# Accuracy and speed of material categorization in real-world images

**Lavanya Sharan**
Department of Brain & Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA

**Ruth Rosenholtz**
Department of Brain & Cognitive Sciences, CSAIL, Massachusetts Institute of Technology, Cambridge, MA, USA

**Edward H. Adelson**
Department of Brain & Cognitive Sciences, CSAIL, Massachusetts Institute of Technology, Cambridge, MA, USA

**It is easy to visually distinguish a ceramic knife from one made of steel, a leather jacket from one made of denim, and a plush toy from one made of plastic. Most studies of material appearance have focused on the estimation of specific material properties such as albedo or surface gloss, and as a consequence, almost nothing is known about how we recognize material categories like leather or plastic. We have studied judgments of high-level material categories with a diverse set of real-world photographs, and we have shown (Sharan, 2009) that observers can categorize materials reliably and quickly. Performance on our tasks cannot be explained by simple differences in color, surface shape, or texture. Nor can the results be explained by observers merely performing shape-based object recognition. Rather, we argue that fast and accurate material categorization is a distinct, basic ability of the visual system.**

## Introduction

Our world consists of surfaces and objects, and often just by looking we can tell what they are made of. Consider Figure 1. The objects are easy to identify: a stuffed toy, a cushion, and curtains. It is also clear that these objects are composed of fabric. We may not be able to name all the fabric types in Figure 1, but we know that the surfaces pictured were made from fabric and not plastic or glass. This ability to identify materials is critical for understanding and interacting with our world (Adelson, 2001). By recognizing what a surface is made of, we can predict how hard, rough, heavy, hot, or slippery it will be and act accordingly. We avoid edges of knives and broken glass but not hemlines of garments. We exert more effort to lift a ceramic plate than a plastic plate. We act more quickly when spills occur on absorbent surfaces like paper or fabric. Material categorization, i.e., being able to tell what things are made of, is a significant aspect of human vision, and as far as we are aware, we were the first to systematically study its basic properties (Sharan, 2009).

Most studies of material appearance have focused on the human ability to estimate specific reflectance properties such as albedo, color, and surface gloss. In order to measure the precise relationship between stimulus properties and perceived reflectance, researchers often use synthetic or highly restricted stimuli so that stimulus appearance can be varied easily. By using such controlled stimuli, a number of facts about reflectance perception have been established. It is known that the perceived reflectance of a surface depends not only on its physical reflectance properties (Gilchrist & Jacobsen, 1984; Pellacini, Ferwerda, & Greenberg, 2000; Xiao & Brainard, 2008), but also on its surface geometry (Bloj, Kersten, & Hurlbert, 1999; Boyaci, Maloney, & Hersh, 2003; VanGorp, Laurijssen, & Dutre, 2007; Ho, Landy, & Maloney, 2008), the illumination conditions (Fleming, Dror, & Adelson, 2003; Maloney & Yang, 2003; Gerhard & Maloney, 2010; Olkonnen & Brainard, 2010; Brainard & Maloney, 2011; Motoyoshi & Matoba, 2011), the surrounding surfaces (Gilchrist et al., 1999; Doerschner, Maloney, & Boyaci, 2010; Radonjić, Todorović, &

Citation: Sharan, L., Rosenholtz, R., & Adelson, E. H. (2014). Accuracy and speed of material categorization in real-word images. *Journal of Vision*, *14*(9):12, 1–24, http://www.journalofvision.org/content/14/9/12, doi:10.1167/14.9.12.

doi: 10.1167/14.9.12 Received March 6, 2014; published August 13, 2014 ISSN 1534-7362 © 2014 ARVO

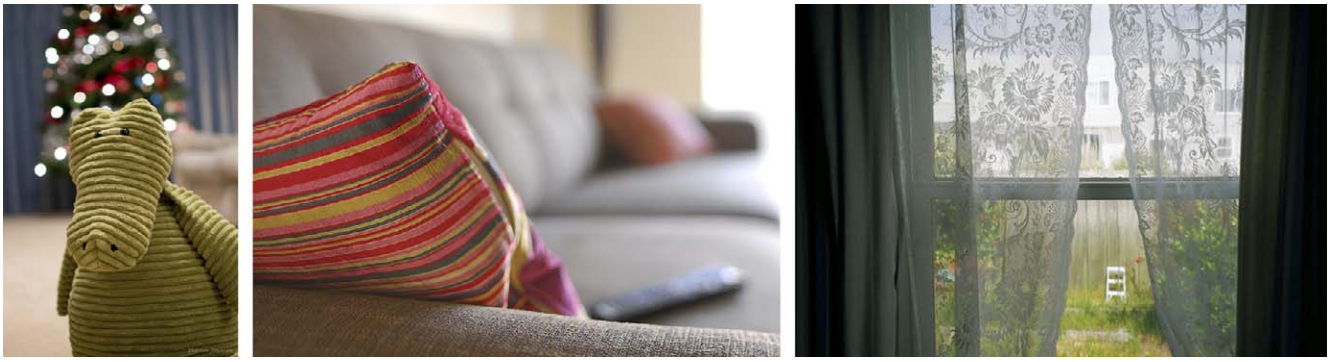


Figure 1. Here are examples of everyday objects that are mainly composed of fabric. We can identify the objects in these images (left to right: stuffed toy, cushion, curtains) just as easily as we can recognize what they are made of. Unlike object and scene categorization, little is known about how we perceive material categories in the real world.

Gilchrist, 2010), the presence of specular highlights (Beck & Prazdny, 1981; Todd, Norman, & Mingolla, 2004; Berzhanskaya, Swaminathan, Beck, & Mingolla, 2005; Kim, Marlow, & Anderson, 2011; Marlow, Kim, & Anderson, 2011) and specular lowlights (Kim, Marlow, & Anderson, 2012), the presence of binocular disparity and surface motion (Hartung & Kersten, 2002; Sakano & Ando, 2010; Wendt, Faul, Ekroll, & Mausfeld, 2010; Doerschner et al., 2011; Kerrigan & Adams, 2013), image-based statistics (Nishida & Shinya, 1998; Motoyoshi, Nishida, Sharan, & Adelson, 2007; Sharan, Li, Motoyoshi, Nishida, & Adelson, 2008), and object identity (Olkkonen, Hansen, & Gegenfurtner, 2008). Recent work has extended this understanding of reflectance perception to include translucent materials (Fleming & Buelthoff, 2005; Motoyoshi, 2010; Fleming, Jakel, & Maloney, 2011; Nagai et al., 2013; Xiao et al., 2014) and real-world surfaces (Obein, Knoblauch, & Viéntot, 2004; Robilotto & Zaidi, 2004, 2006; Ged, Obein, Silvestri, Rohellec, & Viéntot, 2010; Giesel & Gegenfurtner, 2010; Vurro, Ling, & Hurlbert, 2013).

Despite the tremendous progress that has been made on the question of how we estimate *reflectance properties*, little is known about how we recognize *material categories*. How do we know that the surfaces in Figure 1 are made of fabric? What cues do we use to distinguish fabric surfaces from nonfabric surfaces, or for that matter, any given material category from the rest? When we first studied these questions (Sharan, 2009), no one had examined the broad range of visual imagery encountered in real-world materials, of the sort shown in Figure 1. Nothing was known about the accuracy of material categorization, or its speed. Unlike the case of object and scenes (Thorpe, Fize, & Marlot, 1996; Everingham et al., 2005; Grill-Spector & Kanwisher, 2005; Fei-Fei, Fergus, & Perona, 2006; Russell, Torralba, Murphy, & Freeman, 2008; Deng et al., 2009; Greene & Oliva, 2009), there were few suitable datasets (Dana, van Ginneken, Koenderink, & Nayar, 1999; Matusik, Pfister, Brand, & McMillan, 2003) to study material categorization.

In the work described here, we started by collecting a diverse set of real-world photographs in nine common material categories, some of which are shown in Figures 2 and 3. We presented these photographs to human observers in a variety of presentation conditions—unlimited exposures, brief exposures, image-based degradations, etc.—to establish the accuracy and speed of material categorization. We found that observers could categorize high-level material categories reliably and quickly. In addition, we examined the role of surface properties like color and texture, and of object properties like surface shape and object identity. Simple strategies based on color, texture, or surface shape could not account for our results. Nor could our results be explained by observers merely performing shape-based object recognition. We will describe these findings in greater detail in the sections that follow.



Figure 2. An example of a material categorization task. Three of these images contain plastic surfaces while the rest contain nonplastic surfaces. The reader is invited to identify the material category of the foreground surfaces in each image. The correct answers are: (left to right) wood, plastic, plastic, leather, plastic, and glass.

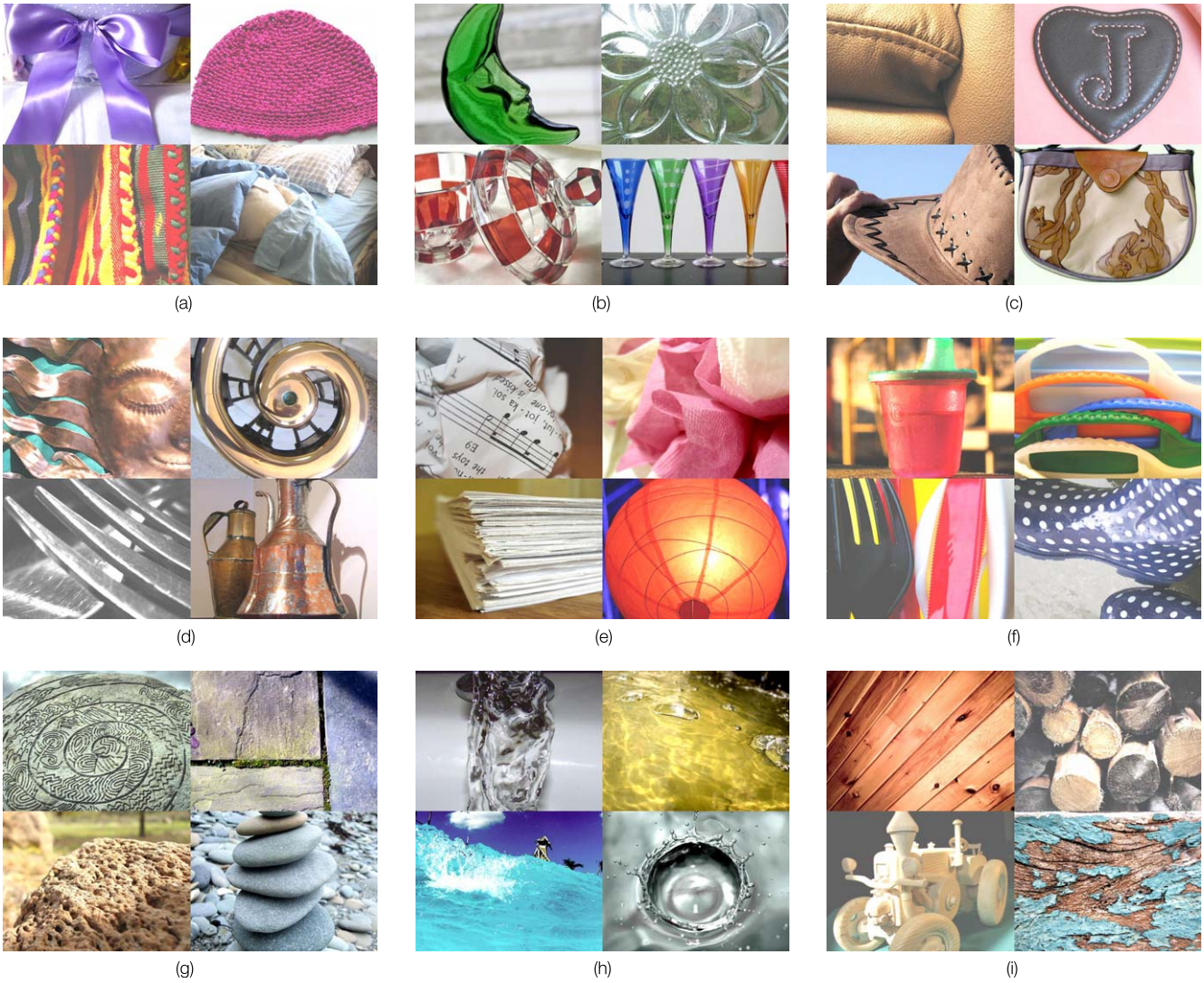



Figure 3. Examples from our image database of material categories: (a) fabric, (b) glass, (c) leather, (d) metal, (e) paper, (f) plastic, (g) stone, (h) water, and (i) wood. We used an intentionally diverse selection of images; each category included a range of illumination conditions, viewpoints, surface geometries, reflectance properties, and backgrounds. This diversity in appearance reduced the chances that simple, low-level information like color could be used to distinguish the categories. In addition, all images in our database were normalized to have the same mean luminance to prevent overall brightness from being a cue to the material category.

