The speed and accuracy of material recognition in natural images

Christiane B. Wiebel · Matteo Valsecchi ·
Karl R. Gegenfurtner

Abstract We studied the time course of material categorization in natural images relative to superordinate and basic-level object categorization, using a backward-masking paradigm. We manipulated several low-level features of the images—including luminance, contrast, and color—to assess their potential contributions. The results showed that the speed of material categorization was roughly comparable to the speed of basic-level object categorization, but slower than that of superordinate object categorization. The performance seemed to be crucially mediated by low-level factors, with color leading to a solid increase in performance for material categorization. At longer presentation durations, material categorization was less accurate than both types of object categorization. Taken together, our results show that material categorization can be as fast as basic-level object categorization, but is less accurate.

Keywords Categorization · Natural image · Scene recognition

Material recognition is an important part of our daily life, but so far little work has methodologically investigated this issue (exceptions include Adelson, 2001, and Hiramatsu, Goda, & Komatsu, 2011). The material properties of objects provide cues as to object identity and are necessary for action planning (Buckingham, Cant, & Goodale, 2009). The importance of material recognition and perception in daily life highlights the necessity for fast and accurate processing of object material properties.

C. B. Wiebel (✉) · M. Valsecchi · K. R. Gegenfurtner
Fachbereich 06, Psychologie und Sportwissenschaft, Abteilung
Allgemeine Psychologie, Justus-Liebig-Universität Gießen,
Otto-Behaghel-Strasse 10F,
35394 Giessen, Germany
e-mail: Christiane.B.Wiebel@psychol.uni-giessen.de

Sharan (2009) suggested that the recognition of material categories in real-world pictures was remarkably fast and accurate. She tested material detection in a rapid serial visual presentation (RSVP) paradigm using images obtained from the Flickr.com material image database (Sharan, Rosenholtz, & Adelson, 2009). The results indicated that observers were able to perform this task well, even with stimulus presentation durations as short as 40 ms. She also investigated rapid material detection in a two-alternative forced choice masking experiment. A sequence of five stimuli—one target or distractor image, and four masking images—were presented to the observers, who were asked to respond to a target material category. Material images that did not belong to the target category were presented as distractors, and images used for the backward masking were chosen from the same database as the target images. The backward-mask images were manipulated using the Portilla–Simoncelli texture algorithm (Portilla & Simoncelli, 2000), so as to match their low-level image statistics to the test stimuli. Targets were presented 50% of the time, and the results showed that observers were able to correctly detect material target categories in 83% of the trials with a stimulus presentation time of 40 ms. For longer presentation times (80 or 120 ms), a significant increase in accuracy was reported.

Although observers were quite accurate for presentation times as short as 40 ms in Sharan's (2009) experiments, it is problematic to evaluate absolute performance in these visual tasks as being objectively "slow" or "fast". Such evaluations would need to be relative to a baseline comparison. Since object categorization provides a stable and objective baseline for RSVP task performance, we chose to use it as a comparison for material categorization performance.

In contrast to Sharan (2009), we aimed to directly compare the full time course of material categorization with that of object categorization in a backward-masking paradigm. Other alternatives have been proposed: For example, Adelson,

Sharan, and Rosenholtz (2011) used objects from two distinct categories that were also made of different materials. In this way, the same stimuli could be classified in two different ways. On the one hand, this approach is, in principle, very elegant and avoids the problem of having to somehow equate two different image sets. On the other hand, it is prone to mutual interference between material and object cues, and also faces the problem that the difficulty of both tasks depends on the particular set of stimuli chosen.

Since the classification and categorization of objects in natural scenes has been extensively studied, we consider it a natural benchmark for fast visual categorization. Rapid detection of objects in natural scenes (Bacon-Mace, Mace, Fabre-Thorpe, & Thorpe, 2005; Delorme, Richard, & Fabre-Thorpe, 2000; Fabre-Thorpe, Delorme, Marlot, & Thorpe, 2001; Kirchner & Thorpe, 2006; Thorpe, Fize, & Marlot, 1996; VanRullen & Thorpe, 2001a) and natural-scene understanding and categorization (Greene & Oliva, 2009; Joubert, Rousselet, Fize, & Fabre-Thorpe, 2007; Renninger & Malik, 2004; Rousselet, Joubert, & Fabre-Thorpe, 2005) have been studied under a wide range of conditions. The remarkable speed of processing in these tasks has been confirmed by electrophysiological measurements. The detection of an animal within a natural scene can be completed within 150 ms (Thorpe et al., 1996). Moreover, not only animals, but also artificial objects such as vehicles, can be processed within this time frame (VanRullen & Thorpe, 2001a, 2001b).

We selected four different superordinate object categories for our object categorization task (Exp. 1a): animals, people, buildings, and means of transport. In a second experiment (Exp. 1b), we used four different basic object categories: horses, cows, goats, and rabbits. Although we predicted that the superordinate categories would be an appropriate domain for comparison, we also wanted to test whether basic-level categorization might be a more adequate comparison to the materials.

Since object categorization is known to depend on low-level features—such as contrast or color—we tested categorization performance on both the original version of the images and a version that was normalized with respect to luminance and contrast. We also ran the experiment using grayscale images in order to assess the role of color as a potentially informative cue in these particular categorization tasks.

Experiment 1

Experiment 1a

Method

Observers A group of 18 naive observers took part in the study (14 females, four males). Their mean age was 22.9 years, and ages ranged from 20 to 26 years. All of the observers had normal or corrected-to-normal vision and participated in all four sessions of the study, but in none of the preliminary experiments.

Stimuli Initially, we used the Flickr material image database (Sharan et al., 2009) that had been used by Sharan (2009) and Wolfe and Myers (2010). But since we aimed to test a larger range of stimulus presentation durations and only close-up images, we started to create our own larger image database of material images. Preliminary results that we obtained on the images from the Flickr material database were fully in line with the findings presented here.

A total of 640 images were used in the study, half of them showing objects and half of them showing material surfaces. Object images representing the four categories (animals, people, buildings, and means of transport) were taken from the commercially available COREL database. Each single category included 80 images. We constructed a new data set for the material images. The pictures were taken under varying indoor and outdoor illumination conditions using a Nikon D70 camera (Nikon, Tokyo, Japan) and belonged to the four categories wood, stone, metal, and fabric. Again, each category consisted of 80 images. Up to 37 images that were not 100% correctly validated were used in some of the experiments by mistake. These were discarded from all analyses later on, which did not have an effect on the significance of any of the results. In the case of the object images, the target object was always embedded in a natural context. In the case of the material images, only a material surface without any context or object information was represented. The images are available online at www.allpsych.uni-giessen.de/MID.

A normalized set of all of the images was created in which the mean luminance was equated and the contrast (pixel-wise standard deviation of luminance) was adjusted so as to yield comparable detection rates across categories (see Figs. 1 and 2). The normalization procedure reduced the contrast of the material images by 6.3% on average, whereas the contrast of object images was reduced by 27.3% on average. Example images are shown in Fig. 1.

Experimental setup The experiments were run in a dimly lit room. Stimuli were shown on a Samsung SyncMaster 2230R7 22-in. monitor with a refresh rate of 120 Hz. The observers were seated approximately 100 cm in front of the screen; the images had a spatial resolution of 512×768 pixels, corresponding to $8.23^\circ \times 12.39^\circ$ of visual angle.

