Journal of Experimental Psychology: General
2016, Vol. 145, No. 1, 95–109

0096-3445/16/$12.00 http://dx.doi.org/10.1037/xge0000130

Mid-Level Perceptual Features Distinguish Objects of Different Real-World Sizes

Bria Long and Talia Konkle
Harvard University

Michael A. Cohen
Massachusetts Institute of Technology

George A. Alvarez
Harvard University

Understanding how perceptual and conceptual representations are connected is a fundamental goal of cognitive science. Here, we focus on a broad conceptual distinction that constrains how we interact with objects—real-world size. Although there appear to be clear perceptual correlates for basic-level categories (apples look like other apples, oranges look like other oranges), the perceptual correlates of broader categorical distinctions are largely unexplored, i.e., do small objects look like other small objects? Because there are many kinds of small objects (e.g., cups, keys), there may be no reliable perceptual features that distinguish them from big objects (e.g., cars, tables). Contrary to this intuition, we demonstrated that big and small objects have reliable perceptual differences that can be extracted by early stages of visual processing. In a series of visual search studies, participants found target objects faster when the distractor objects differed in real-world size. These results held when we broadly sampled big and small objects, when we controlled for low-level features and image statistics, and when we reduced objects to *texforms*—unrecognizable textures that loosely preserve an object's form. However, this effect was absent when we used more basic textures. These results demonstrate that big and small objects have reliably different mid-level perceptual features, and suggest that early perceptual information about broad-category membership may influence downstream object perception, recognition, and categorization processes.

Keywords: object recognition, perceptual and conceptual processing, image statistics, broad category membership, visual search

We can rapidly recognize an incredible number of different objects, effortlessly connecting incoming visual input with high-level conceptual representations, such as an object's identity or category (Grill-Spector & Kanwisher, 2005; Kirchner & Thorpe, 2006). Influential object recognition models posit that this feat is accomplished by extracting a hierarchy of increasingly complex feature representations (e.g., Biederman, 1987; Riesenhuber & Poggio, 1999; Krizhevsky, Sutskever, & Hinton, 2012). Later stages of the hierarchy extract features that are tolerant to identity-preserving transformations, such as changes in location, size, and orientation (DiCarlo & Cox, 2007), thus enabling basic-level object recognition.

Although much research on object recognition has focused on basic-level categorization (e.g., "Is this an apple? Or a hammer?"), less work has focused on how the visual system supports broad conceptual distinctions between objects (e.g., "Is this alive? Is this a tool?"). Intuitively, objects from a particular broad category, such as all manmade objects, come in so many different shapes and sizes that there may be no consistent perceptual features diagnostic of this broad category. Thus, broad object category information might reside only in "semantic" levels of representation.

Alternatively, there may be reliable mid-level perceptual (not semantic) features that differentiate between broad classes of stimuli. Mid-level perceptual features include textural and shape information that preserve local corners, junctions, and contours (e.g., Freeman & Simoncelli, 2011). These features occupy an intermediate status in the visual feature hierarchy, as they are more complex than low-level features like contrast and spatial frequency, but simpler than high-level features, which capture recognizable object parts or entire objects. As such, these features have the potential to carry information about broad category membership.

Bria Long and Talia Konkle, Department of Psychology, Harvard University; Michael A. Cohen, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology; George A. Alvarez, Department of Psychology, Harvard University.

Bria Long, George A. Alvarez, and Talia Konkle developed the study concept. Data collection was conducted by Bria Long and T. Zuluaga. Analyses were conducted by Bria Long. George A. Alvarez and Michael A. Cohen assisted with data analysis and study design. Bria Long drafted the manuscript, and all other authors provided revisions and approved the final draft of the manuscript. J. Freeman provided the code that was used to generate the texform and texture stimuli. This work was supported by a National Science Foundation CAREER grant (BCS-0953730) to George A. Alvarez; U.S. Department of Health and Human Services, National Institutes of Health (NIH) Ruth L. Kirschstein National Research Service Award (NRSA; F32EY022863) to Talia Konkle; and a National Institutes of Health (NIH) Ruth L. Kirschstein National Research Service Award (NRSA; F32EY024483) to Michael A. Cohen.

Correspondence concerning this article should be addressed to Bria Long, Department of Psychology, Harvard University, 33 Kirkland Street, Cambridge, MA, 02140. E-mail: brialong@fas.harvard.edu

Here we focused on one particular broad category distinction, real-world size, and asked if mid-level perceptual features carry information about this distinction. The real-world size of objects has been posited as a core feature of object representation (Konkle & Oliva, 2011), as it constrains which object interactions are appropriate, is automatically accessed during object recognition (Setti, Caramelli, & Borghi, 2009; Sereno, O'Donnell, & Sereno, 2009; Rubinsten & Henik, 2002; Konkle & Oliva, 2012a), and is an organizing property of inanimate object responses in the ventral visual cortex (Konkle & Oliva, 2012b; Konkle & Caramazza, 2013). Further, it has been suggested that objects of different sizes may have different shape and textural properties driven by ecological constraints (Haldane, 1928; Konkle & Oliva, 2012b). However, it is currently unknown whether there are mid-level perceptual features that differentiate the broad classes of big and small objects. If these features exist, they would be useful for speeding basic-level categorization, and generalizing properties to newly learned objects.