Since we first presented these findings (Sharan, Rosenholtz, & Adelson, 2009), others have validated our results and gone on to demonstrate that while material categorization is fast and accurate, it is less accurate than basic-level object categorization (Wiebel, Valsecchi, & Gegenfurtner, 2013) and that visual search for material categories is inefficient (Wolfe & Myers, 2010). It has been shown that correlations exist between material categories and perceived material qualities such as glossiness, transparency, roughness, hardness, coldness, etc. (Hiramatsu, Goda, & Komatsu, 2011; Fleming, Wiebel, & Gegenfurtner, 2013). Newer databases have been developed that capture the appearance of real-world materials beyond high-level category labels (fabric synset of Deng et al., 2009; Bell, Upchurch, Snavely, & Bala, 2013). In our own subsequent work, we built a computer vision system to recognize material categories in real-world images and showed that even the best-performing computer vision systems lag human performance by a large margin (Liu, Sharan, Rosenholtz, & Adelson, 2010; Hu, Bo, & Ren, 2011; Sharan, Liu, Rosenholtz, & Adelson, 2013). We will return to the implications of these recent developments in the Discussion section.

## Are material categorization tasks easy or hard?

The first question we set out to answer was: Are material categorization tasks easy or hard? Should we expect observers to be good at them or be surprised that they can do the tasks at all? Consider the images in Figure 2. Three of these images contain surfaces made from plastic while the rest contain surfaces made from nonplastic materials. Identifying the images that contain plastic surfaces from ones that do not is not obviously straightforward. The surfaces in Figure 2 differ not only in their reflectance properties, but also in their three-dimensional (3-D) shapes, physical scale, object associations, and even the ways in which they are illuminated and imaged. The plastic surfaces of the bag handles, the sippy bottle, and the toy car look quite different from each other. They also bear many similarities in appearance to the nonplastic surfaces. The glasses are multi-colored like the bag handles and translucent like the sippy bottle. The plastic and nonplastic surfaces belong to similar object categories—bags, containers for liquids, and toy vehicles. Given the variations in the appearance within material categories and the similarities across categories, we should expect material categorization to be challenging.

We evaluated performance at material categorization in two ways. In the Material RT experiment, we measured the reaction time (RT) required to make a categorization response. In the Material RapidCat experiment, we measured categorization accuracy under conditions of rapid exposure, in an effort to compare the time course of material category judgments to those for object and scenes. In both experiments, observers were presented photographs from our database of material categories, one at a time, and were asked to report if surfaces belonging to a target material category were present. Performance was averaged over several choices of target material categories and all photographs within a category.

## Stimuli

Color photographs of surfaces belonging to nine material categories were acquired from the photo sharing website, Flickr.com, under various forms of Creative Commons Licensing. The nine categories are shown in Figure 2. Appendix A describes an additional experiment that validates this specific choice of material categories. For each category, we collected 100 images in total, 50 close-ups of surfaces and 50 regular views of objects. Each image contained surfaces belonging to a single material category in the foreground and was selected manually from approximately 50 candidates to ensure a range of illumination conditions, viewpoints, surface geometries, backgrounds, object associations, and material subcategories within each category. All images were cropped to 512 × 384 pixel resolution and normalized to equate mean luminance.

In studies of visual recognition, the accuracy and speed of category judgments are often intimately connected to the choice of stimuli (Johnson & Olshausen, 2003). To ensure that our observers were judging the material category, and not simply low-level image characteristics such as color (brown surfaces are usually wooden) or power spectrum (energy in higher spatial frequencies denotes fabric), we aimed to have our stimuli capture the natural range of material appearances. Consider the fabric selection in Figure 3. The images of the satin ribbon, the crocheted nylon cap, the woven textiles, and the flannel bedding look very different from each other. The four fabric surfaces have different material properties, are of different colors and sizes, and have distinct uses as objects. And yet, it is clear that these surfaces belong to the fabric category and not any of the other eight categories in Figure 3.

## Observers

Thirteen observers with normal or corrected-to-normal vision participated, eight in the Material RT experiment and five in the Material RapidCat experiment. All of them gave informed consent and were compensated monetarily.

## Procedure

### General methods

For all experiments unless noted otherwise, stimuli were displayed centrally on an LCD monitor (1024 × 768 pixels, 75 Hz) against a midgray background using the Psychophysics Toolbox for MATLAB (Brainard, 1997).

### Material RT experiment

Observers were asked to make a material discrimination (e.g., plastic vs. nonplastic), as quickly and as accurately as possible. As illustrated in Figure 4a, each trial started with the fixation symbol (+), and after observers initiated the trial with a key press, a photograph from our database appeared. Observers indicated the presence or absence of a target material category with key presses. Reaction times greater than 1 s, which accounted for 1% of the trials, were discarded. Auditory feedback, in the form of beeps, signaled an incorrect or slow response. To account for

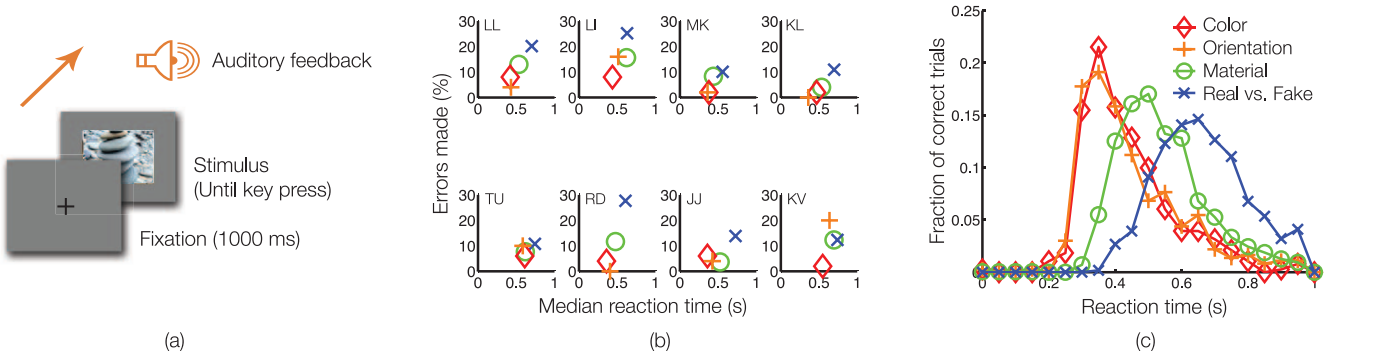



Figure 4. Measuring the accuracy and speed of material categorization with RTs. (a) On each trial, observers indicated the presence or absence of a target category (e.g., stone) with a key press. Auditory feedback was provided, and RTs greater than 1 s were discarded. (b) Errors made by the observers are plotted against their median RTs for the baseline categorization tasks (red and orange; Material RT experiment), material categorization task (green; Material RT experiment), and real versus fake task (blue; Real-Fake RT experiment); there is no evidence of a speed-accuracy trade-off. Chance performance corresponds to 50% error. (c) RT distributions for correct trials, averaged across eight observers, are shown here for each type of task. Compared to baseline tasks, material categorization is slower by approximately 100 ms.

the minimum time taken to make a decision and execute the motor response, observers were asked to complete two easy categorization tasks that served as baselines: discriminating a red versus blue disc and a line tilted at 45° versus −45°. In all three tasks, the target (e.g., red disc or an image containing paper) was present in 50% of the trials.

Each observer completed 400 trials: 50 trials of red versus blue, 50 trials of 45° versus −45°, and 300 trials of material categorization divided equally between three target categories. Trials were blocked by task. Material categorization trials were further blocked by target category. For each target material category, the distracters were selected uniformly from the other material categories in our database. The order of tasks and target material categories were counterbalanced between observers. The stimuli for the material discrimination task subtended 15° × 12°. The stimuli for the baseline tasks were smaller, of size 4° × 4°. Before starting the experiment, observers were shown examples of the judgments they were expected to make and were given a brief practice session.

### *Material RapidCat experiment*

Photographs from our database were presented for 40, 80, or 120 ms and were immediately followed by perceptual masks, as illustrated in Figure 5a. On each trial, the task was to report whether the photograph, of size 15° × 12°, belonged to a target material category (e.g., fabric). Observers pressed keys to indicate target presence. The target category was present in half the trials; distracters were drawn randomly from the other eight material categories. Each observer completed 900 trials, 100 for each of the nine target material categories. Trials were blocked by stimulus presentation time and target material category. Images in each material category were divided as evenly as possible amongst targets and distracters as well as across the three presentation times. The presentation order, the split of database into target and distracter images, and the presentation times associated with each target category were counterbalanced between observers. Like Greene and Oliva (2009), we created our masks using the Portilla-Simoncelli texture synthesis method (Portilla & Simoncelli, 2000). The Portilla-Simoncelli method matches the statistics of the mask images to the statistics of the stimulus images at multiple scales and orientations, which allows for more effective masking than the commonly used pink noise masks.

## Results

In all tasks, chance performance corresponded to 50% accuracy.

### *Material RT experiment*

Figure 4b plots error rates versus median RTs for all tasks. From Figure 4b, it is clear that observers were able to complete our tasks. The accuracy averaged across observers was 90.5% on the material categorization task, 95.2% on the color discrimination task, and 93% on the orientation discrimination task. When we considered only the correct trials, median RT taken over all material categories was 532 ms whereas in the baseline conditions, the median RT was 434 ms for the color task and 426 ms for the orientation task. Figure 4c shows the distribution of RTs averaged across observers for correct trials. There was a significant difference in RT between the conditions shown in

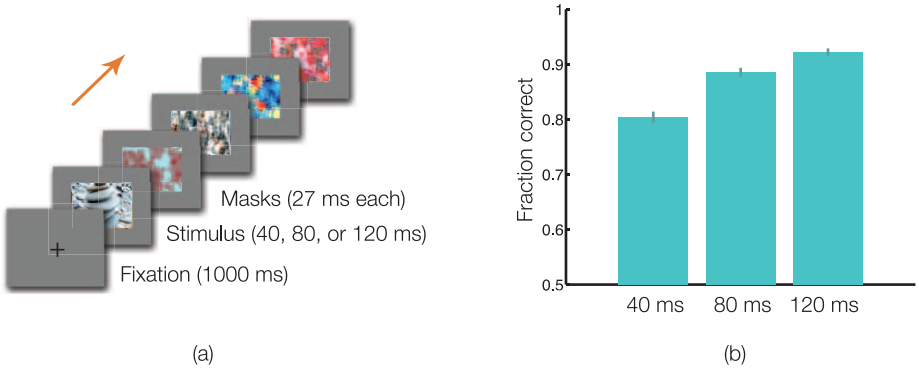



Figure 5. Measuring the accuracy and speed of material categorization with rapid presentations. (a) On each trial, the stimulus was presented for either 40, 80, or 120 ms, and it was followed by four perceptual masks for 27 ms each. Observers indicated the presence or absence of a target category (e.g., stone) with a key press. (b) Accuracy at detecting a given material category, averaged across five observers and nine material categories, is well above chance (0.5) for all three presentation times; this rapid recognition performance is similar to that documented for objects and scenes. Errors represent 1 *SEM*.

Figure 4c, $\chi^2(3) = 21.75$, $p < 0.001$. Post hoc analysis with Wilcoxon signed-rank tests and a Bonferroni correction for alpha revealed significant differences between the material categorization task and the baseline tasks (color: $Z = -2.52$, $p = 0.012$; orientation: $Z = -2.52$, $p = 0.012$). The condition indicated in blue in Figure 4b and c will be described later.

These results demonstrate that observers can make material categorization judgments accurately, even when the images are drawn from a diverse database. However, they take longer to make material category judgments than simpler judgments of color or orientation discrimination. This result is not surprising as the baseline tasks we chose involve a single, simple feature judgment unlike the material discrimination task. Compared to the baseline tasks, the additional time taken to process and respond in the material tasks is approximately 100 ms. Given our initial expectations about the difficulty of material categorization, we find this difference to be small.

### *Material RapidCat experiment*

Our observers achieved 80.2% accuracy even in 40-ms exposures, as shown in Figure 5b. A repeated measures ANOVA determined that accuracy was significantly affected by the image exposure time, $F(2, 8) = 24.92$, $p < 0.001$. Post hoc comparisons using the Tukey HSD test revealed a significant increase in accuracy from 40 ms to 80 ms (88.8%) as well as from 40 ms to 120 ms (92.4%). The accuracy at 120 ms is similar to the accuracy recorded for the material tasks in the Material RT experiment, where image exposures were considerably longer and were determined by observers, $t(11) = -0.86$, $p = 0.411$.

These results establish that material categorization can be accomplished with brief exposures. The performance at 40 ms (80.2%) is similar to that reported for 2-AFC tasks of animal detection, 85.6% at 44 ms (Bacon-Mace, Mace, Fabre-Thorpe, & Thorpe, 2005), and basic-level scene categorization, 75% at 30 ms (Greene & Oliva, 2009), which suggests that the time course of material category judgments is comparable to those for objects and scenes.

It is useful to speculate about observer strategies in the Material RT and Material RapidCat tasks. In deciding whether an image contained a target material, observers may have employed heuristics based on color, reflectance, or shape. For example, wooden surfaces tend to be brown, metal surfaces tend to be shiny, and plastic surfaces tend to be smooth. Alternatively, observers may have employed heuristics based on object knowledge. For example, bottles tend to be made of plastic, handbags of leather, and clothes of fabric. The diversity of our stimuli makes it unlikely that observers can get away with such strategies. However, the possibility that observers were merely recognizing a diagnostic surface property, such as color, or inferring the material category from object knowledge cannot be ruled out without further experiments. To understand the role of surface properties and object knowledge in material categorization, we conducted additional experiments that are described next.