Preliminary experiments: Validation experiment A preliminary validation study was conducted in order to select images that could be consistently assigned to their respective categories. The classifications were performed separately for superordinate object and material images. Each image (in the original version) was displayed on the screen for an
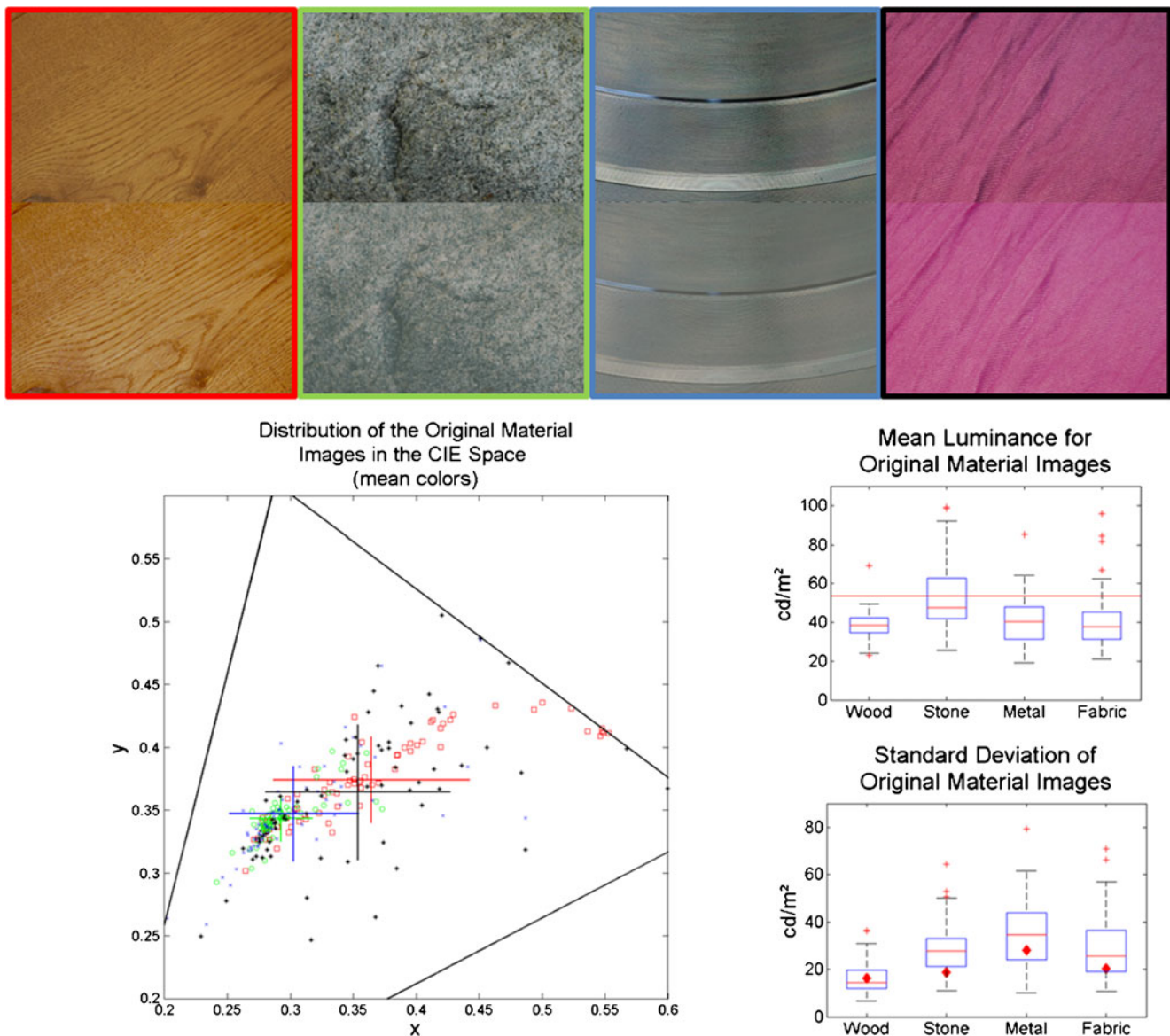
Fig. 1 Example images for the four different material categories, and image statistics for the original material images. The top half of each image shows the image in its original version, and the bottom half shows the image in its normalized version. Color frames (online only) indicate the color codes for the distribution of the original material images in CIE space. The large colored crosses represent the mean and standard deviation of each category's distribution in the CIE space. The

different categories are also indicated by different symbols, as follows: Open squares represent wood; open circles, stone; crosses, metal; asterisks, fabric. The horizontal lines within the mean luminance boxplots (center right panel) indicate the mean luminance after normalization. The diamonds in the standard deviation plot indicate the average contrasts of the single categories after normalization

unlimited viewing time, and four observers were asked to assign each image to one of the given categories. The images were meant to be well discriminable. However, observers were given the option to choose the "I don't know" response instead of picking one of the four categories, if an image looked too ambiguous. A total of 489 material images were tested (115 wood, 125 stone, 128 metal, and 121 fabric) in three different sessions. Of these, 378 images were consistently rated correctly (104 wood, 80 stone, 84 metal, and 110 fabric). That is, approximately 77% of all images were classified correctly.