If big and small objects are distinguished by mid-level perceptual representations, then a given small object should appear more similar to other small objects than big objects, and vice versa. To explore this possibility, we used a visual search paradigm, as the speed of search depends on how similar the target is to the distractors (Duncan & Humphreys, 1989). Specifically, if big and small objects are highly distinguishable in terms of features that guide visual search, then it should be easier to find a small object target among big objects than among other small objects. We tested this possibility by comparing search efficiency between two kinds of displays: mixed displays, in which targets and distractors differed in real-world size, and uniform displays, in which targets and distractors were of the same real-world size. Critically, in all the displays, the items were presented at the same size on the screen; our key manipulation only varied whether the depicted objects were typically big or small in the world.

To explore the perceptual differences between big versus small objects, we constructed four different stimulus sets. In Experiment 1, we widely sampled from the categories of big and small objects to capture the natural variability in the world (Brunswik, 1955). This experiment serves as an existence proof that there are features that distinguish between the broad categories of big and small objects. We replicated and extended this effect in the second study with a smaller set of images, controlled for a wide range of low-level features, such as aspect ratio, extent, and contour variance. In the critical third experiment, we created a "semantic knockout" stimulus set using texturized stimuli that loosely preserve an object's form yet cannot be recognized at the basic-level (*texforms*). Across all three experiments, we found that search was more efficient when targets and distractors differed in real-world size, even when the items themselves were unrecognizable. In the final experiment, we reduced stimuli even further, preserving only basic texture information, and we no longer found this gain in search efficiency.

Together, these results demonstrate that big and small objects differ in terms of mid-level perceptual features that observers can use to guide their attention during visual search. We propose that these features are extracted early in visual processing, prior to object recognition, and therefore may be used to inform downstream recognition and categorization processes.

Experiment 1: Widely Sampled Stimuli

Here, we asked whether objects of the same real-world size are more perceptually similar to each other than to objects of different real-world sizes, even when all objects are presented at the same physical size on the screen. We first tested a large stimulus set of big and small objects to capture the natural variability in object appearance across many real-world objects.

Method

Participants. Thirteen naive subjects (Harvard students or affiliates) participated. Power analyses on a pilot experiment ($N = 8$) with a slightly different stimulus set and variant of the task indicated that 13 participants would allow detection of a similar-sized effect (75% power, .05 α probability). All participants were 18 to 35 years old and had normal or corrected-to-normal visual acuity.

Procedure. Participants performed a visual search task, in which they searched for a target object among a set of distractors (see Figure 1a). On each trial, the exact target stimulus was previewed and presented centrally for 1000 ms. After 500 ms, a search display with either 3 or 9 items was presented. The items were presented at the same physical size on the screen (5.29° × 5.29°), and were randomly positioned to fall within in a 3 × 4 grid with a ±0.94 degree jitter. The target was always present on the display, and the task was to locate the target as quickly as possible. Participants pressed the space bar as soon as they located the target, after which all items were replaced with Xs and participants clicked on the target's location. This procedure enabled us to verify that participants had actually located the target. In the critical manipulation, distractors were either from the same-size category (uniform trials) or the different-size category (mixed trials) as the target (see Figure 1b). During task instructions, no mention was made concerning the real-world size of the stimuli. Trial types were randomly intermixed throughout the session. Feedback was given after every trial, and accuracy was encouraged, as incorrect responses resulted in a 5-s delay before the next trial could be initiated. There were 10 blocks of 72 trials, yielding 90 trials per condition (each combination of set size, real-world target size, and real-world distractor size). Reaction time (RT) and accuracy were recorded.

Stimuli. Images of big objects and small objects, 200 of each, were taken from Google image search and existing image databases (Brady, Konkle, Alvarez, & Oliva, 2009; Konkle & Oliva, 2012b). All small objects were the size of a desk lamp or smaller; all big objects were the size of a chair or bigger. Big and small objects were equalized across luminance and contrast using the Spectrum, Histogram, and Intensity Normalization (SHINE) Toolbox (Willenbockel et al., 2010) and matched such that they did not differ in average area (approximated as the number of nonwhite pixels) or aspect ratio, two-sample t tests, all $p > .1$. Figure 2 (left panel) shows several example stimuli.

Experimental setup. The experiments were run on an Apple iMac computer (1920 × 1200 pixels, 60 Hz) using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) in MATLAB 2010a (MathWorks, Natick, MA). Participants were positioned approximately 57 cm away from the screen, such that 1 cm on the screen was approximately equal to 1 degree of visual angle. Stimuli had an average image background luminance of 69.7 cd/m^2, and were presented on a uniform gray background (170.0 cd/m^2).