## What is the role of surface properties in material categorization?

The material that a surface is made of determines its reflectance properties and to some extent, its geometric

structure (e.g., wax surfaces tend to be translucent and to have rounded edges). Numerous studies have examined how the visual system estimates reflectance and geometric shape properties of surfaces in the world. These studies have shown that human observers can reliably estimate certain aspects of surface reflectance such as color, albedo, and gloss (for reviews, see Gilchrist, 2006; Adelson, 2008; Brainard, 2009; Shevell & Kingdom, 2010; Anderson, 2011; Maloney, Gerhard, Boyaci, & Doerschner, 2011; Fleming, 2012). Observers can also estimate surface shape up to certain ambiguities (for a review, see Todd, 2004). Based on these findings, one might ask: Are judgments of material categories merely judgments of surface properties such as reflectance and shape?

We addressed this question by measuring the contributions of four surface properties—color, gloss, texture, and shape—to material category judgments. We modified the photographs in our database, as described in Stimuli, to either emphasize or deemphasize these surface properties. We then presented the modified images to observers and compared the material categorization performance on the modified images to that on the original photographs. The differences in performance reveal the role of each surface property in high-level material categorization. Similar analyses have been used to understand the cues underlying rapid animal detection in natural scenes (Nandakumar & Malik, 2009; Velisavljevic & Elder, 2009).

In the Material Degradation I experiment, we deemphasized information about color, texture, and gloss, as shown in Figure 6 (first row). If observers are unable to identify material categories in the presence of these significant degradations, it would imply that material categorization is based on simple feature judgments of color, texture, and gloss. In the Material Degradation II experiment, we emphasized information about shape, texture, and color by removing information about all other aspects of surface appearance, as shown in Figure 6 (second and third rows). If observers are able to identify material categories in such images, it would imply that cues based on shape, texture, and color, in isolation, are sufficient for material category recognition.

While designing the manipulations presented in Figure 6, we had to balance many constraints such as the importance of a particular surface property for material categorization and the ease with which it could be manipulated in a photograph. For a diverse database such as ours and in the absence of any knowledge of imaging conditions, it is difficult to isolate the contribution of a given surface property to surface appearance. Indeed, separating the contributions of surface reflectance, 3-D shape, and illumination in a single image, even in controlled conditions, is an active research topic in computer vision (Tominaga & Tanaka, 2000; Boivin & Gagalowicz, 2001; Dror, Adelson, & Willsky, 2001; Tappen, Freeman, & Adelson, 2005; Grosse, Johnson, Adelson, & Freeman, 2009; Romeiro & Zickler, 2010; Barron & Malik, 2011, 2013).

## Stimuli

Photographs from our database were used to generate stimuli for all conditions of the Material Degradation experiments.

### *Material Degradation I experiment*

There were three viewing conditions: Grayscale, Grayscale Blurred, and Grayscale Negative. Color photographs were converted to grayscale for the Grayscale condition, which tested the necessity of color information. The grayscale photographs were then either low-pass filtered to remove high spatial frequencies (Grayscale Blurred) or contrast reversed (Grayscale Negative). Removing high spatial frequencies, we hoped would impair texture recognition; reversing the contrast, we hoped would impair gloss estimation (Beck & Prazdny, 1981; Todd et al., 2004; Berzhanskaya et al., 2005; Motoyoshi et al., 2007; Sharan et al., 2008; Gerhard & Maloney, 2010; Kim et al., 2011, 2012). The pixel resolution was set to 512 × 384 in all three conditions.

### *Material Degradation II experiment*

There were six viewing conditions, two for each type of surface property, one richer than the other: Shape I and II, Texture I and II, and Color I and II. For each surface property, we aimed to preserve information about that property while suppressing information about the other two (with exceptions for Texture I and II because texture, by definition, includes color). The pixel resolution varied from 384 × 384 (Texture I, Texture II, Color II) to 512 × 384 (Shape I, Shape II) to 750 × 150 (Color I). We will now describe each condition in greater detail.

### *Shape*

We preserved two aspects of surface shape in our stimuli: silhouette and shading. In the Shape I condition, we presented the binary silhouette associated with each image. For certain images, these silhouettes were quite informative because they conveyed the object identity (e.g., a wine glass), and thereby, the material identity (e.g., glass). For other images, such as those shown in Figure 6 (Shape I), they were much less

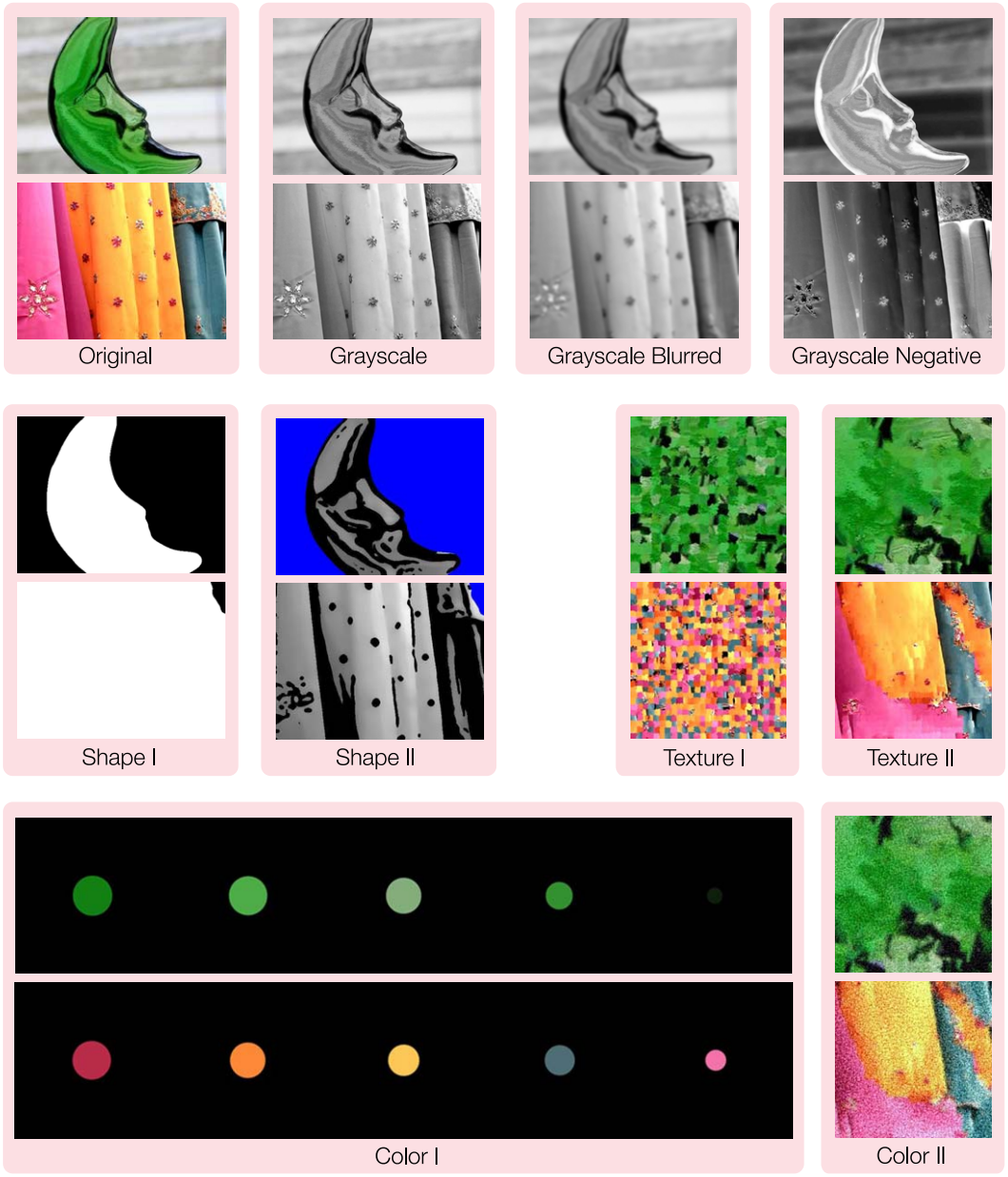



Figure 6. Stimuli used in Material Degradation experiments. In Degradation I experiment (Grayscale, Grayscale Blurred, and Grayscale Negative), we manipulated each photograph in our database to degrade much of the information from a particular surface property, while preserving the information from other properties. In Degradation II experiment (Shape I and II, Texture I and II, Color I and II), we did the reverse. Shown here are two examples from our database, a glass decoration and embroidered garments, along with the three manipulations of Degradation I and the six manipulations of Degradation II experiments. The Grayscale manipulation, for instance, tests the necessity of color information. Grayscale Blurred degrades texture information, and Grayscale Negative makes it hard to see reflectance cues like specular highlights and shadows. The Texture I and II manipulations preserve local spatial-frequency content, including color variations, but lose overall surface shape. The Shape I and II manipulations preserve either the global silhouette of the object or a line drawing-like sketch of the surface shape. The Color I and II manipulations preserve aspects of the distribution of colors within the material, but lose all shape information, and all or much of the texture.

informative. To remedy this, we presented graded, grayscale shading information in addition to silhouettes in the Shape II condition. The richer Shape II stimuli provided more cues to the surface shape without revealing other aspects of appearance such as color and texture.

Stimuli for the Shape I condition were created manually using the Quick Selection tool in Adobe

Photoshop CS3 (Adobe Systems, Inc., San Jose, CA). For each image, regions containing the material of interest were selected and set to white. The remaining regions, which contained the background, were set to black. Silhouettes were generated in this manner for all images in our database.

Stimuli for the Shape II condition were derived as follows: (a) each color image was converted to grayscale and filtered using a median filter, of support $20 \times 20$ pixels, to remove high frequency texture information; (b) the Stamp Filter in Adobe Photoshop CS3 was used to mark strong edges in each original image; and (c) the results of (a) and (b) were combined with the silhouettes from Shape I to create images such as the ones shown in Figure 6 (Shape II). Note that steps (a) and (b) destroy the original appearance of surface edges, including occluding contours; in doing so, our intention was to suppress color and texture cues in our Shape stimuli.

### Texture

We use the term texture to describe two-dimensional (2-D) or wallpaper textures as well as 3-D textures (Dana et al., 1999; Pont & Koenderink, 2005). Both 2-D and 3-D textures are a significant aspect of surface appearance. Unfortunately, for most images in our database, it is difficult to manipulate texture appearance, either 2-D or 3-D, while discounting the effects of illumination, viewpoint, and surface shape. Our solution has been to present texture information locally, in the form of image patches derived from the original photographs. Presenting these local patches in a globally scrambled manner as shown in Figure 6 (Texture I and II), we hoped would interfere with other cues based on surface gloss and surface shape.

Stimuli for the Texture conditions were created using a patch-based texture synthesis method from the field of computer graphics (Efros & Freeman, 2001). For each image, regions containing the material, identified using the silhouettes of Shape I, were divided into either $16 \times 16$ (Texture I) or $32 \times 32$ (Texture II) pixel patches. These patches were then scrambled and recombined in a way that minimized differences in pixel intensity values across seams. The resulting quilted image contained nearly the same information, at the pixel level, as the original surface. However, the scrambling step removed much of the large-scale geometric structure within the material. In Texture I, the scrambling was completely randomized. In Texture II, the scrambling was partially randomized; the first patch was selected at random, and then each subsequent patch was selected so as to best fit its existing neighbors. These differences between Texture I and Texture II stimuli are demonstrated in Figure 6. Texture II stimuli look less fragmented than Texture I, and therefore, they allow observers to glean texture information over larger regions.

It is worth noting that patch-based synthesis methods work best for 2-D, wallpaper-like textures (e.g., head-on view of pattern paper). For more complex textures (e.g., garments close-up in Figure 6), synthesis results do not preserve the full textural appearance. Furthermore, patch-based synthesis methods hallucinate extended edges that do not exist in the original images; this effect is more obvious in the fully randomized Texture I stimuli than in the partially randomized Texture II stimuli.

### Color

We wanted to convey two aspects of surface color appearance: first, the amount of each color (e.g., how much pink), and second, the spatial relationships for each color (e.g., pink occurs next to yellow). We accomplished this by taking inspiration from work on summarization in information visualization (Hearst, 1995; Color I) and by reusing our existing stimuli (Color II). In the Color I condition, we created abstract visualizations that depicted the dominant colors in an image. In the Color II condition, we added noise to our Texture II stimuli to suppress texture information. The resulting stimuli allowed observers to learn not only the dominant colors in the original image, as in the Color I condition, but also the spatial distribution of colors.

Stimuli for the Color I condition were created as follows. For each image, regions containing the material were identified using the silhouettes from Shape I. Next, the R, G, and B values for pixels in these regions were clustered using the k-means algorithm in MATLAB to identify between three to five dominant colors. Finally, a visualization was created where each dominant color was represented by a circle of that color, with radius proportional to the size of its cluster in RGB space.