Preliminary experiments: Contrast detection thresholds The aim of this preliminary experiment was to determine the contrast values that would produce the same detection rates for each of the eight image categories. For this purpose, another four observers were asked to detect the presentation of the images in between two different masks. The contrasts were manipulated over a range of eight fixed conditions, representing 10%–80% of the average contrast in the original image database. The mean luminance of the images was set to the same level as in the normalization procedure, and the images were presented for 8 ms. Masks were generated
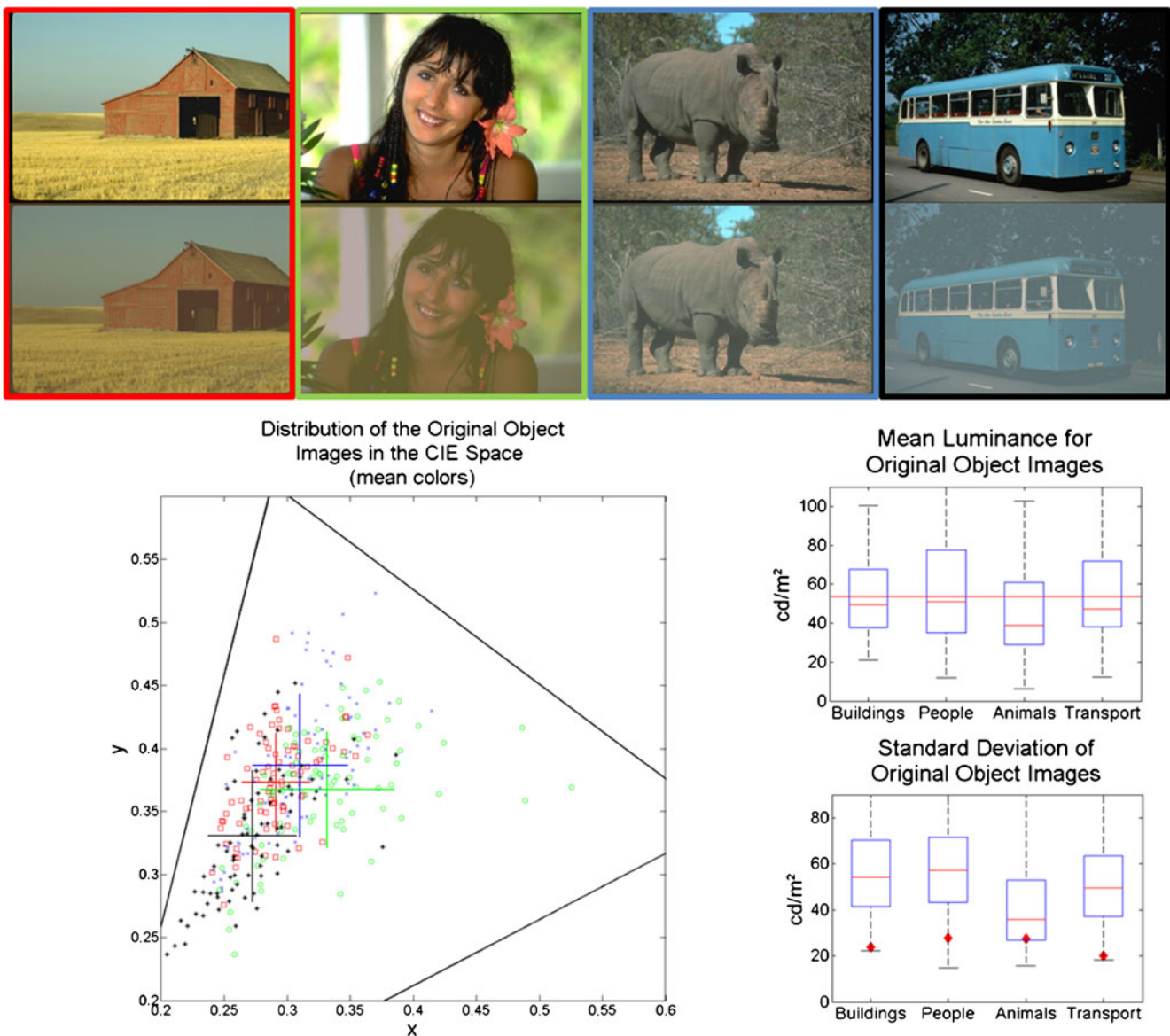
Fig. 2 Example images for the four different object categories, and image statistics for the original object images. The top row shows the images in their original versions, and the second row shows the images in their normalized versions. Color frames (online only) indicate the color codes for the distribution of the original material images in CIE space. The large colored crosses represent the mean and standard deviation of each category's distribution in the CIE space. The different categories are also indicated by different symbols, as follows: Open squares represent buildings; open circles, people; crosses, animals; asterisks, means of transport. The horizontal lines within the mean luminance boxplots (center right panel) indicate the mean luminance after normalization. The diamonds in the standard deviation plot indicate the average contrasts of the single categories after normalization

online by superimposing polygons with colors randomly sampled from the whole image database. The task was repeated in five sessions, in order to obtain 50 trials for each condition. The 50 % threshold values were derived from psychometric functions calculated for each category separately. To create the normalized image sets, the images for each category were set to the contrast value that defined the 50 % detection threshold in this experiment.

Procedure Each of the four combinations of task (material vs. object categorization) and image type (normalized vs.

original) was tested in a separate session. The order of the sessions was balanced between observers, and eight different stimulus presentation times were used, ranging from 8 to 67 ms. Images were centrally displayed in front of a uniformly gray background (see Fig. 3). The images were followed by an online-generated mask (as described above), and the four response categories were subsequently presented in the four corners of the screen. Observers were asked to give their responses by pressing the corresponding button on a standard response pad. A training session consisting of 20 trials with a 1-s image presentation time was
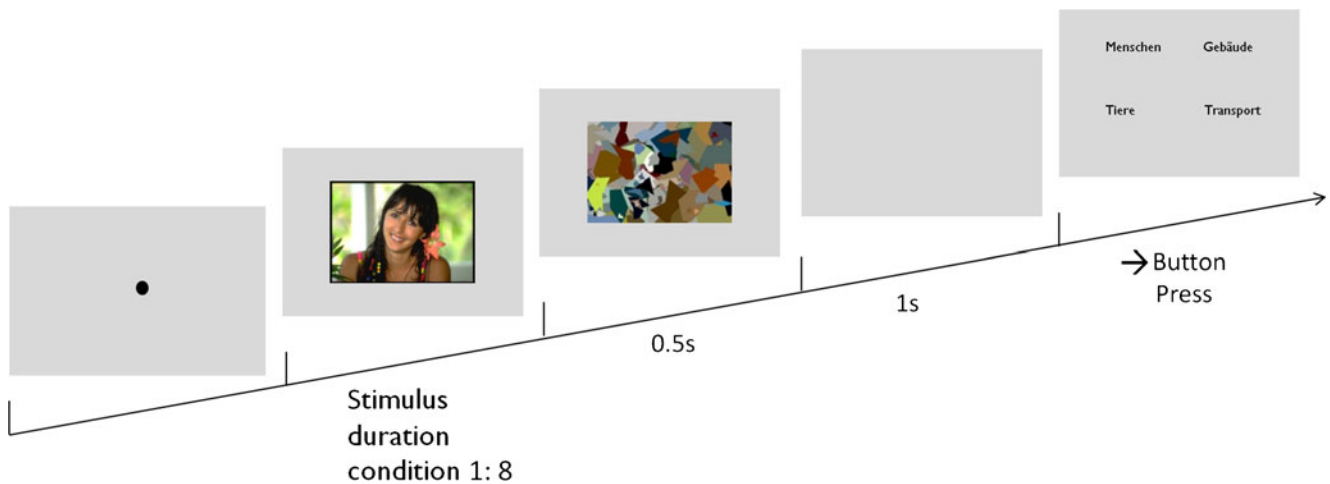
Fig. 3 Schematic procedure of one trial. A fixation dot was shown on a gray background for 1 s before the presentation of the stimulus. Eight different stimulus durations were tested (8, 17, 25, 33, 42, 50, 58, and 67 ms). The stimulus image was followed by a mask for 0.5 s. After one more second, the four response alternatives were presented on the screen, and the observer was asked to make a response

run first. The images included in the training session were not included in the experimental run.

Data analysis Psychometric functions were calculated for each participant and each experimental condition separately using psignifit version 2.5.6 (http://bootstrap-software.org/psignifit/; Wichmann & Hill, 2001). The functions had the following form:

$$\psi(x; \alpha, \beta, \gamma, \lambda) = \gamma + (1 - \gamma - \lambda)\phi(x; \alpha, \beta), \qquad (1)$$

where ψ stands for the cumulative Gaussian function with mean α and standard deviation β. The parameters γ and λ define the lower and upper asymptotes, with γ fixed at the 25% chance level. Two parameters were derived from each fit and used for further analysis. The mean parameter α was taken as an index for the speed of categorization, corresponding to the presentation time at which performance reached a threshold level of 62.5% correct, given that the upper asymptote was at 100%. The upper-asymptote parameter λ was used as an overall index for the accuracy of categorization, corresponding to the performance reached at an infinitely long presentation duration. In practice, it was close to the performance for the longest presentation duration that we used, 68 ms.

Results and discussion

Two questions were examined in this experiment: First, whether material categorization could be as fast as superordinate object categorization, and second, whether the two tasks would yield the same asymptotic level of accuracy.