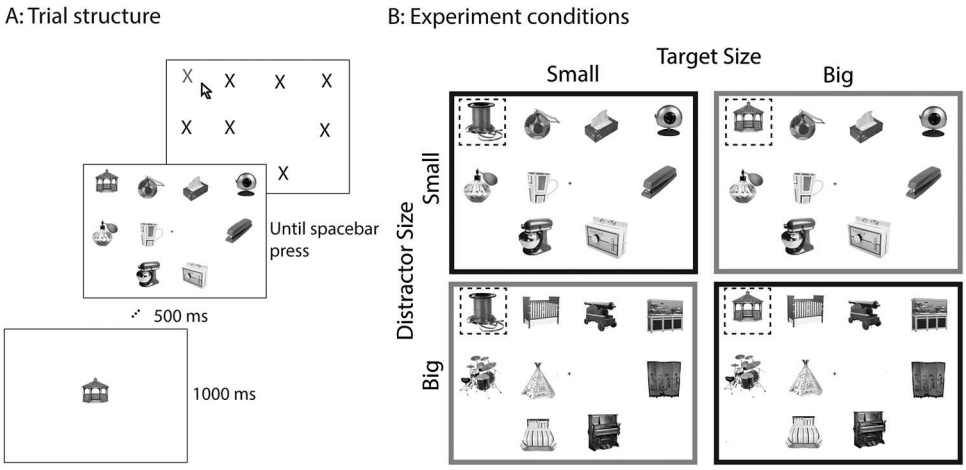

Figure 1. (A) An example trial is shown. A target stimulus is presented for 1000 ms, and then after a 500 ms blank delay, a search display appeared where target and distractor position varied randomly. Participants pressed a spacebar as soon as they found the target, after which all images turned into Xs and they selected the target with the mouse. (B) Example displays are shown for each condition at Set Size 9. The real-world size of the target and distractors was varied to create mixed displays (gray border) and uniform displays (black border). Note that stimuli are shown here in grayscale on a white background for visualization purposes; in the actual experiment, stimuli were always contrast and luminance matched and presented on a gray background.

Outlier removal. RTs were trimmed to exclude trials in which participants incorrectly identified the target or responded in less than 300 ms. We further excluded trials that fell outside 3 *SD*s from the median deviation of the median (Rousseeuw & Croux, 1993), computed separately for each combination of subject, set size, display type, and real-world target size. Overall, 12.5% (*SD* = 4.1%) of the trials were excluded.

Results

The search times for both the mixed and uniform displays are depicted in Figure 2 (right panel). We calculated the search efficiency for each of these conditions, based on the slope of the line relating RT to set size. This slope is a measure of the cost, in RT, for each additional distractor in the display. Thus, steeper slopes indicate less efficient processing. Our main question of interest was whether people search more efficiently in the mixed displays (when the distractors are from the different-size category than the target) than in the uniform displays (when the distractors and target are from the same-size category). The results show that visual search was more efficient in mixed displays than in uniform displays, uniform slope: $M = 50.40$, $SD = 15.10$; mixed slope: $M = 43.58$, $SD = 12.49$, $t(12) = 2.03$, $p = .065$.

This result was also confirmed by a three-way repeated-measures analysis of variance (ANOVA) on RT, with set size (3, 9), real-world target size (big, small), and display type (uniform, mixed) as factors. Participants responded faster when the target was small, $F(1, 12) = 32.5$, $p < .001$, $\eta_p^2 = 0.73$, and when the displays were mixed, $F(1, 12) = 11.9$, $p < .01$, $\eta_p^2 = 0.50$. Most important, the interaction between set size and display type was significant, $F(1, 12) = 5.08$, $p < .05$, $\eta_p^2 = 0.30$, indicating that the increase in RT with additional distractors was reliably lower for mixed trials than for uniform trials. These results demonstrate that there were consistent differences between big and small objects that observers can use to improve visual search performance.

Finally, we also found that this difference in search slopes was greater when the target was a small object than when the target was a big object, 3-way interaction, $F(1, 12) = 33.1$, $p < .001$, $\eta_p^2 = 0.73$. Post hoc tests revealed that search slopes differed between mixed and uniform conditions when the target was a small object, $t(12) = 5.17$, $p < .001$, but did not differ when the target was a big object, $t(12) = -.65$, $p = .53$. Thus, search was most efficient when the target was a small object and distractors were big objects. Search asymmetries are common in visual search tasks (Wolfe, 2001) and suggest asymmetric overlap in object features (e.g., that these small objects have features that separate them from big objects, but that the big objects share many of their features with small objects). Although these asymmetries likely depend on the stimulus set (see Experiments 2–4), they are consistent with the conclusion that small and big objects are distinguished by differences in perceptual features, and could provide insight into how big and small objects overlap in feature space.

The big object category in this experiment contained objects with a very wide range of sizes, from chairs to buildings. This range raised the possibility that only a subset of the biggest objects, namely the buildings, were driving the effects we observed. To test this possibility, we removed any trial in which a building appeared (25 images) as either a target or as a distractor, excluding 36.25% of trials. Eliminating trials in which a building appeared as a big object did not change the pattern of results: Set Size × Display Type interaction, $F(1, 12) = 14.2$, $p < .01$, $\eta_p^2 = 0.54$, suggesting that this effect cannot be attributed to the distinction between buildings and objects.