Stimuli for the Color II condition were created in a manner identical to that for Texture II, except that pink noise was added as a final step.

## Observers

Fourteen observers with normal or corrected-to-normal vision participated in the Material Degradation experiments: five in Degradation I, five in Degradation II, and four in a task that measured baseline performance on the original, nondegraded images. All observers gave informed consent and were compensated monetarily.

## Procedure

In all conditions, observers were asked to identify the material category for a given image. They were provided the list of nine categories in our database, and they indicated their selections with key presses. In each condition, observers were informed about the stimulus generation process, and they were encouraged to reason about what they saw. For example, a lot of brown and orange in the Color or Texture conditions could denote wood, whereas a lot of blue could denote water.

### *Baseline*

Observers were asked to categorize the original photographs in our database. The entire database was divided into three nonoverlapping sets of 300 photographs each. Two observers categorized one set each and the rest categorized two sets each. All photographs were displayed until observers made a response, allowing them as much as time as they needed to examine the photographs. Presentation order was counterbalanced between observers.

### *Material Degradation I experiment*

Observers were shown the Grayscale, Grayscale-Blurred, or Grayscale-Negative images, such as the ones in the top row of Figure 6, and they were asked to perform 9-AFC material categorization. All images were presented for 1000 ms, after which observers indicated their responses with a key press. We used the same split of the database into three nonoverlapping sets as in the baseline task. Two observers categorized two sets each in the Grayscale condition. Three different observers categorized one set each in the Grayscale-Blurred condition. The same three observers categorized images in the Grayscale-Negative condition; two of these observers completed one set each while the third observer completed two sets. The order of viewing conditions and images associated with each condition were counterbalanced between observers.

### *Material Degradation II experiment*

Observers were shown images from all six conditions shown in the lower rows of Figure 6, and they were asked to perform 9-AFC material categorization. On each trial, observers viewed the stimulus image for as long as needed, before indicating with a key press a nine-way response. As the manipulations in this experiment were more severe than the ones in Degradation I, exposure durations were not time limited. The original database was divided into six nonoverlapping sets of 150 photographs each. Each observer completed six sets in total, one set per viewing condition. The order of viewing conditions and the images associated with each condition were counterbalanced between observers.

## Results

In all conditions, chance performance corresponded to 11% accuracy.

### *Baseline*

The accuracy at categorizing the original photographs with unlimited viewing time was 91%. The performance in this condition is the highest that we can expect observers to achieve, and it serves as a baseline for comparison.

### *Material Degradation I experiment*

As shown in Figure 7a, observer performance was robust to the loss of color (Grayscale: 86.6%) and to additional losses of high spatial frequencies (Grayscale Blurred: 75.5%) and contrast polarity (Grayscale Negative: 73.8%). On comparing the accuracy in each condition to baseline performance, we found that removing color did not significantly affect accuracy, Grayscale: $t(4) = 1.007$, $p = 0.371$. However, removing high spatial frequencies or inverting contrast polarity in the grayscale photographs led to a significant decrease in accuracy, Grayscale Blurred: $t(5) = 3.688$, $p = 0.014$; Grayscale Negative: $t(5) = 4.268$, $p = 0.008$; Bonferroni correction yields 0.017 significance value. These results indicate that the simple cues we manipulated have only a minor influence on material categorization. There is sufficient information in the blurred and negative photographs for observers to make categorization judgments successfully.

### *Material Degradation II experiment*

As shown in Figure 7b, observers performed poorly in all conditions (e.g., 32.9% for Shape II). A repeated measures ANOVA determined that there was a significant effect of the viewing condition on accuracy, $F(5) = 9.772$, $p < 0.001$. Post hoc tests using the Bonferroni correction revealed a significant increase in accuracy from 23.3% in the Texture I condition to 31.7% in the Texture II condition ($p = 0.018$). While there is an overall influence of the viewing condition, the main result in Figure 7b is that, unlike rapid scene and object recognition, fast discrimination of a single, simple feature (e.g., color, Oliva & Schyns, 2000; spatial frequency content, Schyns & Oliva, 1994; Johnson & Olshausen, 2003; texture and outline shape, Velisavl-

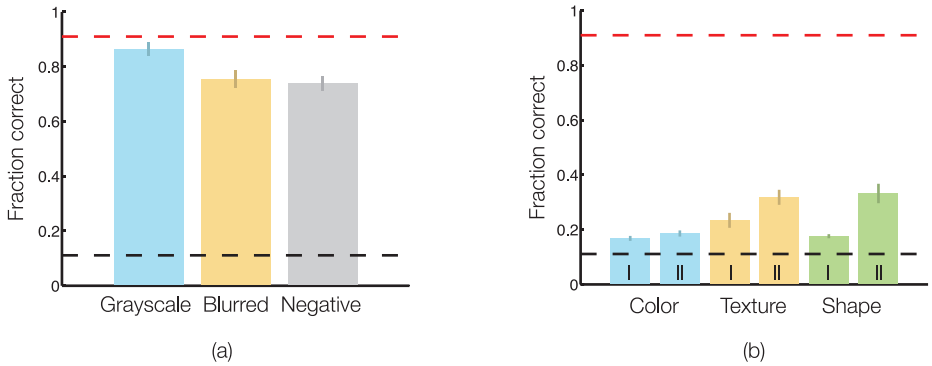



Figure 7. Simple features, by themselves, cannot explain performance. These graphs show the accuracy at material categorization for all conditions in Material Degradation Experiments (a) I and (b) II, averaged across observers and material categories. In the Degradation I experiment, removing color information did not affect performance, while removing high spatial frequencies and inverting the contrast polarity did, but only to a small extent. Observers were still able to identify the material categories in the degraded images shown in the top row of Figure 6. In the Degradation II experiment, when observers were presented mainly one type of information, either shape, texture, or color, they performed poorly. For comparison, accuracy on the original photographs in the baseline condition (0.91) is indicated in red and chance performance (0.11) is indicated in black. Error bars represent 1 *SEM* across three observers in the Grayscale and Grayscale Blurred conditions, two observers in the Grayscale Negative condition, and five observers for all conditions in (b).

jevic & Elder, 2009) that we have tested for does not explain the quality or speed of material category judgments. This is not to say that observers do not use cues based on color, texture, and shape for material categorization, but that observers at minimum require multiple cues.

A closer look at Shape I and Shape II results revealed that there was no effect of the type of view (close-up vs. regular) on categorization accuracy, Shape I, $t(4) = 0.716$, $p = 0.513$; Shape II, $t(4) = 0.211$, $p = 0.843$. Even when objects were clearly visible in regular views (e.g., moon-shaped ornament in Figure 6, Shape I & II), object-based cues did not seem to assist the material category judgment. This result suggests one of two possibilities: (a) object knowledge is not useful for material categorization; or (b) object knowledge is useful for material categorization but observers were unable to utilize it in the Shape conditions. To account for the second possibility, we conducted another experiment with a new set of stimuli.

## What is the role of object knowledge in material categorization?

The material(s) that an object is made of are usually not arbitrary. Keys are made of metallic alloys, candles of wax, and tires of rubber. It is reasonable to believe that we form associations between objects and the materials they are made of. Given the object identity (e.g., book), we can easily infer the material identity (e.g., paper) on the basis of such learned associations. Is it possible then that material category judgments are simply derived from object knowledge? Perhaps, the speeds we have measured for material categorization are merely a consequence of fast object recognition (Biederman, Rabinowitz, Glass, & Stacy, 1974; Potter, 1975, 1976; Intraub, 1981; Thorpe et al., 1996) followed by inferences based on object-material relationships?

There are a few problems with this line of reasoning. First, for most man-made objects, the object identity does not uniquely determine the material identity. For example, chairs can be made of wood, metal, or plastic. Second, for most natural objects, the object identity is confounded with the material identity. For example, bananas are made of "banana-stuff," and nearly all objects made of "banana-stuff" are bananas. Third, there is evidence from patients with visual form agnosia that the ability to identify materials can be independent of the ability to identify objects (Humphrey, Goodale, Jakobson, & Servos, 1994). Finally, half the stimuli used in our experiments were close-up shots, where the object identity was difficult to discern, making it unlikely that object knowledge could be used to identify the material category.

In general, object identities are based on 3-D shape properties as well as material properties. Although shape properties usually dominate the definition of an object, material properties are an important aspect of object identity. For example, a marble toy, a ball bearing element, and an olive that have roughly the same shape and size are easy to distinguish on the basis of their material appearance. It is conceivable that instead of material category judgments being derived from object knowledge, perhaps, object category



Figure 8. Stimuli used in Real versus Fake experiments. Here are some examples from our image database of real and fake objects. (Top row, from left to right) A knit wool cupcake, flowers made of fabric, and plastic fruits. (Bottom row, from left to right) Genuine examples of a cupcake, a flower, and fruits. We attempted to balance the real and fake categories for content by including similar variations in shape, color, object type, and backgrounds. The fake objects in our database were composed of materials such as fabric, plastic, glass, clay, paper, etc., and they ranged from easily identifiable fakes, such as the knit cupcake shown here, to harder cases that potentially required scrutiny. By presenting images from our database to observers and asking them to make real versus fake discriminations, it was possible to dissociate shape-based object recognition from material recognition.

judgments are derived from material knowledge. To understand the relationship between object categorization and material categorization better, we conducted an experiment where the object identity could be dissociated from material properties.

There is a vast industry dedicated to creating fake objects like fake leather garments, fake flowers, or fake food for restaurant displays. These fake items and their genuine versions are useful for our purposes as they differ mainly in their material composition, but not their shape-based object category. Consider the examples shown in Figure 8. Recognizing the objects in these images as fruits or flowers is not sufficient. One has to assess whether the material appearance of these objects is consistent with standard expectations for that object category. The real versus fake discrimination can be viewed as a material category judgment, even if it is a subtler judgment than, say, plastic versus non-plastic. In the experiments that follow, we presented photographs of real and fake objects to observers and asked them to identify the object category and the real versus fake category. By comparing observers' performance on these two tasks, we can determine whether material categorization is simply derived from shape-based object recognition.

In the Real-Fake RapidCat experiment, stimuli were presented for different exposure durations, and the time course for the object categorization was compared to that for the real versus fake discrimination. In Real-Fake RT experiment, RTs were measured for the real versus fake discrimination, and the results thus obtained were compared to those of the Material RT experiment.

## Stimuli

We collected real and fake examples for three familiar object categories: desserts, flowers, and fruits. As shown in Figure 8, the fake examples were made from materials like fabric, plastic, clay, etc. We collected 300 color photographs from the photo sharing website, Flickr.com, under various forms of Creative Commons licensing. There were 100 examples in each object category—50 real and 50 fake. We attempted to balance lighting, background, and color cues as well as the content of these real and fake images. For example, fake flowers, like real flowers, could appear next to real leaves outdoors, whereas fake fruits, like real fruits, could appear in fruit baskets indoors. All photographs were cropped down to 1024

× 768 pixel resolution to ensure uniformity, and they were normalized to equate mean luminance.

As the real versus fake discrimination can be subtle, we screened this set of 300 photographs further to eliminate difficult cases. We asked four naive observers to rate all photographs on a 5-point scale, from "definitely fake" to "definitely real," in unlimited presentation time conditions. Photographs that received unambiguous ratings of realness and fakeness (125 in total, 67 real and 58 fake, divided roughly evenly amongst object categories) were then used for the Real-Fake RapidCat and Real-Fake RT experiments. The four observers whose ratings were used to screen the original set of photographs did not participate in these experiments.

## Observers

Fifteen observers with normal or corrected-to-normal vision participated in the Real-Fake experiments: seven in Real-Fake RapidCat and eight in Real-Fake RT. The observers who participated in the Real-Fake RT experiment were the same as the ones who participated in Material RT experiment. All observers gave informed consent, and they were compensated monetarily.

## Procedure

### *Real-Fake RapidCat experiment*

Photographs of real and fake objects were presented in a rapid presentation paradigm similar to the Material RapidCat experiment. Each photograph was presented for a brief duration (40, 320, or 2560 ms) and was followed by a colored pink noise mask for 1000 ms. Observers were asked to identify the object category (dessert, fruit, or flower) and to judge the authenticity (real or fake) of the object(s) in the photograph. To accommodate the harder material discrimination task, presentation times were longer than in the Material RapidCat experiment. The set of 125 unambigously real and fake photographs was split into three nonoverlapping subsets. Each block consisted of a six-way categorization (dessert/fruit/flower × real/fake) of one of these subsets for one setting of stimulus presentation time. All observers viewed each image only once and completed three blocks each. The order of presentation times and images associated with each presentation time were counterbalanced between observers. All stimuli were presented centrally against a midgray background and subtended 26° × 19.5°. For this experiment, unlike the rest, stimuli were displayed on a different LCD monitor (1280 × 1024 pixels, 75 Hz).