We found that superordinate object categorization was faster than material categorization. The threshold level of performance indicated by the mean parameter α in the object task was reached after 16 ms, whereas in the material task it took 26 ms. Normalization affected the object images to a greater extent than the material images. The normalization led to a substantial decrease in speed for the object task, as the time needed to obtain a threshold level of performance increased to 25 ms for those stimuli, while the speed of processing did not change much in the material task after normalization (29 ms). Consequently, observers were equally fast in categorizing the normalized object and material images. However, object categorization was more accurate than material categorization, independent of normalization: The maximum accuracy for the 68-ms presentation time was above 98% correct for object images, independent of normalization. In the material task, on the other hand, only 84%-correct performance was reached for the original, and 80% for the normalized, material images. These results are illustrated in Figs. 4 and 5.

Two 2×2 repeated measures analyses of variance (ANOVAs; Task × Image Type) on the mean parameters and on the upper-asymptote parameters were conducted. The analysis of the mean parameters showed a main effect of the Task factor, $F(1, 17) = 15.940$, $p = .001$, $\eta_p^2 = .484$, as well as a significant effect of the Image Type factor, $F(1, 17) = 25.399$, $p < .001$, $\eta_p^2 = .599$, and a significant interaction between the two, $F(1, 17) = 6.945$, $p < .05$, $\eta_p^2 = .290$. The effect was further analyzed by two dependent t tests for each image type separately. Here we found a significant difference only in the case of the original images, $t(17) = 7.103$, $p < .001$, while in the case of the normalized images, the effect did not reach significance.

The analogous 2×2 repeated measures ANOVA (Task × Image Type) for the upper-asymptote parameters revealed a significant main effect of task, $F(1, 17) = 94.099$, $p < .001$,
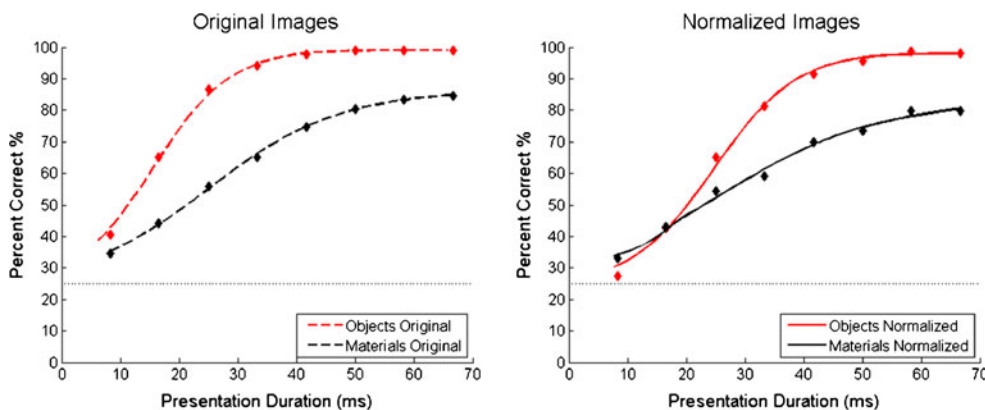
Fig. 4 Superordinate object categorization versus material categorization. The left panel shows the data for the original-image condition, whereas the right panel shows the data for the normalized-image condition. The two curves in each plot represent the averages of the psychometric functions fitted on the data of each single observer, where lighter symbols (red online) show data from the object condition, and black symbols indicate the data from the material image condition. Diamonds represent the respective mean percentage-correct data for each time point measured in the experiment

η_p^2 = .847. No significant effect for image type and no interaction were found.

This raises the question of whether these results could be generalized to other object categories, especially at a level of abstraction that might be more comparable to the material categories at hand. Thus, we conducted a second experiment in which we explored the categorization of four different animal categories.

Experiment 1b

According to Rosch, Mervis, Gray, Johnson, and Boyes-Braem (1976), object categories can be defined at three different levels of abstraction: the basic level (dog), the superordinate level (animal), and the subordinate level (poodle). Rosch et al. found a reaction time advantage for verifying that an object belonged to a basic-level target category, as compared to the superordinate and subordinate levels. Later on, Grill-Spector and Kanwisher (2005) found that object detection and basic object categorization required the same amount of processing time. This contrasted with subordinate categorization, which required longer reaction times in order to produce the same amount of accuracy. The authors showed in a second experiment that correct object detection and basic-level categorization were interdependent.

However, several subsequent studies have challenged this view (for reviews of the whole topic, see Fabre-Thorpe, 2011; Mack & Palmeri, 2011). Mace, Joubert, Nespoulous, and Fabre-Thorpe (2009) reported increased reaction times in a go/no-go task if the target represented a basic-level category rather than a superordinate category. Thus, varying the level of abstraction in the objects should result in a variation of task difficulty.
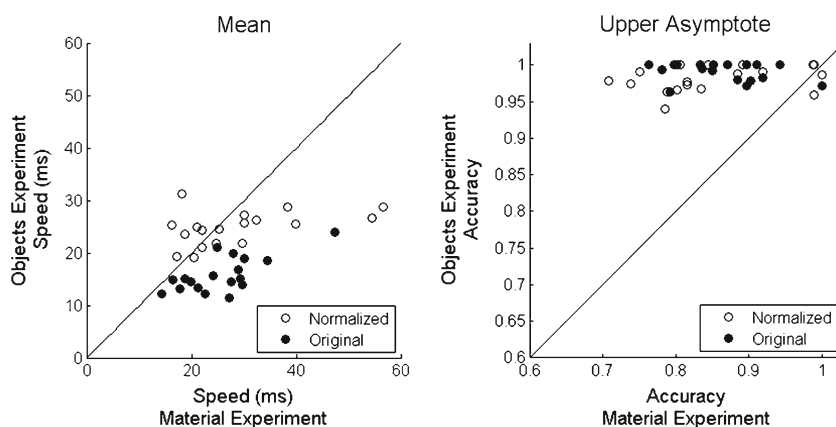