Experiment 2: Controlled Stimuli

In the first study, we broadly sampled from the set of big and small objects. However, this stimulus set was not controlled for a

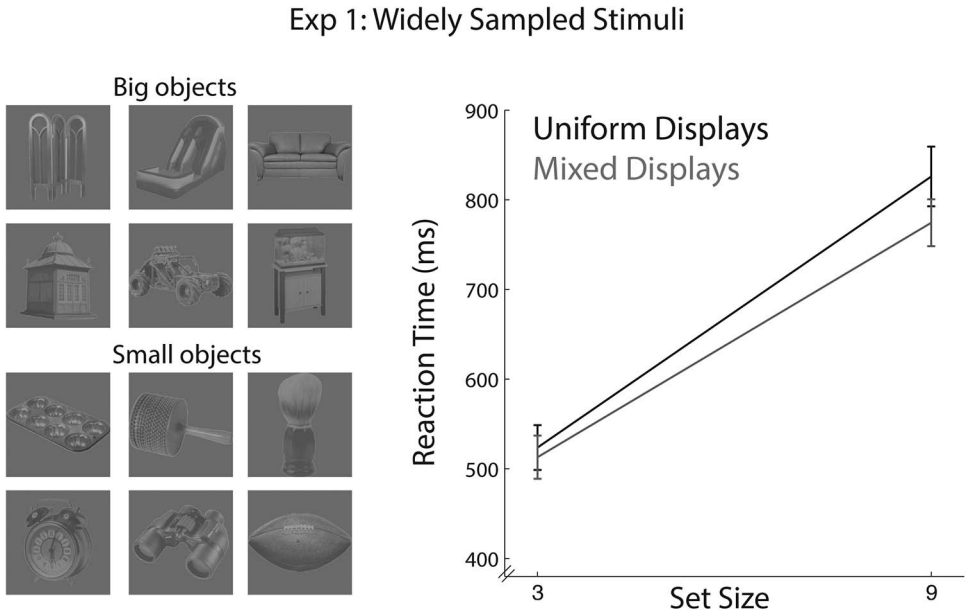

Figure 2. Experimental stimuli and results are shown for Experiment 1. The left panel shows examples of 6 big objects and 6 small objects. The right panel shows reaction time data (ms), plotted as a function of set size. Mixed displays, where target and distractors differed in real-world size, are plotted with gray lines; Uniform displays, where the target and distractors were from the same real-world size, are plotted with black lines. Data are collapsed across the real-world size of the target item. Error bars represent 95% within-subject confidence intervals (Morey, 2008).

number of possible differences between big and small objects that could influence visual search performance, ranging from differences in low-level image statistics to differences in conceptual similarity. In Experiment 2, we selected a highly controlled set of big and small objects that were matched in terms of several low-level properties (e.g., spatial frequency and orientation content, extent, and boundary contour variance), and experience-based properties (e.g., object familiarity and typicality). If any of these factors accounted for the results of Experiment 1, then the difference between mixed and uniform trials should be eliminated with this controlled stimulus set.

Method

Participants. Fourteen naive subjects (Harvard students or affiliates) participated. One participant was excluded for not following task instructions (pressing the response button before the search display appeared). All participants were 18 to 35 years old and had normal or corrected-to-normal visual acuity.

Stimuli. Small objects were chosen to have a canonical orientation (Palmer, Rosch, & Chase, 1981), and buildings were no longer included in the set of big objects. Contour variance was measured by computing the standard deviation of the distance from the centroid of each object (Gonzalez, Woods, & Eddins, 2009) to each point on the objects contour, as previous research has indicated this factor may influence visual search (Naber, Hilger, & Einhäuser, 2012). Object extent was taken as the ratio of the area of the object to its rectangular bounding box (Gonzalez et al., 2009). We also measured image area (percentage of nonwhite pixels within a square frame) and aspect ratio (max height/max width in the picture plane). Finally, an Amazon Mechanical Turk (mTurk) experiment was conducted to obtain typicality and familiarity rankings for each object on a 4-point Likert scale.

Sixty final objects (30 big objects, 30 small objects) were chosen so that the two sets did not differ on any of the above features, two-sample t tests, all $p > .4$. These objects and backgrounds were then matched in terms of their intensity histograms (luminance and contrast) and power spectra (power at each orientation and spatial frequency) using the SHINE Toolbox, (Willenbockel et al., 2010). These images were set to an average luminance of 95.8 cd/m^2, presented on a lighter gray background (170.0 cd/m^2) to ensure they segmented from the background easily. Example stimuli are shown in Figure 3 (left panel).

Given the smaller stimulus set, trials were counterbalanced so that each object appeared as a target equally often in all conditions. All other procedures were the same as in Experiment 1.

Results

RTs were trimmed following the same procedure as in Experiment 1, excluding 9.95% of the trials ($SD = 4.12\%$). The results of Experiment 2 are plotted in Figure 3 (right panel). Overall, we observed the same pattern of results as in Experiment 1, even with this highly controlled stimulus set. That is, visual search was more efficient for mixed displays, when targets and distractors differed in real-world size, relative to uniform displays, when targets and distractors were of the same real-world size, uniform slope: $M = 78.26$, $SD = 21.11$, mixed slope: M = 65.41, $SD = 19.44$, $t(12) = 4.75$, $p < .001$.