### *Real-Fake RT experiment*

The observers and procedures of the Material RT experiment were employed to measure the speed of the real versus fake discrimination (see Figure 4a). Fake objects served as targets and the genuine objects from the same category served as distracters. Trials were blocked by the object category, and target prevalence was 50%. Four observers viewed all 125 images in the unambiguously real and fake set, while the rest (TU, RD, JJ, and KV in Figure 4b) viewed half of them. All observers viewed each image only once. Presentation order of object categories was counterbalanced between observers. All stimuli in this experiment were of size 15° × 12°.

## Results

In all cases, chance performance for the real versus fake discrimination corresponded to 50% accuracy.

### *Real-Fake RapidCat experiment*

Figure 9 plots the performance on both tasks, object categorization and real versus fake. The accuracy of object categorization was very high (94.6% at 40 ms), and it was not affected by exposure duration (40 ms vs. 2560 ms, $t[6] = -0.636$, $p = 0.548$). The accuracy of real versus fake discrimination, on the other hand, was much lower (74.9% at 40 ms), and a repeated measures ANOVA determined that there was a significant effect of exposure duration, $F(2, 12) = 18.38$, $p < 0.001$. Post hoc comparisons using the Tukey HSD test revealed a significant increase in accuracy from 40 ms to 320 ms (88.9%) as well as from 40 ms to 2560 ms (90.4%). Clearly, shape-based object information, while no doubt useful and available to observers, is insufficient for the real versus fake discrimination. These results tell us that: (a) Material recognition can be fast, even when its rapidity is dissociated from that of shape-based object recognition, and (b) observers can make fine material discriminations even in brief presentations.

### *Real-Fake RT experiment*

The results indicated in blue in Figure 4b and c correspond to this experiment. The accuracy of the real versus fake discrimination was 83.6%, and the median RT on correct trials was 703 ms. The real versus fake discrimination was faster than we had expected, but it was significantly slower than the material categorization task in Material RT experiment ($Z = -2.521$, $p = 0.012$). This outcome is not surprising because the real versus fake task is

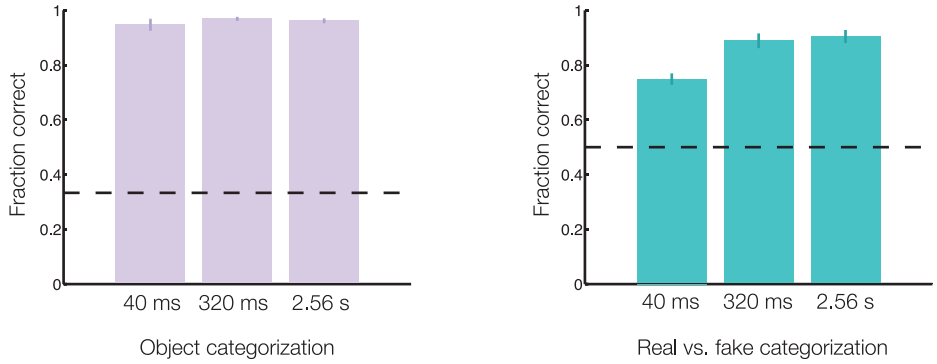



Figure 9. Shape-based object knowledge is insufficient for material recognition. Accuracy at identifying the object in each image as dessert, fruit, or flower (left panel) and as a real or fake example of that object (right panel) is shown here as a function of stimulus duration. Observers were able to identify the object category accurately in all three exposure conditions. Their performance at identifying real versus fake was lower, and it improved with longer exposures. The pattern of these results tells us that shape-based object identity is insufficient for the real versus fake material discrimination and that observers can make fine material discriminations even in brief presentations. Error bars represent 1 *SEM* across seven observers. Chance performance (black lines) is 0.33 for the object task and 0.5 for the real versus fake task.

inherently a harder discrimination task than the one used in the Material RT experiment.

## Discussion

Our world contains both objects and materials, but somehow objects dominate the study of visual recognition (Adelson, 2001). Perhaps this is because there is a strong bias towards referencing entities defined by shape (Landau, Smith, & Jones, 1988). We have names for nearly 1,500 basic-level object categories (Biederman, 1987) but nowhere near as many names for material categories. In an additional experiment, reported in Appendix A, we asked naive observers to name materials in 1,000 photographs of daily scenes. They were unable to name more than 50 materials, even when the granularity of their responses was ignored (e.g., fabric and silk were counted separately). In informal pilot experiments, we found that experienced observers (e.g., artists) were unable to name many more materials, even though they were able to describe precisely how a surface might look, feel, or react to forces. Patterson and Hays (2012) have obtained a similar count for material names (38) by studying a much larger collection of scenes.

Our goal in this paper has been to establish some of the basic properties of material recognition. We built a database with photographs from nine material categories, selecting images covering a wide range of material appearances. We found that material categories can be identified quickly and accurately, and that this rapid identification cannot be accomplished with simple cues, requiring that observers engage mechanisms of some sophistication. We argue that material categorization should be considered as a distinct task, different from the recognition of shapes, textures, objects, scenes, or faces.

### Material categorization can be fast

To measure the speed of material categorization, we used several tasks, paralleling tasks that have been used to study the rapid recognition of objects and scenes. We found that observers can categorize materials in exposures as brief as 40 ms, in the presence of backwards masking. This was true for categorizing one material out of many (Material RapidCat experiment), as well as for distinguishing real versus fake materials of a given object category (Real-Fake RapidCat experiment). The 40-ms value does not directly correspond to the duration of processing, but it is comparable to the brief exposures needed for object and scene categorization (Bacon-Mace et al., 2005; Greene & Oliva, 2009). We also measured RTs for material categorization and compared them to those for a pair of simple baseline tasks (Material RT experiment). Median RT for material categorization was approximately 100 ms slower than the baseline tasks of color and orientation discrimination, suggesting that 100 ms extra processing is needed for material recognition to occur. Median RT for real versus fake was 150 ms slower still (Real-Fake RT experiment), indicating that this task is more challenging.

Since we first presented these findings (Sharan et al., 2009), others have further probed the rapidity of material categorization. Wiebel et al. (2013) compared the time course of material categorization directly to that of object categorization. They found that material categorization can be as fast as basic-level object

categorization, and when low-level image properties (e.g., mean luminance, contrast) are normalized, it can even be as fast as superordinate-level object categorization. Wiebel et al. used close-ups of relatively flat surfaces from four material categories (wood, metal, fabric, and stone), and their observers performed four-way categorization tasks at exposures briefer than ours (e.g., 8 ms). While it is difficult to compare estimates of speed across studies (ours: 80.2% accuracy in 40-ms exposures, chance = 50%; theirs: 62.5% in 25-ms exposures, chance = 25%), it is safe to conclude that material categorization can be rapid for a range of experimental conditions.

We have also examined the relative speeds of object and material categorization in subsequent work (Xiao, Sharan, Rosenholtz, & Adelson, 2011). However, unlike Wiebel et al. (2013), we used the same set of images for the object and material tasks, which ensured that low-level image properties stayed the same in both conditions. We found that object categorization (glove vs. handbag) was faster for close-up views and material categorization (leather vs. fabric) was faster for regular views. These results do not contradict Wiebel et al.'s findings. There are many differences between our stimuli and theirs. It is fair to say that there is no clear answer yet; object and material categorization can both be fast, and depending on the stimuli (Johnson & Olshausen, 2003), one can be faster than the other.

## Material categorization is not merely based on simple feature judgments

Speed estimates are necessarily confounded with task difficulty, i.e., choice of image database (Johnson & Olshausen, 2003; Wichmann, Drewes, Rosas, & Gegenfurtner, 2010). One could make an easy database in which all the metal pictures had sharp angles, all the glass pictures had swooping curves, and all the wood was brown. Observers would be quite fast at telling the material categories apart, possibly because certain low-level features based on primitive color or outline shape would suffice for material recognition. We constructed an intentionally diverse database to avoid such confounds, and further, we tested if simple feature judgments could explain the speeds we have measured.

We modified the images in our database to deemphasize simple features such as color and high spatial frequencies (Material Degradation I). By deemphasizing a particular feature, we could determine how critical that feature was for material categorization. We found that observers did not critically depend on any one feature that we tested for, consistent with Wiebel et al.'s (2013) findings about the role of color in material categorization. It is worth noting that while simple features may not be critical for material categorization, they can play an important role in other material judgments (e.g., material change identification, Zaidi, 2011; fabric classification, Giesel & Zaidi, 2013).

Wolfe and Myers (2010) have asked whether material categories, as an attribute, can guide attention as efficiently as color, size, orientation, etc. They used close-up images from our database as stimuli and asked observers to perform visual search tasks such as finding a plastic target in a display of stone distracters. A variety of set sizes and target-distracter combinations were tested. The result was the same in each case: Search was inefficient. Wolfe and Myers concluded that material categories do not guide attention as efficiently as simpler attributes. We interpret this result to mean that material categorization is a distinct task, and especially for our stimuli, it cannot be explained in terms of simple feature judgments.

## Material categorization is not merely a surface property judgment

To allow judgments of material categories, our visual system has to process surface appearance to extract material-relevant information. The nature of this processing remains largely unknown (Gilchrist, 2006; Adelson, 2008; Brainard, 2009; Shevell & Kingdom, 2010; Anderson, 2011; Maloney et al., 2011; Fleming, 2012, 2014). In our experiments, we examined the role of four surface properties in material categorization: color, gloss, texture, and shape. Observers had no trouble judging the material category when one or more surface properties were deemphasized (Material Degradation I); however, they struggled when only one surface property was emphasized (Material Degradation II). These results tell us that the knowledge of color, gloss, texture, or shape, in itself, is not sufficient for material categorization. This is not to suggest that these properties are not important for the task, but that they are not useful individually.

It is possible that instead of relying on any one surface property, the visual system combines estimates of different surface properties to decide the material category. For example, a plastic versus nonplastic judgment is more easily made when color (uniform, saturated colors), gloss (shiny), surface shape (smooth, rounded), and texture (lack of strong textures) are considered together. In order to study the joint influence of different surface properties, one requires the ability to independently manipulate those properties. The images in our database were acquired in unknown conditions, which makes such manipulations impossible. Even if our images had been acquired in carefully controlled conditions, varying each surface property independently and smoothly is beyond the

capabilities of current image processing algorithms (Barron & Malik, 2013).

An alternative would have been to use synthetic stimuli, generated using computer graphics. Synthetic stimuli are useful because they allow smooth variations along perceptual dimensions of interest such as albedo (Nishida & Shinya, 1998), color (Yang & Maloney, 2001; Boyaci et al., 2003; Delahunt & Brainard, 2004), gloss (Pellacini et al., 2000; Fleming et al., 2003), or shape (Fleming, Torralba, & Adelson, 2004). Unfortunately, current computer graphics tools cannot easily represent the range of material categories that we have studied, or capture the richness and the diversity of appearance within each category (Jensen, 2001; Ward et al., 2006; Dorsey, Rushmeier, & Sillion, 2007; Igarashi, Nishino, & Nayar, 2007; Schröder, Zhao, & Zinke, 2012).

Fleming et al. (2013) have examined the relationship between surface appearance and material categories in another way. They used a subset of our database and asked observers to rate nine surface properties (e.g., glossiness, roughness, fragility) for each image. They found that the ratings of different properties form a distinctive signature for each material category (e.g., glass images were rated to be glossy, transparent, hard, cold, and fragile). In theory, such signatures could be used to infer the material category. However, as Fleming et al. point out, correlation does not imply causation. Despite the tight coupling between surface properties and material categories, it is unclear if material categorization requires the estimation of diagnostic surface properties.

In companion papers (Liu et al., 2010; Sharan et al., 2013), we have studied material categorization from a computational perspective. We developed a state-of-the-art computational model that takes as input an image and produces as output a material category label. This model computes local image features that loosely correlate with different surface properties: color (RGB pixel patches), texture (SIFT features, Lowe, 2004), shape (curvature of strong edges), and reflectance (region analysis of strong edges). The model then combines these features using supervised learning methods. When it is trained on half of our database, it is able to categorize images in the other half with 55.6% accuracy. This performance is well above chance (10%; a tenth category Foliage was added to the database) and higher than that of competing models (23.8%, Varma & Zisserman, 2008; 54%, Hu et al., 2011).

Although our modeling effort is nowhere close to explaining, or even matching, human performance (84.9% at 10-way categorization, unlimited exposures), it supports the conclusions of the present study in two ways. One, material categorization is a distinct task. Models that succeed at texture recognition (Varma & Zisserman, 2008) fail at material categorization. Success on our task requires strategies different from those used in texture or object recognition. Two, simple low-level features, of the type used in our model, are not sufficient for material categorization. This is not to suggest that these features are not useful (accuracy for individual features: 25%–40%), but that they offer limited clues to the high-level material category (combined accuracy: 55.6%).

## Material categorization is not merely derived from high-level object recognition

The relationship between objects and materials is not arbitrary. Most objects (e.g., spoons) are made of certain materials (e.g., plastic, porcelain, metal, wood) and not others (e.g., paper, fabric). To understand the role of object knowledge in material categorization, we used a new set of images designed to dissociate object and material identities. For each image, observers made an object judgment (dessert, flowers, or fruit) and a material judgment (real vs. fake). Object judgments were faster and more accurate than real versus fake judgments, which suggests that object categorization engages different mechanisms from material categorization. This conclusion is borne out in subsequent work (Xiao et al., 2011; Wiebel et al., 2013) for different choices of object and material judgments.

