect when it follows a trial with a similar scene, then the tendency for good image pairs to have high similarity could contribute to the good/bad effect observed in these experiments. That is, the apparent advantage for good scenes in Experiments 1 through 3 may reflect the fact that they have been more effectively primed by recent images. To test this potential explanation for the good/bad effect, the relationship between participants' accuracy and the similarity of successive trials was assessed.

First, using the same similarity analysis described above, the similarity values of same-category image pairs were analyzed to determine whether good pairs were actually more similar than bad pairs. Similarity values were selected from all pairs of images that have the same category and representativeness value. Across all such pairs, good image pairs were more similar than bad image pairs; Welch's two-sample t test revealed that this difference was significant, $t(18{,}247) = 47.54$, $p = 2.2 \times 10^{-16}$. If similarity-based priming is observed, this result suggests that this priming could be related to the effect of representativeness on accuracy, because good images would be more likely to follow a similar scene than bad exemplars.

A hierarchical logistic regression with a fixed effect of similarity and random intercept for each participant was fit to observers' accuracy on intact trials. The first trial from each block was excluded from this analysis, as we did not expect priming effects to endure through the short break participants were allowed to take between blocks. We found that similarity between a given image and the previous trial's image was a reliable predictor of accuracy ($\beta = 0.50$, $Z = 6.78$, $p = 1.2 \times 10^{-11}$). To determine the relationship between similarity and the good/bad effect, representativeness was next included in the model. Both factors were statistically significant predictors of accuracy (good/bad: $\beta = 0.33$, $Z = 7.98$, $p = 1.58 \times 10^{-15}$; similarity: $\beta = 0.52$, $Z = 7.02$, $p = 2.3 \times 10^{-12}$). The addition of the representativeness factor resulted in a model with higher explanatory power than the model fit to similarity alone, as shown by a likelihood ratio test, $\chi^2(1) = 34.63$, $p = 4.0 \times 10^{-9}$. In other words, although priming from similar images results in greater accuracy on the intact versus scrambled image discrimination task, the fact that additional variance in accuracy can be explained by representativeness suggests that priming cannot completely account for the good/bad effect on accuracy.

Finally, to compare the priming and typicality effects, we fit a mixed-effects logistic regression to accuracy on intact trials that modeled representativeness, similarity, and typicality as fixed effects and allowed for a random intercept for participants. This model revealed significant effects of representativeness ($\beta = 0.27$, $Z = 6.28$, $p = 3.5 \times 10^{-10}$), similarity ($\beta = 0.43$, $Z = 5.70$, $p = 1.2 \times 10^{-8}$), and typicality ($\beta = 1.19$, $Z = 6.72$, $p = 1.8 \times 10^{-11}$). This finding suggests that all of the factors considered here exert an influence on scene detection. We also considered the effects of priming, typicality, and representativeness on the clarity ranking observed in Experiment 2. This was evaluated by fitting a hierarchical linear regression on the rankings participants gave to intact images, which had reliable effects of representativeness ($\beta = 0.17$, $t = 4.66$, $p = 1.4 \times 10^{-4}$), similarity ($\beta = 0.06$, $t = 3.41$, $p = 2.7 \times 10^{-3}$), and typicality ($\beta = 0.10$, $t = 5.08$, $p = 5.3 \times 10^{-5}$).

General discussion

Using a two-alternative forced-choice discrimination task, we found that intact photographs of bad exemplars of natural scene categories were more likely than good category exemplars to be mistaken for phase-scrambled images; that is, good exemplars were actually easier to see as coherent images than bad ones. The similar pattern of results observed across all three experiments shows that the good exemplar advantage is not dependent on being explicitly instructed to consider the category of an image (Experiment 1). Importantly, these data suggest not only that detection and categorization co-occur (Grill-Spector & Kanwisher, 2005) but also that categorization actually influences detection; whether the image is a good exemplar or not actually influences how readily participants detect the presence of a coherent image. Such a result is consistent with recent predictive coding and "frame-and-fill" models that posit that hypotheses generated in higher areas help to shape activity in earlier areas (Bullier, 2001; Chen et al., 2007; Hochstein & Ahissar, 2002; Panichello, Cheung, & Bar, 2012; Rao & Ballard, 1999); such hypotheses (or templates) are more likely to be in line with good exemplars and thus result in smaller prediction errors when the image is representative of its category.