Fig. 5 In both panels, the results for the object task can be found on the y-axis, whereas the results for the material task can be found on the x-axis. The left panel shows the estimated means of the psychometric curves, indicating the speed of categorization. The right panel shows the estimated upper asymptotes, indicating the accuracy of categorization. Each data point represents one participant, with the filled circles representing the original image condition and the open circles the normalized image condition. The black unity lines indicate equal values for both tasks
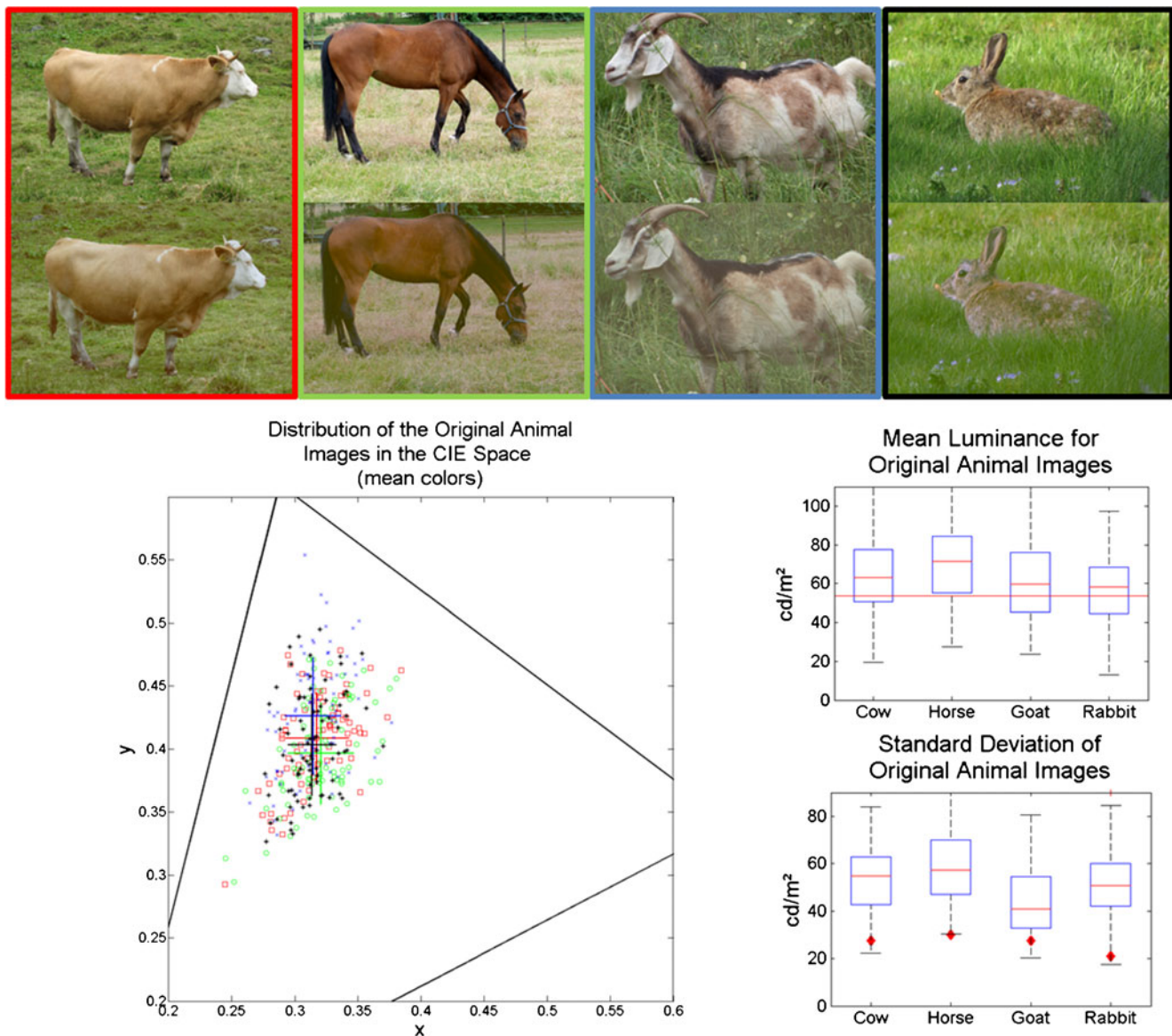
Distribution of the Original Animal
Images in the CIE Space
(mean colors)

Mean Luminance for
Original Animal Images

Standard Deviation of
Original Animal Images

Fig. 6 Example images for the four different animal categories, and image statistics for the original animal images. The top row shows the images in their original versions, and the second row shows the images in their normalized versions. Color frames (online only) indicate the color codes for the distribution of the original material images in CIE space. The large colored crosses represent the mean and standard deviation of each category's distribution in the CIE space. The different categories are also indicated by different symbols, as follows: Open squares represent cows; open circles, horses; crosses, goats; asterisks, rabbits. The horizontal lines within the mean luminance boxplots (center right panel) indicate the mean luminance after normalization. The diamonds in the standard deviation plot indicate the average contrasts of the single categories after normalization

The four object categories (animals, people, buildings, and means of transport) used in Experiment 1a had depicted different superordinate categories, but crucially, the material images had a smaller between-category variability than did the object images used in that experiment, even though they should have constituted quite favorable examples. The object and material images also corresponded to different spatial scales (close-up material images vs. objects embedded in scenes). Adelson et al. (2011) suggested that a difference in spatial scale might play an important role,

indicating that close-up images of materials are more adequate. This is in line with Sharan (2009), who showed that close-up images of materials produced higher accuracies than did those in which the material had to be identified within diverse objects.

Thus, to test whether the advantage of object categorization over material categorization would hold for different levels of object task difficulty, we chose four basic categories (horses, cows, goats, and rabbits) from the superordinate category "animals" for comparison to the material task

in Experiment 1b. According to the results reported above, we would predict that performance would be less accurate in the animal categorization task than in the superordinate object categorization task. Moreover, it was of particular interest whether performance in the much harder animal categorization task would then be comparable to the performance in the material categorization task, or whether we would still find an advantage of our basic-level object categories over the materials.

Method

The setup and the procedure used in this experiment were the same as those in Experiment 1a. To achieve 50% contrast detection levels for normalization, the same contrast detection experiment that we described above was run using the animal and material images. Normalization was undertaken given the same mean luminance value as in Experiment 1a and the contrast levels derived from the contrast detection experiment run for Experiment 1b.

Observers A group of 18 new subjects (all female) participated in the experiment. Their mean age was 22 years, and ages ranged from 18 to 30 years. All of the observers had normal or corrected-to-normal vision. Again, all of the observers took part in four sessions, one for each of the four combinations of task (material vs. animal categorization) and image type (normalized vs. original), but none took part in the preliminary experiment.

Stimuli We collected 320 animal images from the Internet, representing the four categories horses, cows, rabbits, and

goats. The whole animal was represented in the image, and the background was kept relatively similar in all samples. Example images can be found in Fig. 6. We also created a second normalized set of the images, using the same procedure as described before. In Fig. 6, normalized images and image statistics for the different animal categories can be seen.

Results and discussion

The results revealed that observers were equally fast in categorizing material and basic-level object images. This finding emerged independent of normalization. In both tasks, a threshold level of performance was reached after 25 ms. The contrast normalization, which equally affected both image types, led to minor delays of about 7 ms. In terms of accuracy, basic-level object categorization was significantly better than material categorization. This effect was also independent of normalization. Performance in the animal task reached an asymptote of about 90% correct, which was 2% higher than the asymptote in the material task for the original images, and 11% higher than the asymptote for the normalized images. These data are shown in Figs. 7 and 8.

The same kinds of analysis were conducted as in Experiment 1a. A 2×2 repeated measures analysis on the mean parameters revealed no main effect of task and no significant interaction, but a main effect of image type did emerge, $F(1, 17) = 15.413$, $p < .001$, $\eta_p^2 = .476$. Performing the same ANOVA on the upper-asymptote parameters revealed a main effect of task, $F(1, 17) = 13.240$, $p < .05$, $\eta_p^2 = .438$, but no main effect of image type and no interaction.
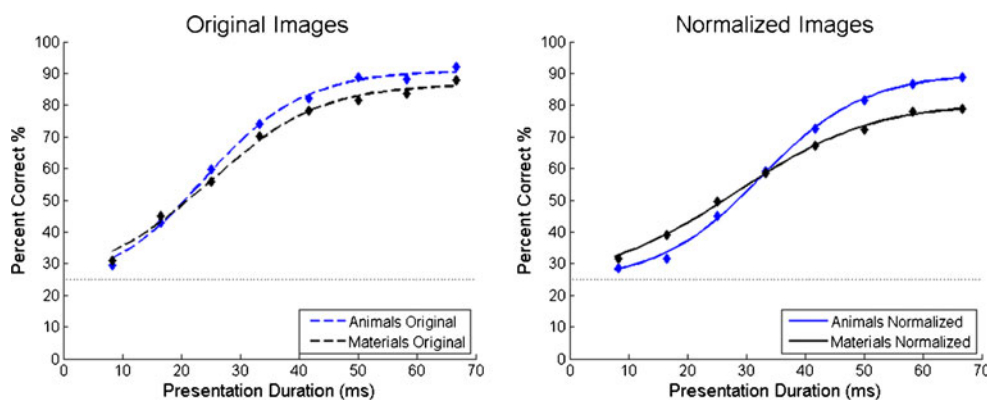