Further support for the distinction between object and material categorization comes from an early study of a visual form agnosic (Humphrey et al., 1994) and more recent fMRI investigations (Newman, Klatzky, Lederman, & Just, 1995; Cant & Goodale, 2007; Cavina-Pratesi, Kentridge, Heywood, & Milner, 2010; Cant & Goodale, 2011; Hiramatsu et al., 2011). Humphrey et al. reported that patient DF, who suffers from profound visual form agnosia, was able to identify materials (e.g., brass) even though she was unable to identify objects (e.g., ashtray). In subsequent fMRI studies with neurologically normal participants, Cant and Goodale (2007, 2011) have shown that 3-D shape properties, which determine object identity, are processed in a different brain region from material properties. Behavioral reports support these findings (Cant, Large, McCall, & Goodale, 2008), although there is clear evidence of an interaction: Material properties influence the perception of 3-D shape (Fleming et al., 2004; Ho et al., 2008; Fleming, Holtmann-Rice, & Buelthoff, 2011; Wijntjes, Doerschner, Kucukoglu, & Pont, 2012) and vice versa (Bloj et al., 1999; Boyaci et al., 2003; VanGorp et al., 2007; Ho et al., 2008; Olkonnen & Brainard, 2011; Marlow, Kim, & Anderson, 2012; Vurro et al., 2013).

## Conclusions

We studied judgments of high-level material categories using real-world images. The images we used capture the natural range of material appearances. We found that material categories can be identified quickly, requiring only 100 ms more than simple baseline tasks, and accurately, achieving 80.2% accuracy even in 40-ms exposures (chance = 50%). This performance cannot be explained in terms of simple feature judgments of color, gloss, texture, or shape. Nor can it be explained as a consequence of shape-based object recognition. We argue that fast and accurate material categorization is a distinct, basic ability of the visual system.

Put together, recent work (Sharan, 2009; Liu et al., 2010; Wolfe & Myers, 2010; Hu et al., 2011; Xiao et al., 2011; Fleming et al., 2013; Sharan et al., 2013; Wiebel et al., 2013) has taught us a great deal about material categorization that was not known before. Despite this progress, there remain many unanswered questions. Surface appearance clearly determines the material category, but how? Does the visual system combine an assortment of surface properties to categorize materials (Fleming et al., 2013)? Do the mechanisms underlying material categorization differ from those underlying object categorization? Or for that matter, scene categorization? The relationships between materials and objects, and even materials and scenes, exhibit certain regularities (Steiner, 1998), many of which are yet to be understood.

*Keywords: material perception, material categories, material properties, real-world stimuli*

## Acknowledgments

We thank Aseema Mohanty for help with database creation, Alvin Raj for image quilting code that was used to generate Texture manipulations, Ce Liu for help with the study described in Appendix A, Aude Oliva for use of eye-tracking equipment that was used in preliminary experimentation, and for discussions: Aude Oliva, Michelle R. Greene, Barbara Hidalgo-Sotelo, Molly Potter, Nancy Kanwisher, Jeremy Wolfe, Roland Fleming, Shin'ya Nishida, Isamu Motoyoshi, Micah K. Johnson, Alvin Raj, Ce Liu, James Hays, and Bei Xiao. This work was supported by NIH grants R01-EY019262 and R21-EY019741 and a grant from NTT Basic Research Laboratories. The photographs in Figures 1 through 6, and 8 were acquired from the website Flickr.com courtesy of the following users:

Figure 1. Darren Johnson, David Martyn Hunt, and Michael Fraley

Figure 2. Fabian Bromann, tanakawho, Randy Robertson, Crimson & Clover Vintage, Robert Brook, and millicent_bystander

Figure 3. Siti Saad, Breibeest, Ute, liz west, Bart Everson, Luz, La Petit Poulailler, sozoooo, Nelson Alexandre Rocha, Janine, Damien du Toit, Crimson & Clover Vintage, Audrey, Ctd 2005, Kevin Dooley, apintogsphotos, Mikko Lautamäki, Nanimo, lotyloty, Jenni Douglas, Randy Robertson, tanakawho, liz west, Carolyn Williams, Michel Filion, sallypics, Stefan Andrej Shambora, Ryan McDonald, Paulio Geordio, tico_24, Ishmael Orendain, mrphishphotography, spDuchamp, spDuchamp, Fabian Bromann, and Leonardo Aguiar

Figure 4a, Figure 5a. Ryan McDonald

Figure 6. Bart Everson and Muhammad Ghouri

Figure 8. The Shopping Sherpa, Kerry Breyette, Jubei Kibagami, Wendy Copley, Mykl Roventine, and edvvc

All photographs were acquired under the Creative Commons (CC) BY 2.0 License with the following exceptions: (i) cut-glass bowl in Figure 3b (CC BY-NC-ND 2.0); and (ii) fake cupcake (CC ND 2.0), real cupcake (CC BY-NC-SA 2.0), and real flowers (CC BY-NC-SA 2.0) in Figure 8. The image of the fake cupcake was magnified for clarity.

Commercial deposit: none.
Corresponding author: Lavanya Sharan.
Email: sharan@alum.mit.edu.
Address: Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge MA, USA.

## References

Adelson, E. H. (2001). On seeing stuff: The perception of materials by humans and machines. In B. E. Rogowitz & T. N. Pappas (Eds.), *SPIE: Vol. 4299. Human vision and electronic imaging VI* (pp. 1–12). doi:10.1117/12.784132.

Adelson, E. H. (2008). Image statistics and surface perception. In B. E. Rogowitz & T. N. Pappas (Eds.), *Human vision and electronic imaging XIII, Vol.* 6809 (pp. 1–9). doi:10.1117/12.784132.

Anderson, B. (2011). Visual perception of materials and surfaces. *Current Biology*, *21*(24), R978–R983.

Bacon-Mace, N., Mace, M. J., Fabre-Thorpe, M., & Thorpe, S. J. (2005). The time course of visual processing: Backward masking and natural scene categorization. *Vision Research*, *45*(11), 1459–1469.

Barron, J. T., & Malik, J. (2011). High-frequency shape and albedo from shading using natural image

statistics. In *IEEE conference on computer vision and pattern recognition* (pp. 2521–2528). doi:10.1109/cvpr.2011.5995392.

Barron, J. T., & Malik, J. (2013, May). *Shape, illumination, and reflectance from shading* (Technical Report No. UCB/EECS-2013-117). Retrieved from the University of California at Berkeley, Electrical Engineering and Computer Science, http://www.cs.berkeley.edu/~barron/BarronMalikTR2013.pdf.

Beck, J., & Prazdny, S. (1981). Highlights and the perception of glossiness. *Perception & Psychophysics*, *30*(4), 401–410.

Bell, S., Upchurch, P., Snavely, N., & Bala, K. (2013). Opensurfaces: A richly annotated catalog of surface appearance. *Proceedings: ACM Transactions on Graphics - Proceedings of ACM SIGGRAPH 2013, 34*(4), 1–11. doi:10.1145/2461912.2462002.

Berzhanskaya, J., Swaminathan, G., Beck, J., & Mingolla, E. (2005). Remote effects of highlights on gloss perception. *Perception*, *34*(5), 565–575.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*(2), 115–147.

Biederman, I., Rabinowitz, J. C., Glass, A. L., & Stacy, E. W. (1974). On information extracted from a glance at a scene. *Journal of Experimental Psychology*, *103*(3), 597–600.

Bloj, M., Kersten, D., & Hurlbert, A. C. (1999). Perception of three-dimensional shape influences color perception through mutual illumination. *Nature*, *402*, 877–879.

Boivin, S., & Gagalowicz, A. (2001). Image based rendering of diffuse, specular and glossy surfaces from a single image. In *Proceedings of 28th annual conference on computer graphics and interactive techniques* (pp. 107–116).

Boyaci, H., Maloney, L. T., & Hersh, S. (2003). The effect of perceived surface orientation on perceived surface albedo in binocularly viewed scenes. *Journal of Vision*, *3*(8):2, 541–553, http://www.journalofvision.org/content/3/8/2, doi:10.1167/3.8.2. [PubMed] [Article]

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.

Brainard, D. H. (2009). Bayesian approaches to color vision. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (4th ed., pp. 395–408). Cambridge, MA: MIT Press.

Brainard, D. H., & Maloney, L. T. (2011). Surface color perception and equivalent illumination models. *Journal of Vision*, *11*(5):1, 1–18, http://www.journalofvision.org/content/11/5/1, doi:10.1167/11.5.1. [PubMed] [Article]

Cant, J. S., & Goodale, M. A. (2007). Attention to form or surface properties modulates different regions of human occipitotemporal cortex. *Cerebral Cortex*, *17*, 713–731.

Cant, J. S., & Goodale, M. A. (2011). Scratching beneath the surface: New insights into the functional properties of the lateral occipital area and parahippocampal place area. *Journal of Neuroscience*, *31*(22), 8248–8258.

Cant, J. S., Large, M. E., McCall, L., & Goodale, M. A. (2008). Indepdent processing of form, color, and texture in object perception. *Perception*, *37*(1), 57–78.

Cavina-Pratesi, C., Kentridge, R. W., Heywood, C. A., & Milner, A. D. (2010). Separate processing of texture and form in the ventral stream: Evidence from fmri and visual agnosia. *Cerebral Cortex*, *20*(2), 433–446.

Dana, K. J., van Ginneken, B., Koenderink, J. J., & Nayar, S. (1999). Reflectance and texture of real world surfaces. *ACM Transactions on Graphics*, *18*, 1–34.

Delahunt, P. B., & Brainard, D. H. (2004). Color constancy under changes in reflected illumination. *Journal of Vision*, *4*(9):8, 764–778, http://www.journalofvision.org/content/4/9/8, doi:10.1167/4.9.8. [PubMed] [Article]

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009,* 248–255. doi:10.1109/cvpr.2009.5206848.

Doerschner, K., Fleming, R. W., Yilmaz, O., Schrater, P. R., Hartung, B., & Kersten, D. (2011). Visual motion and the perception of surface material. *Current Biology*, *21*(23), 2010–2016.

Doerschner, K., Maloney, L. T., & Boyaci, H. (2010). Perceived glossiness in high dynamic range scenes. *Journal of Vision*, *10*(9):11, 1–11, http://www.journalofvision.org/content/10/9/11, doi:10.1167/10.9.11. [PubMed] [Article]

Dorsey, J., Rushmeier, H., & Sillion, F. (2007). *Digital modeling of material appearance*. Burlington, MA: Morgan Kauffman.

Dror, R., Adelson, E. H., & Willsky, A. S. (2001). Recognition of surface reflectance properties from a single image under unknown illumination. *Proceedings of the IEEE workshop on identifying objects across variations in lighting: psychophysics and computation*, Kauai, Hawaii.

Efros, A., & Freeman, W. T. (2001). Image quilting for texture synthesis and transfer. *Proceedings of ACM SIGGRAPH, 2001,* 341–346.

Everingham, M., Zisserman, A., Williams, C., Gool, L. V., Allan, M., Bishop, C., ... Zhang, J. (2005). The 2005 pascal visual object classes challenge. In J. Quiñonero-Candela, I. Dagan, B. Magnin, & F. d'Alché-Buc (Eds.), *Maching learning challenges, evaluating predictive uncertainty, visual object classication and recognizing textual entailment. Series: Lecture notes in artificial intelligence* (pp. 117–176).

Fei-Fei, L., Fergus, R., & Perona, P. (2006). One-shot learning of object categories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *28*(4), 594–611.

Fleming, R. W. (2012). Visual heuristics in the perception of glossiness. *Current Biology*, *22*(20), R865–R866.

Fleming, R. W. (2014). Visual perception of materials and their properties. *Vision Research*, *94*, 62–75.

Fleming, R. W., & Buelthoff, H. (2005). Low-level image cues in the perception of translucent materials. *ACM Transactions on Applied Perception*, *2*(3), 346–382.

Fleming, R. W., Dror, R. O., & Adelson, E. H. (2003). Real-world illumination and the perception of surface reflectance properties. *Journal of Vision*, *3*(5):3, 347–368, http://www.journalofvision.org/content/3/5/3, doi:10.1167/3.5.3. [PubMed] [Article]

Fleming, R. W., Holtmann-Rice, D., & Buelthoff, H. H. (2011). Estimation of 3-D shape from image orientations. *Proceedings of the National Academy of Sciences*, *108*(51), 20438–20443.

Fleming, R. W., Jakel, F., & Maloney, L. T. (2011). Visual perception of thick transparent materials. *Psychological Science*, *22*(6), 812–820.