These data raise the interesting question of how the good versus bad exemplars are having their effect on perception. Does the good exemplar advantage reflect better templates, built up over the observer's lifetime, or might the good versus bad effect reflect a perceptual advantage accrued within the context of the experiment? We used a measure of scene similarity to examine these two possibilities. We demonstrated that accuracy

in the intact/scrambled detection task was higher for more typical scenes (i.e., those that were most similar to other same-category images). This effect may be driven by neuronal tuning or learned statistical regularities present within visual cortex, which enable the system to more efficiently represent good scenes compared with bad scenes. Previous work has investigated the patterns of neural activation associated with scene categorization, evaluating in several regions of interest both the ability to decode the category from their activity and the similarity of the decoded information to human categorization performance (Walther et al., 2009). Above-chance decoding was observed in a number of areas, including the primary visual cortex and the parahippocampal place area (Epstein & Kanwisher, 1998), with the latter having both the highest decoding accuracies and the best match with observers' behavior in a scene categorization task. More recently, it has been shown that in these same areas, the patterns of activity evoked by good category exemplars is decoded more accurately than that elicited by bad exemplars (Torralbo et al., 2013), implying that good exemplars result in a more robust neural representation of scene category. In the parahippocampal place area, good scenes not only were decoded more accurately but also elicited a lower blood-oxygen-level dependent (BOLD) response than bad exemplars, which is consistent with a more efficient representation.

Although such changes in neuronal tuning may develop over long timescales, we also used scene similarity to investigate a possible role of priming in a scene detection task. Repetition priming is known to enhance perception (Tulving & Schacter, 1990), and consistent with these effects, we showed that intact images were more likely to be detected when they were preceded by similar images. Although we have documented both typicality and priming effects on scene perception, models including these effects failed to explain participants' behavior as well as models that included representativeness; thus, although these factors may contribute to our effect, neither can fully account for it. We note also that using the same intact/scrambled paradigm, we have shown that improbable images (e.g., people in pink rabbit suits pulling suitcases in an airport) are also detected less readily than probable images (e.g., people in typical clothing pulling suitcases in an airport; Greene, Botros, Beck, & Fei-Fei, 2015). Because these images were not drawn with respect to any category, priming between images is highly unlikely and certainly no more likely between probable than improbable images. Indeed, using a support vector machine classifier, we were unable to distinguish between improbable and probable images on the basis of color histograms, scene gist features, edge density, and multiscale Gabor filters. Thus, consistent with the data shown here, it would seem that there is a role for expectedness in detection, in the learned statistical regularity sense (i.e., it need not be a conscious or top-down expectation).

Regardless of the mechanism, we have shown that good exemplars of natural scene categories are "seen" better than bad category exemplars, implying that categorization actually influences detection. Importantly, participants were not asked to categorize or even recognize the scenes and were instead just asked whether the images were intact or not. This finding implies that the simple apprehension of coherent visual information is more strongly influenced by category membership than previously believed. Moreover, because category representativeness is presumably learned, these data suggest that experience and expectation affect simple detection, a prediction made by bidirectional models of vision (Bar et al., 2006; Bullier, 2001; Chen et al., 2007; Hochstein & Ahissar, 2002; Panichello et al., 2012; Rao & Ballard, 1999).

Keywords: scene perception, categorization, detection, similarity

Acknowledgments

This work was funded by National Institutes of Health Grant (R01 EY 019429) Office of Naval Research Multidisciplinary University Research Initiative (N000141410671) to L. F.-F. and D. M. B.

Commercial relationships: none.
Corresponding author: Diane M. Beck.
Email: dmbeck@illinois.edu.
Address: Department of Psychology, University of Illinois, Champaign, IL, USA.

References

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., ... Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences, USA*, *103*, 449–454.

Bowers, J., & Jones, K. (2008). Detecting objects is easier than categorizing them. *Quarterly Journal of Experimental Psychology*, *61*, 552–557.

Bregman, A. S. (1981). Asking the "what for" question in auditory perception. In Kubovy M.and Pomerantz J. R. (Eds.), *Perceptual organization* (pp. 99–118). Hillsdale, NJ: Lawrence Erlbaum.

Bullier, J. (2001). Integrated model of visual processing. *Brain Research Reviews*, *36*, 96–107.

Chen, C.-M., Lakatos, P. S., Shah, A. S., Mehta, A. D., Givre, S. J., Javitt, D. C., & Schroeder, C. E. (2007). Functional anatomy and interaction of fast and slow visual pathways in macaque monkeys. *Cerebral Cortex*, *17*, 1561–1569.