These observations were confirmed with a three-way repeated-measures ANOVA. Observers responded faster when the target was a small object, $F(1, 12) = 25.6$, $p < .001$, $\eta_p^2 = 0.68$, and on

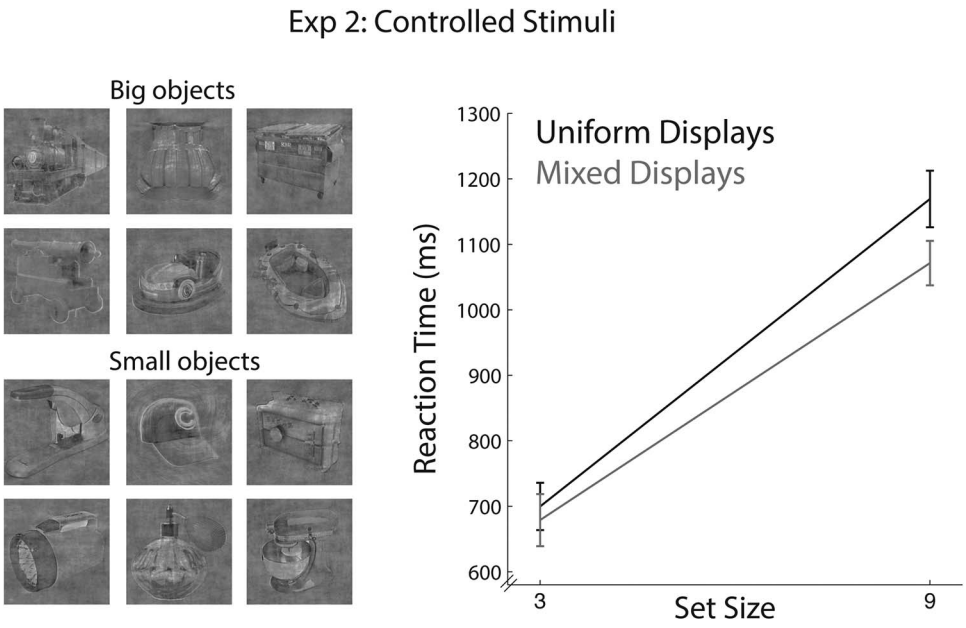

Figure 3. Experimental stimuli and results are shown for Experiment 2. The left panel shows examples of 6 big objects and 6 small objects. The right panel shows reaction time data (ms), plotted as a function of set size. Mixed displays, where the target and distractors differed in real-world size, are plotted with gray lines; Uniform displays, where the target and distractors were from the same real-world size, are plotted with black lines. Data are collapsed across the real-world size of the target item. Error bars represent 95% within-subject confidence intervals (Morey, 2008).

mixed displays, $F(1, 12) = 35.2$, $p < .001$, $\eta_p^2 = 0.75$. Observers were again more efficient at searching in the mixed relative to uniform displays, $F(1, 12) = 23.7$, $p < .001$, $\eta_p^2 = 0.66$. Unlike Experiment 1, this effect was not modulated by whether the target was a big or small object, $F(1, 12) = 1.9$, $p = .19$.

Thus, the ability to find a target faster when distractors are of a different real-world size (mixed displays) does not appear to be driven by low-level image and basic contour features, which were matched in this stimulus set. Given the reduced number of stimuli in Experiment 2, these effects were also confirmed using linear mixed-effects models to ensure that the results generalized across items and participants (see Appendix, Section I).

Experiment 3: Texture-Form Stimuli

The previous results demonstrated that there are robust differences between big and small objects that can be used to facilitate visual search. These differences cannot be explained by simple low-level image features, and as such, points to a difference in mid-level features as a guiding factor (Treisman & Gelade, 1980; Itti & Koch, 2000; Wolfe, 1994; Duncan & Humphreys, 1989). However, in both experiments, big and small objects were also recognizable and therefore also differed in their semantic content. Thus, the search efficiency differences we observed in the previous experiments could be due to semantic interference (Telling, Kumar, Meyer, & Humphreys, 2010; Moores, Laiti, & Chelazzi, 2003). On such an account, similar-sized objects might impede search performance differentially on uniform displays because they are more semantically related to each other.

To examine this possibility, we generated a "semantic knockout" stimulus set by creating images of big and small objects that loosely preserve an object's form and feature differences between objects, but which are not recognizable. We used a texture synthesis algorithm to create stimuli that match the first- and second-order statistics of a target image within a series of receptive field-like pooling windows (Freeman & Simoncelli, 2011). By pooling image statistics within separate windows, these stimuli capture texture in a way that preserves the coarse form of the object (*texforms*). Assuming these texforms preserve the features that guide visual search (Rosenholtz, Huang, & Ehinger, 2012; Alexander, Schmidt, & Zelinsky, 2014), these stimuli should generate the same pattern of results as the original objects. In contrast, the semantic interference account predicts that we should no longer find a difference in search efficiency because the stimuli are unrecognizable.

Method

Participants. Participants were 16, naive Harvard students or affiliates, aged 18 to 35 years. Three participants did not complete the experiment; their data were never analyzed. All participants had normal or corrected-to-normal visual acuity.