Fleming, R. W., Torralba, A., & Adelson, E. H. (2004). Specular reflections and the perception of shape. *Journal of Vision*, *4*(9), 798–820, doi:10.1167/4.9.10. [PubMed] [Article]

Fleming, R. W., Wiebel, C., & Gegenfurtner, K. R. (2013). Perceptual qualities and material classes. *Journal of Vision, 13*(8):9, 1–20, http://www.journalofvision.org/content/13/8/9, doi:10.1167/13.8.9. [PubMed] [Article]

Ged, G., Obein, G., Silvestri, Z., Rohellec, J. L., & Viéntot, F. (2010). Recognizing real materials from their glossy appearance. *Journal of Vision*, *10*(9):18, 1–17, http://www.journalofvision.org/content/10/9/18, doi:10.1167/10.9.18. [PubMed] [Article]

Gerhard, H. E., & Maloney, L. T. (2010). Detection of light transformations and concomitant changes in surface albedo. *Journal of Vision*, *10*(9):1, 1–14, http://www.journalofvision.org/content/10/9/1, doi:10.1167/10.9.1. [PubMed] [Article]

Giesel, M., & Gegenfurtner, K. R. (2010). Color appearance of real objects varying in material, hue and shape. *Journal of Vision*, *10*(9):10, 1–21, http://www.journalofvision.org/content/10/9/10, doi:10.1167/10.9.10. [PubMed] [Article]

Giesel, M., & Zaidi, Q. (2013). Frequency-based heuristics for material perception. *Journal of Vision*, *13*(14):7, 1–19, http://www.journalofvision.org/content/13/14/7, doi:10.1167/13.14.7. [PubMed] [Article]

Gilchrist, A. L. (2006). *Seeing black and white*. New York: Oxford University Press.

Gilchrist, A. L., & Jacobsen, A. (1984). Perception of lightness and illumination in a world of one reflectance. *Perception*, *13*, 5–19.

Gilchrist, A. L., Kossyfidis, C., Bonato, F., Agostini, T., Cataliotti, J., Li, X., ... Economou, E. (1999). An anchoring theory of lightness perception. *Psychological Review*, *106*, 795–834.

Greene, M. R., & Oliva, A. (2009). The briefest of glances: The time course of natural scene understanding. *Psychological Science*, *20*(4), 464–472.

Grill-Spector, K., & Kanwisher, N. (2005). Visual recognition: As soon as you see it, you know what it is. *Psychological Science*, *16*(2), 152–160.

Grosse, R., Johnson, M. K., Adelson, E. H., & Freeman, W. T. (2009). A ground-truth dataset and baseline evaluations for intrinsic image algorithms. *Proceedings of the 2009 IEEE International Conference on Computer Vision, 2009,* 2335–2342. doi:10.1109/ICCV.2009.54594328.

Hartung, B., & Kersten, D. (2002). Distinguishing shiny from matte. *Journal of Vision*, *2*(7):551, http://www.journalofvision.org/content/2/7/551, doi:10.1167/2.7.551. [Abstract]

Hearst, M. (1995). Tilebars: Visualization of term distribution information in full text information access. In I. R. Katz, R. Mack, L. Marks, M. B. Rosson, & J. Nielsen (Eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 59–66). doi:10.1145/223904.223912.

Hiramatsu, C., Goda, N., & Komatsu, H. (2011). Transformation from image-based to perceptual representation of materials along the human ventral visual pathway. *Neuroimage*, *57*(2), 482–494.

Ho, Y. X., Landy, M. S., & Maloney, L. T. (2008).

Conjoint measurement of gloss and surface texture. *Psychological Science*, *19*(2), 196–204.

Hu, D., Bo, L., & Ren, X. (2011). Toward robust material recognition for everyday objects. In J. Hoey, S. McKenna, & E. Trucco (Eds.), *Proceedings of the British machine vision conference, 2011,* (pp. 48.1–48.11).

Humphrey, G. K., Goodale, M. A., Jakobson, L. S., & Servos, P. (1994). The role of surface information in object recognition: Studies of a visual form agnosic and normal subjects. *Perception*, *23*, 1457–1481.

Igarashi, T., Nishino, K., & Nayar, S. K. (2007). The appearance of human skin: A survey. *Foundations and Trends in Computer Graphics and Vision*, *3*, 1–95.

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

Jensen, H. W. (2001). *Realistic image synthesis using photon mapping*. Wellesley, MA: AK Peters.

Johnson, J. S., & Olshausen, B. A. (2003). Timecourse of neural signatures of object recognition. *Journal of Vision*, *3*(7):4, 499–512, http://www.journalofvision.org/content/3/7/4, doi:10.1167/3.7.4. [PubMed] [Article]

Kerrigan, I. S., & Adams, W. J. (2013). Highlights, disparity, and perceived gloss with convex and concave surfaces. *Journal of Vision*, *13*(1):9, 1–10, http://www.journalofvision.org/content/13/1/9, doi:10.1167/13.1.9. [PubMed] [Article]

Kim, J., Marlow, P. J., & Anderson, B. (2011). The perception of gloss depends on highlight congruence with surface shading. *Journal of Vision*, *11*(9):4, 1–19, http://www.journalofvision.org/content/11/9/4, doi:10.1167/11.9.4. [PubMed] [Article]

Kim, J., Marlow, P. J., & Anderson, B. L. (2012). The dark side of gloss. *Nature Neuroscience*, *15*, 1590–1595.

Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, *3*, 299–321.

Liu, C., Sharan, L., Rosenholtz, R., & Adelson, E. H. (2010). Exploring features in a bayesian framework for material recognition. In *2010 IEEE conference on computer vision and pattern recognition* (pp. 239–246). doi:10.1109/CVPR.2010.5540207.

Lowe, D. G. (2004). Distinctive image-features from scale-invariant keypoints. *International Journal of Computer Vision*, *60*(2), 91–110.

Maloney, L. T., Gerhard, H. E., Boyaci, H., and Doerschner, K. (2011). Surface color perception and light field estimation in 3d scenes. In L. R. Harris & M. R. M. Jenkin (Eds.), *Vision in 3D environments* (pp. 65–88). Cambridge, UK: Cambridge University Press.

Maloney, L. T., & Yang, J. N. (2003). The illumination estimation hypothesis and surface color perception. In R. Mausfeld & D. Heyer (Eds.), *Colour: Connecting the mind to the physical world* (pp. 335–358). Oxford, UK: Oxford University Press.

Marlow, P. J., Kim, J., & Anderson, B. (2012). The perception and misperception of specular surface reflectance. *Current Biology*, *22*(20), 1909–1913.

Marlow, P. J., Kim, J., & Anderson, B. L. (2011). The role of brightness and orientation congruence in the perception of surface gloss. *Journal of Vision*, *11*(9):16, 1–12, http://www.journalofvision.org/content/11/9/16, doi:10.1167/11.9.16. [PubMed] [Article]

Matusik, W., Pfister, H. P., Brand, M., & McMillan, L. (2003). A data-driven reflectance model. In 30th annual conference on computer graphics and interactive techniques. *ACM Transactions on Graphics (TOF) - Proceedings of ACM SIGGRAPH 2003, 22*(3), 759–769.

Motoyoshi, I. (2010). Highlight-shading relationship as a cue for the perception of translucent and transparent materials. *Journal of Vision*, *10*(9):6, 1–11, http://www.journalofvision.org/content/10/9/6, doi:10.1167/10.9.6. [PubMed] [Article]

Motoyoshi, I., & Matoba, H. (2011). Variability in constancy of the perceived surface reflectance across different illumination statistics. *Vision Research*, *53*, 30–39.

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

Nagai, T., Ono, Y., Tani, Y., Koida, K., Kitazaki, M., & Nakauchi, S. (2013). Image regions contributing to perceptual translucency: A psychophysical reverse-correlation study. *i-Perception*, *4*(6), 407–428, doi:10.1068/i0576.

Nandakumar, C., & Malik, J. (2009). Understanding rapid category detection via multiply degraded images. *Journal of Vision*, *9*(6):19, 1–8, http://www.journalofvision.org/content/9/6/19, doi:10.1167/9.6.19. [PubMed] [Article]

Newman, S. D., Klatzky, R. L., Lederman, S. J., & Just, M. A. (1995). Imagining material versus geometric properties of objects: An fmri study. *Cognitive Brain Research*, *23*, 235–246.

Nishida, S., & Shinya, M. (1998). Use of image-based information in judgments of surface-reflectance

properties. *Journal of Optical Society of America A*, *15*, 2951–2965.

Obein, G., Knoblauch, K., & Viéntot, F. (2004). Difference scaling of gloss: Nonlinearity, binocularity, and constancy. *Journal of Vision*, *4*(9):4, 711–720, http://www.journalofvison.org/content/4/9/4, doi:10.1167/4.9.4. [PubMed] [Article]

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

Olkkonen, M., Hansen, T., & Gegenfurtner, K. R. (2008). Color appearance of familiar objects: Effects of object shape, texture, and illumination changes. *Journal of Vision*, *8*(5):13, 1–16, http://www.journalofvision.org/content/8/5/13, doi:10.1167/8.5.13. [PubMed] [Article]

Olkkonen, M., & Brainard, D. H. (2010). Perceived glossiness and lightness under real-world illumination. *Journal of Vision*, *10*(9):5, 1–19, http://www.journalofvision.org/content/10/9/5, doi:10.1167/10.9.5. [PubMed] [Article]

Olkkonen, M., & Brainard, D. H. (2011). Joint effects of illumination geometry and object shape in the perception of surface reflectance. *i-Perception*, *2*, 1014–1034.

Patterson, G., & Hays, J. (2012). Sun attribute database: Discovering, annotating, and recognizing scene attributes. In *2012 IEEI conference on computer vision and pattern recognition* (pp. 2751–2758). doi:10.1109/CVPR.2012.6247998.

Pellacini, F., Ferwerda, J. A., & Greenberg, D. P. (2000). Towards a psychophysically-based light reflection model for image synthesis. In *Proceedings of ACM SIGGRAPH 2000* (pp. 55–64).

Pont, S. C., & Koenderink, J. J. (2005). Bidirectional texture contrast function. *International Journal of Computer Vision*, *62*(1/2), 17–34.

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

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

Potter, M. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning & Memory*, *2*, 509–522.

Radonjić, A., Todorović, D., & Gilchrist, A. (2010). Adjacency and surroundedness in the depth effect on lightness. *Journal of Vision*, *10*(9):12, 1–16, http://www.journalofvision.org/content/10/9/12, doi:10.1167/10.9.12. [PubMed] [Article]

Robilotto, R., & Zaidi, Q. (2004). Limits of lightness identification of real objects under natural viewing conditions. *Journal of Vision*, *4*(9):9, 779–797, http://www.journalofvision.org/content/4/9/9, doi:10.1167/4.9.9. [PubMed] [Article]

Robilotto, R., & Zaidi, Q. (2006). Lightness identification of patterned three-dimensional, real objects. *Journal of Vision*, *6*(1):3, 18–36, http://www.journalofvision.org/content/6/1/3, doi:10.1167/6.1.3. [PubMed] [Article]

Romeiro, F., & Zickler, T. E. (2010). Blind reflectometry. In K. Daniildis, P. Maragos, & N. Paragios (Eds.), *11th European conference on computer vision, Proceedings, Part I,* September 5–11, 2010, Heraklion, Crete, Greece.

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

Russell, B., Torralba, A., Murphy, K. P., & Freeman, W. T. (2008). Labelme: A database and web-based tool for image annotation. *International Journal of Computer Vision*, *77*(1–3), 157–173.

Sakano, Y., & Ando, H. (2010). Effects of head motion and stereo viewing on perceived glossiness. *Journal of Vision*, *10*(9):15, 1–14, http://www.journalofvision.org/content/10/9/15, doi:10.1167/10.9.15. [PubMed] [Article]

Schröder, K., Zhao, S., & Zinke, A. (2012). Recent advances in physically-based appearance modeling of cloth. In *SIGGRAPH Asia 2012 Courses* (pp. 1–52). New York: ACM.

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time and spatial scale dependent scene recognition. *Psychological Science*, *5*, 195–200.

Sharan, L. (2009). *The perception of material qualities in real-world images* (Unpublished doctoral dissertation). Massachusetts Institute of Technology, Boston, MA.

Sharan, L., Li, Y., Motoyoshi, I., Nishida, S., & Adelson, E. H. (2008). Image statistics for surface reflectance perception. *Journal of Optical Society of America A*, *25*(4), 846–865.

Sharan, L., Liu, C., Rosenholtz, R., & Adelson, E. H. (2013). Recognizing materials using perceptually inspired features. *International Journal of Computer Vision*, *103*(3), 348–371.

Sharan, L., Rosenholtz, R., & Adelson, E. H. (2009). Material perception: What can you see in a brief glance? *Journal of Vision*, *9*(8):784, http://www.journalofvision.org/content/9/8/784, doi:10.1167/9.8.784. [Abstract]

Shevell, S. K., & Kingdom, F. A. A. (2010). Color in complex scenes. *Annual Review of Psychology*, *59*, 143–166.

Steiner, J. (1998). *Look-alikes* (1st ed.). New York: Megan Tingley.

Tappen, M., Freeman, W. T., & Adelson, E. H. (2005). Recovering intrinsic images from a single image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *27*(9), 1459–1472.

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

Todd, J. T. (2004). The visual perception of 3d shape. *Trends in Cognitive Sciences*, *8*(3), 115–121.

Todd, J. T., Norman, J. F., & Mingolla, E. (2004). Lightness constancy in the presence of specular highlights. *Psychological Science*, *15*, 33–39.