Fig. 7 Basic-level object categorization versus material classification. The left panel shows the original-image data, whereas the right panel shows the normalized-image data. Lines represent the averages of the psychometric functions fitted on the data of each single observer in the different conditions. Lighter symbols (blue online) show data from the animal condition, while black symbols indicate the data from the material image condition. Diamonds represent the respective mean percentage-correct data for each time point measured in the experiment
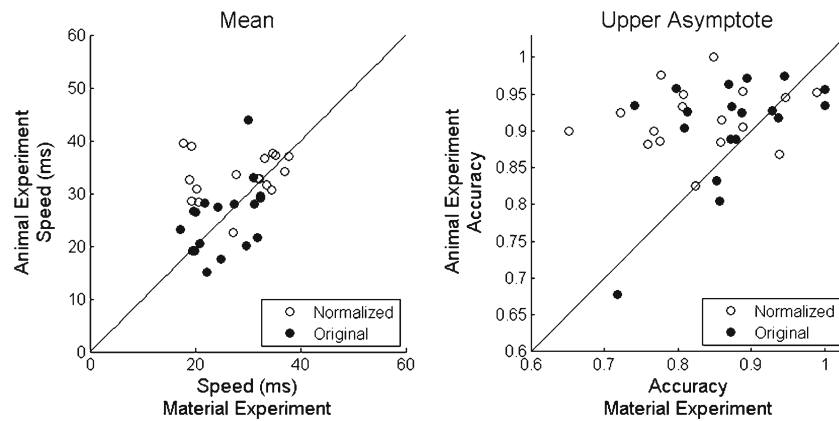
Fig. 8 In both panels, results for the animal task can be found on the *y*-axis, whereas the *x*-axis shows the results for the material task. The left panel shows the estimated means of the psychometric curves, and the right panel shows the estimated upper asymptotes. Each data point represents one participant. Filled circles show data for the original image conditions, and open circles show data for the normalized image conditions. The black lines indicate equal values for both tasks

We also replicated the well-known advantage of superordinate object categorization over basic-level object categorization: Superordinate object categorization was not only significantly faster, but also significantly more accurate than basic-level object categorization (see Fig. 9). On average, the threshold level of performance was reached about 8–9 ms earlier for superordinate than for basic-level objects. The accuracy for basic-level object categorization was also lower by about 9% in the normalized and 7% in the original images than the accuracy for superordinate object categorization.

These results were reflected in two 2×2 mixed design ANOVAs (Task × Image Type) on the mean and upper-asymptote parameters. In both analyses, a main effect of task was found, $F(1, 34) = 45.793$, $p < .001$, $\eta_p^2 = .574$; $F(1, 34) = 19.390$, $p < .001$, $\eta_p^2 = .363$. Besides, a main effect for the factor image type was found for the mean parameters, $F(1, 34) = 85.595$, $p < .001$, $\eta_p^2 = .716$, but not for the upper-asymptote parameters. No interactions occurred.

Experiment 1b showed that the extent to which the ability to categorize material categories differs from the ability to categorize objects crucially depends on the selected comparison level. If the task required a basic-level classification of the object images, object and material categorization were equally fast. Independent of the objects' abstraction level, accuracy was always significantly higher for categorizing objects compared to materials.
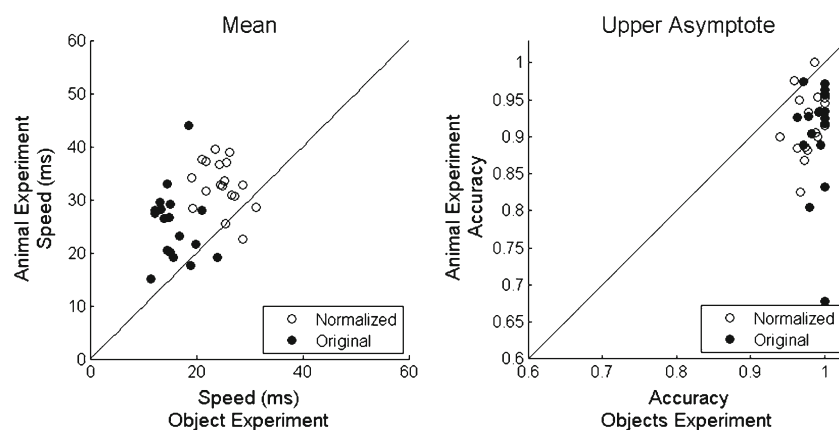

Fig. 9 Both panels show the results for the object task on the *x*-axis and the results for the animal task on the *y*-axis. The left panel shows the estimated means of the psychometric curves, and the right panel shows the estimated upper asymptotes. Filled circles indicate data points for the original image conditions, and open circles indicate data points for the normalized image conditions. Each data point represents one participant. The black lines indicate equal values for both tasks

In agreement with our hypothesis, performance was worse in the basic-level object categorization task than in the superordinate object categorization task. Superordinate categorization was roughly 10 ms faster and 10% more accurate than basic object categorization.

Experiment 2

Since low-level cues seem to play an important role for the categorization task, we chose to examine the influence of color on categorization. In principle, mean chromaticity alone, as indexed by the average CIE xy coordinates of the images (see Figs. 1, 2, and 6), could be potentially useful for classifying the images into the different categories. A linear discriminant analysis, implemented using the classify function of MATLAB (The MathWorks, Natick, MA, USA), performed better than the 25% chance level for classifying object images (47%), material images (38%), and animal images (35%).

The role of color in natural-scene processing and object detection has been assessed by a number of studies before. Whereas color cues were shown to contribute to enhanced recognition memory of objects in natural scenes (Gegenfurtner & Rieger, 2000; Wichmann, Sharpe, & Gegenfurtner, 2002), Delorme et al. (2000) showed that color cues played a rather minor role in a go/no-go object detection task. In a scene-naming and verification task, however, Oliva and Schyns (2000) found a reaction time advantage for color diagnostic scenes.

Therefore, we set out to test the contribution of color to our categorization tasks (material vs. object vs. animal categorization) more directly, by repeating the experiments with grayscale images. Once again, we created normalized versions of the grayscale pictures. According to the literature, no large effects of color information would be expected for the superordinate object and basic-level animal tasks (Delorme et al., 2000). However, color seems to be relatively diagnostic for at least some of the materials, and it could play a role in material categorization.

Method

The experiment was run with the same procedure and under the same conditions as described for Experiments 1a and 1b.

Observers A total of 18 new observers participated in the experiment (14 females, four males). Their ages ranged from 19 to 32 years, with a mean age of 24.3 years. All of the observers had normal or corrected-to-normal vision. The subjects participated in all six sessions of the experiment, one for each of the six combinations of task (material vs. animal vs. object categorization) and image type (normalized vs. original), but in none of the previious experiments.

Stimuli The stimuli were created by converting the images to grayscale. This was done by calculating a weighted sum of the RGB components of each pixel, with the weights chosen by the relative luminance of the three monitor phosphors. In order to normalize the images, we ran the same contrast detection experiment described before using the grayscale images.