Stimuli. Synthesized versions of the big and small objects were generated by initializing Gaussian white noise images and iteratively adjusting them (using a variant of gradient descent) to conform to the modeled parameters of the original image (Freeman & Simoncelli, 2011, see Appendix, Section II). This produced images that were nearly always unrecognizable, while preserving

mid-level image statistics in each pooling window. Amazon Mechanical Turk norming studies were run to select a subset of 60 images for which the original objects were unidentifiable, even when answers were coded generously (e.g., "stove" was accepted as a correct response for "jukebox" because it is the same sized object with a similar shape). In our final subset of 60 items, the average identification accuracy was 2.83%, $SD = 4.02\%$ ($N = 30$). Example stimuli are shown in Figure 4 (left panel).

Procedure. Search displays with 3 or 8 items were presented in a circle around fixation at 7.4 degrees of eccentricity, and subtended 5.1 × 5.1 degrees of visual angle (192 × 192 pixels). As eccentricity is a parameter in the texture synthesis algorithm, we generated one set of texforms at a single eccentricity. Each texform stimulus was presented inside of a black outline to ensure that it was clearly visible from the background. The overall luminance of the texforms and background were matched ($M = 77.6$ cd/m^2). All other aspects of the experimental design were identical to Experiment 2.

To ensure that the texforms were not recognizable for the participants who completed the visual search task, observers were presented with two follow-up tasks at the end of the visual search experiment. First, they were asked: "In this experiment, there were two groups of images. On some trials, the image you were looking for was from a different category than the other images, and on other trials, all of images were from the same category. Please guess what the two categories could be." The choices were "1–Animals/Objects, 2–Tools/Non-Tools, 3–Natural/Unnatural, 4–Edible/Non-Edible, 5–Big/Small, 6–Familiar/Unfamiliar, 7–I have no idea." Second, subjects completed an unspeeded, randomized questionnaire in which they were asked to guess the identity of each texform.

Results

RTs were trimmed with the same procedure, excluding 14.9% ($SD = 3.9\%$) of the trials. Overall, we found the same pattern of results with unrecognizable texform stimuli as with intact objects (see Figure 4). That is, visual search was more efficient in mixed displays than in uniform displays, uniform slope $M = 91.14$, $SD = 32.72$; mixed slope $M = 78.69$, $SD = 35.69$, $t(12) = 3.27$, $p < .01$. A repeated-measures ANOVA confirmed that participants responded faster on mixed trials, $F(1, 12) = 20.2$, $p < .001$, $\eta_p^2 = 0.63$ and searched more efficiently in mixed displays, $F(1, 12) = 8.68$, $p = .01$, $\eta_p^2 = 0.42$. We also confirmed these results using linear mixed-effects models, which showed that the improved efficiency for mixed displays generalized across both items and participants (see Appendix, Section III).

The efficiency advantage on mixed trials did not differ depending on whether the target was a small object texform or a big object texform, 3-way interaction, $F(1, 12) = 0.80$, $p = .39$, $\eta_p^2 = 0.06$. Although the interaction was not significant, numerically the effect appears bigger for big object targets than small object targets, which is opposite to the trends observed in Experiment 1 and 2. To determine whether these differences were consistent, we ran two replication experiments, which again had no statistically significant interactions, although the same opposing assymetries were present (see Appendix, Sections IV, V). Thus, the results suggest there is a weak but potentially consistent difference between texforms and their original images: the feature overlap betweeen big versus small texforms maybe be subtly different than the feature overlap between big versus small objects.

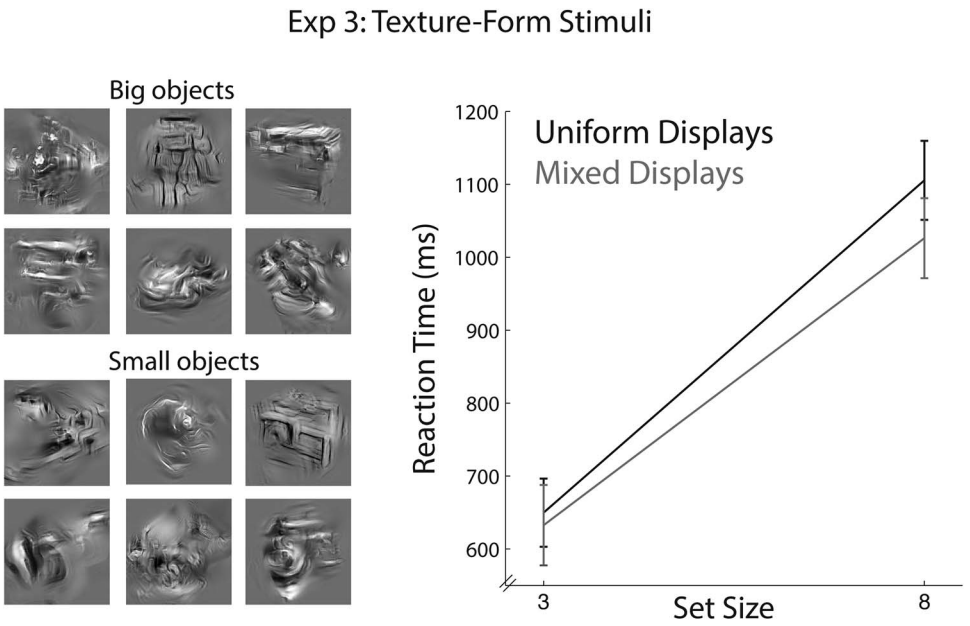

Figure 4. Experimental stimuli and results are shown for Experiment 3. The left panel shows examples of texforms generated from 6 big objects and 6 small objects, corresponding to the original objects in Figure 3. The right panel shows reaction time data (ms), plotted as a function of set size. Mixed displays, where target and distractors differed in real-world size, are plotted with gray lines; uniform displays, where the target and distractors were from the same real-world size, are plotted with black lines. Data are collapsed across the real-world size of the target item. Error bars represent 95% within-subject confidence intervals (Morey, 2008).