Driver, J., & Baylis, G. C. (1996). Edge-assignment and figure–ground segmentation in short-term visual matching. *Cognitive Psychology*, *31*, 248–306.

Epstein, R., & Kanwisher, N. G. (1998). A cortical representation of the local visual environment. *Nature*, *392*, 598–601.

Greene, M. R., Botros, A. P., Beck, D. M., & Fei-Fei, L. (2015). What you see is what you expect: Rapid scene understanding benefits from prior experience. *Attention, Perception, & Psychophysics*, *77*, 1239–1251.

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

Grill-Spector, K., & Kanwisher, N. G. (2005). Visual recognition: As soon as you know it is there, you know what it is. *Psychological Science*, *16*, 152–161.

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, *36*, 791–804.

Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences, USA*, *99*, 9596–9601.

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 and Performance*, *33*, 1431–1450.

Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. *Proceedings of the National Academy of Sciences, USA*, *110*, 14196–14201.

Mack, M. L., Gauthier, I., Sadr, J., & Palmeri, T. J. (2008). Object detection and basic-level categorization: Sometimes you know it is there before you know what it is. *Psychonomic Bulletin & Review*, *15*, 28–35.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York: Freeman.

Nakayama, K., He, Z. J., & Shimojo, S. (1995). Visual surface representation: A critical link between lower-level and higher-level vision. *Visual Cognition*, *2*, 1–70.

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.

Palmer, S., & Rock, I. (1994a). On the nature and order of organizational processing: A reply to Peterson. *Psychonomic Bulletin & Review*, *1*, 515–519.

Palmer, S., & Rock, I. (1994b). Rethinking perceptual organization: The role of uniform connectedness. *Psychonomic Bulletin & Review*, *1*, 29–55.

Panichello, M. F., Cheung, O. S., & Bar, M. (2012). Predictive feedback and conscious visual experience. *Frontiers in Psychology*, *3*, 620.

Peterson, M. A., & Gibson, B. S. (1994). Must figure-ground organization precede object recognition? An assumption in peril. *Psychological Science*, *5*, 253–259.

Peterson, M. A., Harvey, E. M., & Weidenbacher, H. J. (1991). Shape recognition contributions to figure-ground reversal: Which route counts? *Journal of Experimental Psychology: Human Perception and Performance*, *17*, 1075–1089.

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive field effects. *Nature Neuroscience 2*, 79–87.

Rosch, E. (1975). Universals and cultural specifics in human categorization. In Brislin, R. W. Bochner, S. & Lonner W. J. (Eds.), *Cross-cultural perspectives on learning* (pp. 177–206). Beverly Hills, CA: Sage.

Rosch, E., Simpson, C., & Miller, R. S. (1976). Structural bases of typicality effects. *Journal of Experimental Psychology Human Perception and Performance*, *2*, 491–502.

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

Rubin, E. (1958). Figure and ground. In Beardslee D. C. W. (Ed.), *Readings in perception* (pp. 194–203). New York: Van Nostrand.

Sadr, J., & Sinha, P. (2004). Object recognition and random image structure evolution. *Cognitive Science*, *28*, 259–287.

Schwaninger, A., Vogel, J., Hofer, F., & Schiele, B. (2006). A psychologically plausible model for typicality ranking of natural scenes. *ACM Transactions on Applied Perception*, *3*(4), 333–353.

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

Torralbo, A., Walther, D. B., Chai, B., Caddigan, E., Fei-Fei, L., & Beck, D. M. (2013). Good exemplars

of natural scene categories elicit clearer patterns than bad exemplars but not greater BOLD activity. *PLOS One*, *8*, e58594.

Tulving, E., & Schacter, D. L. (1990). Priming and human memory systems. *Science*, *247*, 301–306.

VanRullen, R., & Thorpe, S. J. (2001a). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artifactual objects. *Perception*, *30*, 655–668.

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

Vogel, J., Schwaninger, A., Wallraven, C., & Bulthoff, H. H. (2007). Categorization of natural scenes: Local versus global information and the role of color. *ACM Transactions on Applied Perception*, *4*(3), Article 19.

Walther, D. B., Caddigan, E., Fei-Fei, L., & Beck, D. M. (2009). Natural scene categories revealed in distributed patterns of activity in the human brain. *Journal of Neuroscience*, *29*, 10573–10581.

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception and Psychophysics*, *33*, 113–120.