Results and discussion

In this experiment, we dealt with the role of color in the categorization of the different image classes. We directly compared the results of the color experiments with the results of the grayscale experiment for every single task. Performance did not reach a level of accuracy in each condition to allow us to successfully fit the data for each single observer. Therefore, we analyzed only the average percentage-correct data across observers. In addition, we calculated comparisons between the color and grayscale conditions at a fixed presentation time of 68 ms.

Color improved performance in all three tasks. Crucially, color led to 7% better performance with the original material images and to 17% better performance with the normalized material images. We found an improvement for color in the superordinate and basic-level object tasks as well, but it depended on image normalization; significance was reached for the normalized images, while only a trend emerged for the original animal images. These results are shown in Fig. 10.

Three 2×2 mixed-design ANOVAs (Color×Image Type) were calculated for each task, and the data were transformed for statistical analyses, using the arcsine of the square root to correct for violations of variance homogeneity in some conditions.
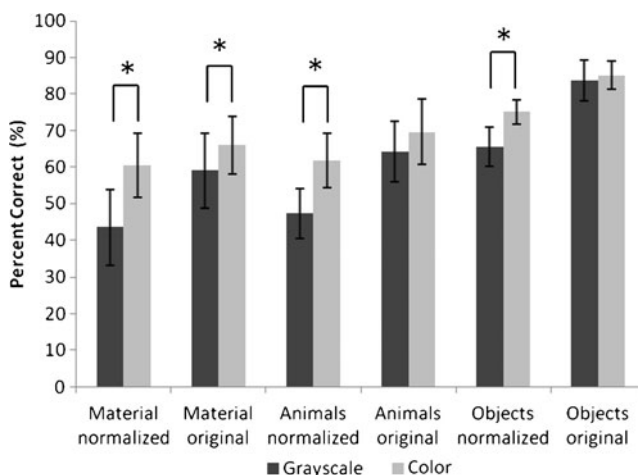

Fig. 10 Mean accuracy results for all color and grayscale conditions are shown for each task and each image type separately. Asterisks indicate significant differences between the indicated conditions

Each ANOVA revealed significant effects of both color [material task, $F(1, 52) = 26.486$, $p < .001$, $\eta_p^2 = .337$; animal task, $F(1, 34) = 15.951$, $p < .001$, $\eta_p^2 = .319$; object task, $F(1, 34) = 21.942$, $p < .001$, $\eta_p^2 = .392$] and image type [material task, $F(1, 52) = 75.366$, $p < .001$, $\eta_p^2 = .592$; animal task, $F(1, 34) = 178.972$, $p < .001$, $\eta_p^2 = .840$; object task, $F(1, 34) = 358.268$, $p < .001$, $\eta_p^2 = .913$], as well as a significant interaction between the two factors [material task, $F(1, 52) = 16.360$, $p < .001$, $\eta_p^2 = .239$; animal task, $F(1, 34) = 21.354$, $p < .001$, $\eta_p^2 = .386$; object task, $F(1, 34) = 35.235$, $p < .001$, $\eta_p^2 = .509$]. Subsequent independent t tests for each image type showed a significant effect of color for the original image condition only in the material task, $t(52) = 2.747$, $p < .05$. For the normalized images, a significant effect of color was found in all tasks [material task, $t(52) = 6.289$, $p < .001$; animal task, $t(34) = 6.00$, $p < .001$; object task, $t(34) = 8.44$, $p < .001$].

At an intermediate presentation times of 68 ms, material categorization accuracy in the grayscale condition was substantially less than in the object task, independently of normalization. While superordinate object categorization was still close to perfect even when color information was discarded (92% in the normalized condition, 98% in the original condition), basic-level animal categorization was less good (74% in the normalized condition, 85% in the original condition), but still much better than material categorization (64% in the normalized condition, 80% in the original condition).

This was confirmed by means of a 3×2 repeated measures ANOVA. We compared accuracies between the three tasks (material vs. animal vs. superordinate objects) and the two image types (normalized vs. original) in the grayscale condition only. Overall, a picture similar to that in the color conditions was revealed. We found significant effects of task, $F(2, 34) = 38.323$, $p < .001$, $\eta_p^2 = .693$, and image type, $F(1, 17) = 62.948$, $p < .001$, $\eta_p^2 = .787$, as well as a significant interaction, $F(2, 34) = 5.529$, $p < .05$, $\eta_p^2 = .245$. Here, we only report subsequent t tests for the comparisons between the material task and the object tasks, as these were of major interest for the scope of this study. Paired-sample t tests showed significant differences between the basic-level animal task and the material task for original images, $t(17) = 2.124$, $p < .05$, as well as for normalized images, $t(17) = 2.563$, $p < .05$. The same pattern of results was found for the comparisons between superordinate objects and materials [original images, $t(17) = 7.500$, $p < .001$; normalized images, $t(17) = 6.912$, $p < .001$].

Overall, these results show that observers actually use the color information that is available for material categorization. We found some evidence that color also aids superordinate- and basic-level object categorization, but these results were less clear-cut, since they depended on the image normalization process. Thus, color seems to contribute to material categorization to a greater extent, but it

does not fully explain the differences between the three tasks at intermediate presentation times.

General discussion

In the present study, we explored the time course of material categorization relative to superordinate- and basic-level object categorization. Specifically, we investigated the speed and accuracy of material categorization as compared to object categorization for very fast presentation times. Our results showed that perceptual material categorization can be processed in the same time range as object categorization, but that it is less accurate. Even under quite favorable conditions (i.e., close-up pictures of visually rather dissimilar material categories), material categorization is still more difficult than object discrimination is under quite unfavorable conditions, when basic-level categorization has to be performed for visually similar objects embedded in rather uninformative scenes.

The speed of categorization

At first glance, our results may seem unsurprising. In agreement with Sharan (2009), material classification is fast, in the sense that an above-chance level of performance is reached even with very short presentation durations of 30 ms or less. Performance in the material categorization task is lower than in a classic object categorization task, for which extremely fast processing was found (Bacon-Mace et al., 2005; Delorme, Richard, & Fabre-Thorpe, 2000, 2010; Fabre-Thorpe et al., 2001; Thorpe et al., 1996; VanRullen & Thorpe, 2001a). This dichotomy could be due to differences in the visual stimuli, as images of objects typically have higher contrast than images of material textures. Therefore, although equating the contrast led to a correlated change in the speed of processing, this was mainly caused by a decline in performance for the object images when their contrast was reduced (see also Mace, Thorpe, & Fabre-Thorpe, 2005). In the real world, object and material texture contrasts vary, so these classifications would normally be processed at different speeds. Since our data were dependent on the interaction between luminance and contrast, it is quite likely that basic image statistics might be used for these classifications.

Two different perceptual mechanisms seemed to be at work in our tasks. At the very shortest presentation durations, performance for all stimuli was very dependent on low-level image factors. Therefore, we think that this regime reflects the activity of very early visual processing. Performance here mainly depended on the (dis)similarity of the stimuli. The later stage, at around 50 ms, was nearly independent of low-level image statistics, and probably reflects higher-level visual

processing. The differences that arose here were presumably due to specialized processing for objects and materials.