Follow-up tasks. The results of our follow up tasks suggest that people had little to no awareness of the relevant categorical distinction (big vs. small objects). The most common response was animate versus inanimate, and no subject guessed that real-world size was the relevant distinction. In addition, none of the participants accurately guessed the identity of more than one of the 60 texforms. When coding liberally (e.g., accepting responses somewhat similar to the original object), participants guessed an average of 3.0 out of 60 objects ($SD = 1.68$). Thus, it is unlikely that explicit categorization or identification enabled participants to find target texforms faster in mixed-size trials.

The purpose of using these texforms was to preserve perceptual differences while preventing explicit object recognition. One concern is that participants may have been consistently misidentifying the texforms as another object within the correct size category. To test whether this was true, we asked each participant who performed the search task to guess what each texform was. Participant's responses were then coded liberally for any size information (e.g., "a cheetah's face" and "microbes" were coded as small) when possible, though some participants refused to respond ("I don't know") or gave answers that were uncodable for size (e.g., "something burning," "swirl," "mayan ritual"). These uncodable responses occurred on 24.1% of the trials and were counted as incorrect. Overall, participants were not above chance in naming objects that were the right size, $M = 44.1\%$, $SD = 10.4\%$, $t(12) = -2.0$, $p = .07$. This analysis suggests it is unlikely that participants are able to search more efficiently on mixed-sized displays by recognizing texforms as objects and subsequently leveraging semantic information about their real-world size.

Additional control task. Our follow up tasks suggest that participants could not explicitly recognize the original objects used to generate the texforms. However, it remains possible that these texforms implicitly activated real-world size knowledge, and that this knowledge could lead to implicit semantic influences on visual search. One method for detecting implicit knowledge activation is to use a forced-choice task (e.g., Turk-Browne, Jungé, & Scholl, 2005). Thus, in the above norming study, participants also completed a forced choice task, guessing the real-world size of each texform.

First, participants in our norming study were asked to guess the real-world size of the texforms using a continuous scale from 1 (*as small as a key*) to 7 (*as large as a building*). These responses were binarized and coded for accuracy according to whether the original object was small or big. Participants chose the correct real-world size category of the original objects slightly more often than chance, $M = 59.8\%$ correct, $SD = 6.3\%$, $t(29) = 8.59$, $p < .0001$, see Appendix, Section VI.

We next split our visual search data into two halves as a function of how accurately the target texform was classified as big or small in the norming study. In the top split of the data, target texforms were classified as big or small at a rate above chance, $M = 76.4\%$, $SD = 10.9\%$, two-tailed t test against chance (50%), $t(29) = 13.29$, $p < .0001$, and in the bottom split of the data at a rate below chance, $M = 43.2\%$, $SD = 16.25\%$, $t(29) = -2.28$, $p < .05$. We conducted a four-way ANOVA with factors of set size, display type, real-world target size and data split. If participants were implicitly recognizing the size of the object from the texforms and using this abstract knowledge to guide visual search, then we should see accentuated effects in the top split (and potentially reversed effects in the bottom split).

However, we observed the same pattern of results in both halves of the data: there was no difference in overall RT, $F(1, 12) = .97$, $p = .34$, $\eta_p^2 = 0.08$ and no difference in how efficiently participants found targets on mixed versus uniform displays, $F(1, 12) = 0.92$ $p = .36$, $\eta_p^2 = 0.07$. This analysis suggests that it is unlikely that participants were using implicit knowledge of real-world size to modulate their search efficiency.

Experiment 4: Texture Stimuli

To understand more clearly what critical visual information could distinguish between big and small objects, in the final experiment we generated textures that preserved the same image statistics as those used in Experiment 3, but distributed them across the entire image. That is, the image features were synthesized over one pooling window that included each entire object (Portilla & Simoncelli, 2000; Balas, 2006, see Appendix, Section II). The resulting images do not preserve object form and have little to no perceptible contours (Figure 5, left panel), which can be easily seen by comparing these stimuli with those from Experiment 3 (see Figure 4, left panel).

Method

Participants. Thirteen Harvard affiliates or students again participated in Experiment 4. All participants were 18 to 35 years old and had normal or corrected-to-normal visual acuity.

Procedure. All procedures were identical to Experiment 3, except the stimuli.

Stimuli. Textures were generated using the same algorithm in Experiment 3, except that white noise was coerced to have the same statistics as the original image pooled across the entire image. See Appendix, Section II for details.