Tominaga, S., & Tanaka, N. (2000). Estimating reflection parameters from a single color image. *IEEE Computer Graphics and Applications*, *20*, 58–66.

VanGorp, P., Laurijssen, J., & Dutre, P. (2007). The influence of shape on the perception of material reflectance. *ACM Transactions on Graphics (TOG) - Proceedings of ACM SIGGRAPH 2007, 26*(3), 1–9.

Varma, M., & Zisserman, A. (2008). A statistical approach to material classification using image patch exemplars. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *99*(1), 2032–2047.

Velisavljevic, L., & Elder, J. H. (2009). Cue dynamics underlying rapid detection of animals in natural scenes. *Journal of Vision*, *9*(7):7, 1–20, http://www.journalofvision.org/content/9/7/7, doi:10.1167/9.7.7. [PubMed] [Article]

Vurro, M., Ling, Y., & Hurlbert, A. C. (2013). Memory color of natural familiar objects: Effects of surface texture and 3-d shape. *Journal of Vision*, *13*(7):20, 1–20, http://www.journalofvision.org/content/13/7/20, doi:10.1167/13.7.20. [PubMed] [Article]

Ward, K., Bertails, F., Kim, T.-Y., Marschner, S. R., Cani, M.-P., & Lin, M. C. (2006). A survey on hair modeling: Styling, simulation, and rendering. *IEEE Transactions on Visualization and Computer Graphics*, *13*(2), 213–234.

Wendt, G., Faul, F., Ekroll, V., & Mausfeld, R. (2010). Disparity, motion, and color information improve gloss constancy performance. *Journal of Vision*, *10*(9):7, 1–17, http://www.journalofvision.org/content/10/9/7, doi:10.1167/10.9.7. [PubMed] [Article]

Wichmann, F. A., Drewes, J., Rosas, P., & Gegenfurtner, K. R. (2010). Animal detection in natural scenes: Critical features revisited. *Journal of Vision*, *10*(4):6, 1–27, http://www.journalofvision.org/content/10/4/6, doi:10.1167/10.4.6. [PubMed] [Article]

Wiebel, C., Valsecchi, M., & Gegenfurtner, K. R. (2013). The speed and accuracy of material recognition in natural images. *Attention, Perception, & Psychophysics*, *75*(5), 954–966.

Wijntjes, M., Doerschner, K., Kucukoglu, G., & Pont, S. (2012). Relative flattening between velvet and matte 3-d shapes: Evidence for similar shape-from-shading computations. *Journal of Vision*, *12*(1):2, 1–11, http://www.journalofvision.org/content/12/1/2, doi:10.1167/12.1.2. [PubMed] [Article]

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, 1–9, http://www.journalofvision.org/content/10/9/8, doi:10.1167/10.9.8. [PubMed] [Article]

Xiao, B., & Brainard, D. H. (2008). Surface gloss and color perception of 3d objects. *Visual Neuroscience*, *25*, 371–385.

Xiao, B., Sharan, L., Rosenholtz, R., & Adelson, E. H. (2011). Speed of material vs. object recognition depends upon viewing conditions. *Journal of Vision*, *11*(15): 19, http://www.journalofvision.org/content/11/15/19, doi:10.1167/11.15.19. [Abstract]

Xiao, B., Walter, B., Gkioulekas, I., Zickler, T., Adelson, E. H., & Bala, K. (2014). Looking against the light: How perception of translucency depends on lighting direction. *Journal of Vision*, *14*(3):17, 1–22, http://www.journalofvision.org/content/14/3/17, doi:10.1167/14.3.17. [PubMed] [Article]

Yang, J. N., & Maloney, L. T. (2001). Illuminant cues in surface color perception: Tests of three candidate cues. *Vision Research*, *41*(20), 2581–2600.

Zaidi, Q. (2011). Visual inferences of material changes: Color as clue and distraction. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(6), 686–700.

## Appendix A

We chose nine common materials, as shown in Figure 3, to study material categorization in real-world images. It is reasonable to ask if this specific choice of materials was justified. Are these nine materials the most common in daily life? If not, which materials should we have chosen? Unlike for objects (Rosch, 1978), there are no studies on the familiarity of various

material categories. Therefore, we conducted an exploratory study to determine the most common materials in daily life.

We collected a set of 1,000 photographs of everyday scenes from the photo sharing website, Flickr.com. To construct a set of images that was representative of the visual experience of an average person, we searched for images that conveyed the following settings: street, buildings, office, kitchen, bedroom, bathroom, people, shopping, outdoors, and transportation. In addition to these keywords, we searched for food, drinks, kitchen equipment, clothes, computers, skin, store inside, trees, sky, and seat to include close-up images of the materials in each setting. We collected 50 color images for each of these 20 keywords. The resolution of the images ranged from 300 × 450 to 1280 × 1280 pixels.

Five naive observers were asked to annotate the materials in each image for as many images as they could complete. Observers were asked to focus on the materials that occupied the largest image regions first. They were also asked to provide a category label if they were not sure of precise material name (e.g., metal if confused between tin and aluminum). Three observers annotated nonoverlapping sets of 300, 299, and 221 images, respectively. Responses from these observers were used to create a list of suggestions for the remaining two observers who annotated all 1,000 images. Observers were given as much time as needed per image, and the presentation order of images was randomized for each observer.

The annotations thus obtained were interpreted by experimenters so that misspellings (e.g., mettal) and subordinate category labels (e.g., aluminum) were counted towards the intended category (e.g., metal). The twenty most frequently occurring categories in decreasing order of occurrence were: metal, fabric, wood, glass, plastic, stone, ceramic, greenery, skin, paper, food, hair, air, water, clouds, leather, rubber, mirror, sand, and snow. These results confirm that the nine categories used in our Material experiments were a reasonable choice (ranks 1–6, 10, 14, and 16 out of 20).

While designing this study, we were motivated by an average person's daily visual experience. It is likely that a different selection of scenes would yield a different ordering of material categories. The results presented here are merely a preliminary step towards studying the prevalence and hierarchy of material categories in the real world.

# Appendix B

In the main paper, we focused on establishing some of the basic properties of material categorization. To avoid effects that were driven by a particular material category (e.g., wood) or a particular view type (e.g., close-up), we averaged across material categories and view types in the Material experiments (RT, RapidCat, Degradations I and II). For the interest of the reader, we present here a more complete view of the data. In Figure 10, we replot the data from Figures 4c, 5b, and 7 for each material category and view type. The reader should note, however, that this data may be insufficient for answering questions for which we did not design our experiments. Therefore, the trends observed in Figure 10 have to be interpreted with caution.

### *Influence of material category*

Although we used an intentionally diverse selection of images, there are clear hints of category-specific effects in Figure 10. For example, stone was easier to identify than metal in 40-ms exposures (Figure 10b), whereas fabric was easier to identify than paper in Texture II stimuli (Figure 10d). When we designed our Material experiments, we were not interested in measuring the influence of individual material categories, so we recruited a small number of observers (5–8 observers per experiment) and tested them on a few categories each (3–9 categories per experimental condition). As a result, there is insufficient statistical power to expect differences between material categories to be significant.

### *Influence of view type*

Figure 10 suggests that view type had no influence in any of our Material experiments. Statistical analyses confirmed this observation. For the Material RT experiment, a Wilcoxon signed-rank test determined that there was no significant difference in median RTs for close-up versus regular views ($Z = 15$, $p = 0.7422$). For the Material RapidCat experiment, a two-way repeated measures ANOVA revealed a significant effect of exposure duration, $F(2, 8) = 23.23$, $p < 0.001$, but no significant effect of view type, $F(1, 4) = 2.59$, $p = 0.1828$, and no significant interaction between exposure duration and view type, $F(2, 8) = 0.64$, $p = 0.5523$. For the Material Degradations I experiment, there were no significant differences between close-up and regular views in any of the conditions, Baseline: $t(3) = -2$, $p = 0.1098$; Grayscale: $t(1) = -29$, $p = 0.0222$; Grayscale Blurred: $t(2) = 1$, $p = 0.5582$; Grayscale Negative: $t(2) = -1$, $p = 0.4544$; Bonferroni correction yields 0.0125 significance level. For the Material Degradations II experiment, a two-way repeated measures ANOVA revealed a significant effect of manipulation, $F(2, 8) = 4.61$, $p = 0.0465$, but no significant effect of view type, $F(1, 4) = 0.03$, $p = 0.8662$, and no significant interaction between image manipulation and view type, $F(2, 8) = 0.17$, $p = 0.8448$.

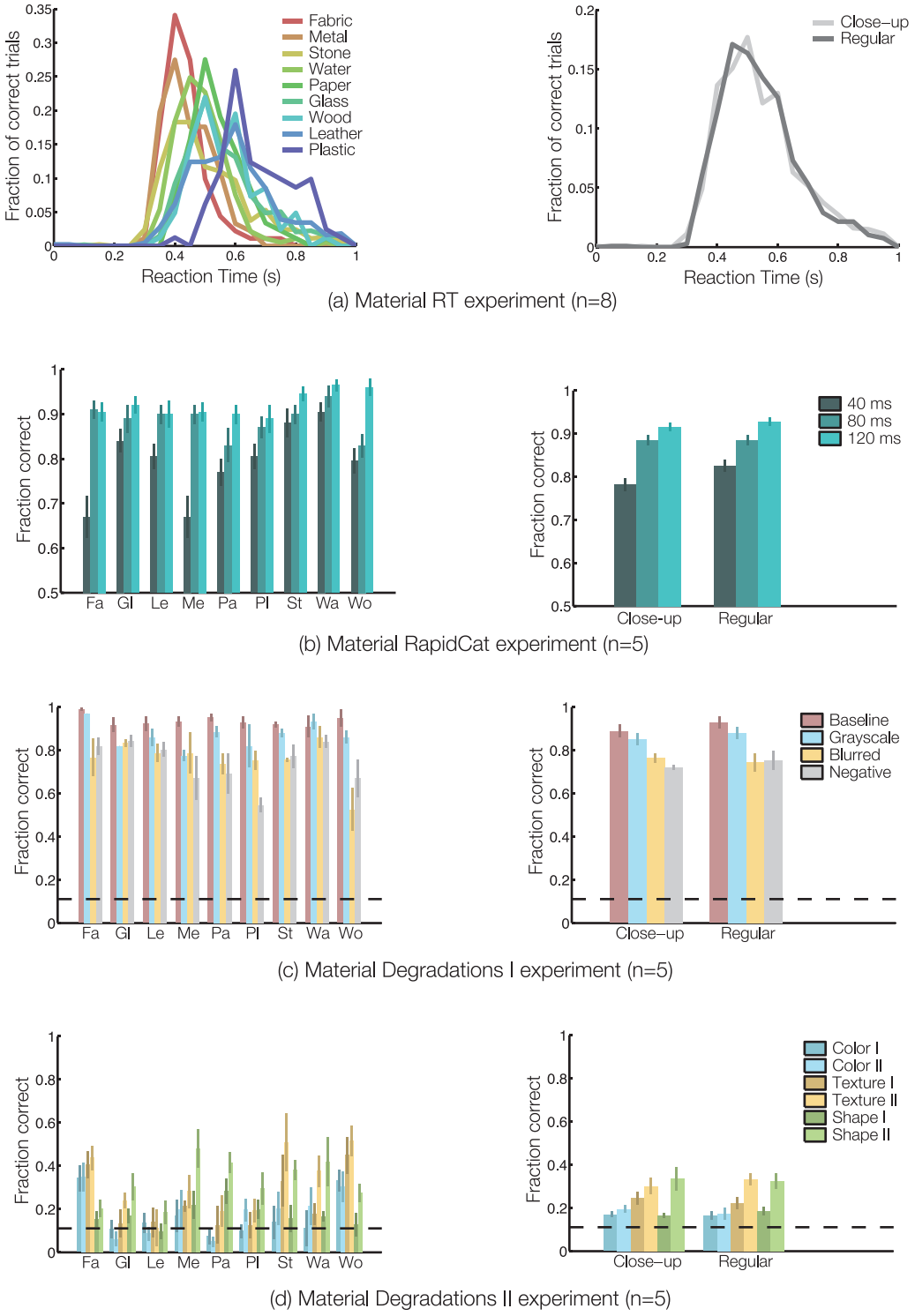



Figure 10. Measuring the influence of material category and view type on material categorization. (a) RT distributions for correct trials, averaged across (left) one to six and (right) eight observers, are shown here for each (left) material category and (right) view type. (b) Accuracy at detecting material categories is shown here for all exposure durations as a function of (left) material category and (right) view type. Chance performance corresponds to 0.5, and error bars represent 1 *SEM* across (left) one to two and (right) five observers. (c–d) Accuracy at nine-way material categorization is shown here for all degradations as a function of (left) material category and (right) view type. Chance performance (0.11) is indicated by dashed black lines, and error bars represent 1 *SEM* across four observers in the Baseline condition, three observers in the Grayscale and Blurred conditions, two observers in the Negative condition, and five observers in all conditions of (d). The influence of material category should be interpreted cautiously as we lack sufficient statistical power (e.g., $n = 1$ for fabric and plastic RT curves). Meanwhile, it is safe to conclude that there were no significant effects of view type.