Influence of low-level image statistics on categorization

The idea that low-level image features can guide performance in ultra-rapid object or scene recognition had previously been implied. Bar (2003) suggested that low spatial frequencies—representing global information about shape in natural images—can activate high-level representations sufficient for an "initial guess" about the identity of a relevant object. This view was supported, for instance, by the findings of Torralba and Oliva (2003), who showed that second-order image statistics are a reliable cue for discriminating manmade from natural scenes, or even for the detection of objects like animals or vehicles within a scene.

This low-level classifier has not been tested for material images before, but, on the basis of computational studies, a number of important features for both texture analysis and synthesis—for example, Portilla and Simoncelli (2000)—as well as for material image classification—for example, Liu, Sharan, Adelson, and Rosenholtz (2010)—have been identified. Low-level image statistics have also been found to be able to capture some of the subjective qualities of certain material properties, for example gloss (Motoyoshi, Nishida, Sharan, & Adelson, 2007). It is intriguing to ask which of these image statistics may contribute to the fast recognition of material images.

Here, we conducted an experiment assessing the role of color as a low-level cue for categorization in our images. Whereas color has previously been shown to play rather a minor role in object categorization (Delorme et al., 2000), we believed that it would have a greater influence on material categorization, due to its relative diagnosticity for single material classes. As predicted, we found that color was more important for materials than for objects. Regardless of the task, the effect of color on categorization performance proved to be stronger for normalized images, which suggests that once the information provided by luminance is reduced, the information conveyed by chromaticity becomes unmasked and can be accessed by observers. This is in line with previous findings showing that accuracies significantly decreased in a material categorization task (Sharan, 2009), as well as in a go/no-go animal detection task (Mace et al., 2005) when other important low-level factors were degraded in an image.

The accuracy of categorization

While fast categorization is likely due to low-level image statistics, a major difference between object and material image categorization performance is apparent at slightly longer presentation durations. At the longest durations used here, we found marked differences in accuracy between the three tasks. Most importantly, material classification was less accurate than the superordinate- and basic-level categorization tasks. This was somewhat unexpected, because we selected material images that would be perfectly classified with much longer presentations. Most likely, different computations are used at a higher level of visual processing.

Neuronal correlates

Our results fit nicely with the recent fMRI findings of Hiramatsu et al. (2011), who investigated how information about material categories is progressively assembled along the ventral stream. They showed that differences in the relevant image statistics between material categories were associated with differences in neural activation in early visual areas. In contrast, perceptual dissimilarities between materials showed only high correlations with the activation in higher visual areas.

Summary

In summary, we have shown that material categorization can be similarly fast as object categorization, but is less accurate. We replicated the advantage of superordinate object categorization over basic-level object categorization (Mace et al., 2009) and demonstrated that material classification was more similar to basic-level than to superordinate object categorization. The speed of categorization under these ultra-short presentation times was crucially mediated by our normalization procedure. This suggested that performance at this stage was mainly driven by low-level image properties, such as luminance and contrast. In an additional experiment, we examined the effect of color on task performance and found that color was more important for material categorization than for object categorization. The accuracy of categorization was barely affected by normalization, suggesting that different processes were at work. Superordinate and basic-level object categorization were both more accurate than material categorization, independent of normalization. Thus, it can be assumed that computations at higher visual processing stages are involved here, which are more time-consuming in material than in object categorization.

References

Adelson, E. H. (2001). On seeing stuff: The perception of materials by humans and machines. In B. E. Rogowitz & T. N. Pappas (Eds.), *Human vision and electronic imaging VI* (pp. 1–12). Bellingham, WA: International Society for Optical Engineering. doi:10.1117/12.429489

Adelson, E., Sharan, L., & Rosenholtz, R. (2011). Material recognition is fast, but not superfast [Abstract]. *Journal of Vision, 11*(11), 402. doi:10.1167/11.11.402

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

Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience, 15,* 600–609.

Buckingham, G., Cant, J. S., & Goodale, M. A. (2009). Living in a material world: How visual cues to material properties affect the way that we lift objects and perceive their weight. *Journal of Neurophysiology, 102,* 3111–3118.

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

Delorme, A., Richard, G., & Fabre-Thorpe, M. (2010). Key visual features for rapid categorization of animals in natural scenes. *Frontiers in Perception Science, 1,* 21. doi:10.3389/fpsyg.2010.00021

Fabre-Thorpe, M. (2011). The characteristics and limits of rapid visual categorization. *Frontiers in Psychology, 2,* 243.

Fabre-Thorpe, M., Delorme, A., Marlot, C., & Thorpe, S. (2001). A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. *Journal of Cognitive Neuroscience, 13,* 171–180.

Gegenfurtner, K. R., & Rieger, J. (2000). Sensory and cognitive contributions of color to the recognition of natural scenes. *Current Biology, 10,* 805–808.

Greene, M. R., & Oliva, A. (2009). The briefest of glances: The time course of natural scene understanding. *Psychological Science, 20,* 464–472. doi:10.1111/j.1467-9280.2009.02316.x

Grill-Spector, K., & Kanwisher, N. (2005). Visual recognition: As soon as you know it is there, you know what it is. *Psychological Science, 16,* 152–160. doi:10.1111/j.0956-7976.2005.00796.x

Hiramatsu, C., Goda, N., & Komatsu, H. (2011). Transformation from image-based to perceptual representation of materials along the human ventral visual pathway. *NeuroImage, 57,* 482–494.

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

Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research, 46,* 1762–1776.

Liu, C., Sharan, L., Adelson, E. H., & Rosenholtz, R. (2010). *Exploring features in a Bayesian framework for material recognition. In 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 239–246). Los Alamitos, CA: IEEE Press. doi:10.1109/CVPR.2010.5540207

Mace, M. J., Joubert, O. R., Nespoulous, J. L., & Fabre-Thorpe, M. (2009). The time-course of visual categorizations: You spot the animal faster than the bird. *PLoS ONE, 4,* e5927. doi:10.1371/journal.pone.0005927

Mace, M. J., Thorpe, S. J., & Fabre-Thorpe, M. (2005). Rapid categorization of achromatic natural scenes: How robust at very low contrasts? *European Journal of Neuroscience, 21,* 2007–2018.

Mack, M. L., & Palmeri, T. J. (2011). The timing of visual object categorization. *Frontiers in Psychology, 2,* 165.

Motoyoshi, I., Nishida, S., Sharan, L., & Adelson, E. H. (2007). Image statistics and the perception of surface qualities. *Nature, 447,* 206–209.

Oliva, A., & Schyns, P. G. (2000). Diagnostic colors mediate scene recognition. *Cognitive Psychology, 41,* 176–210.

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

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

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. doi:10.1016/0010-0285(76)90013-X

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

Sharan, L. (2009). *The perception of material qualities in real-world images*. Unpublished Ph.D. Thesis. Massachussets Institute of Technology, Cambridge, MA.

Sharan, L., Rosenholtz, R., & Adelson, E. H. (2009). What can you see in a brief glance? *Journal of Vision, 9*(8), 784.

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

Torralba, A., & Oliva, A. (2003). Statistics of natural image categories. *Network: Computation in Neural Systems, 14,* 391–412. doi:10.1088/0954-898X_14_3_302

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.

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics, 63,* 1293–1313.

Wichmann, F. A., Sharpe, L. T., & Gegenfurtner, K. R. (2002). The contributions of color to recognition memory for natural scenes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 509–520.

Wolfe, J. M., & Myers, L. (2010). Fur in the midst of the waters: Visual search for material type is inefficient. *Journal of Vision, 10*(9), 8. doi:10.1167/10.9.8