Results

RTs were trimmed using the same procedure as the previous experiments, $M = 17.02\%$, $SD = 6.52\%$. Unlike the previous experiments, we found that visual search was not more efficient in mixed displays than in uniform displays, uniform slope: $M = 112.65$, $SD = 30.25$, mixed slope: $M = 106.31$, $SD = 21.92$, $t(12) = 1.42$, $p = .18$, see Figure 5. A repeated-measures ANOVA confirmed that participants were not faster at finding textures when distractors were generated from objects of a different size category, $F(1, 12) = 1.64$, $p = .23$, $\eta_p^2 = .12$, and this effect did not interact with the number of distractors, $F(1, 12) = 1.19$, $p = .30$, $\eta_p^2 = .09$.

We did not observe the same search advantage when targets and distractors were from different size categories, even though these textures were generated from the exact same images as the stimuli in Experiment 3—the numerical trend was in the same direction, but the difference was not reliable. Thus, it appears that these textures preserve less of the critical feature differences between big and small objects than the texforms used in Experiment 3. Taken together, these results suggest that the spatial organization of these texture statistics is important for capturing the differences between big and small objects.

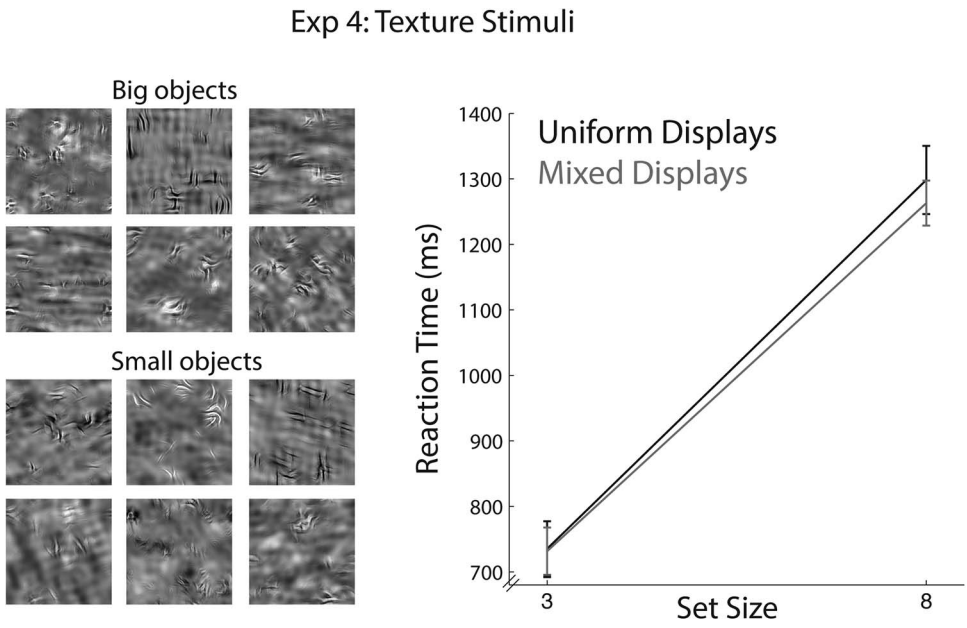

Figure 5. Experimental stimuli and results are shown for Experiment 4. The left panel shows examples of textures generated from six big objects and six small objects, corresponding to the original objects in Figure 3. The right panel shows reaction time data (ms), plotted as a function of set size. Mixed displays, where target and distractors differed in real-world size, are plotted with gray lines; Uniform displays, where the target and distractors were from the same real-world size, are plotted with black lines. Data are collapsed across the real-world size of the target item. Error bars represent 95% within-subject confidence intervals (Morey, 2008).

General Discussion

Here, we explored whether big and small objects have reliable perceptual differences that can be extracted by early stages of visual processing, focusing on real-world object size as a case study of broad category membership. We found that visual search was more efficient when the target and distractors differed in real-world size, both when exemplars were widely sampled (Experiment 1), and when they were more controlled (Experiment 2). Critically, when we reduced the objects to textures that preserved some form information, we still found a gain in search efficiency (Experiment 3), but when we reduced the objects further to textures without form information, this visual search effect was absent (Experiment 4). Together, these results demonstrate that big objects and small objects differ in mid-level perceptual features, which are used to guide attention in visual search. In the following sections, we discuss the nature of these mid-level perceptual features, how the visual system might develop sensitivity to these features, and the implications of these findings for models of object recognition and categorization.

Features of Big Versus Small Objects

The present results demonstrate that big and small objects classes are distinguishable by mid-level perceptual features—but what exactly is the nature of these feature differences? Based on the image synthesis model we used (Freeman & Simoncelli, 2011), we know that these features are related to differences in local texture and contour statistics, including the presence of junctions, corners, and parallel lines. Further, we know that these features may contain coarse shape information, because the basic textures—which did not preserve any coarse shape information—did not generate a reliable category search advantage. These findings suggest that the key differences are in how texture statistics are spatially organized.

Although understanding exactly which features embedded within the model parameters separate big and small object classes is beyond the scope of the current paper, it is nevertheless useful to consider some intuitive possibilities. One possibility is that the relevant mid-level perceptual features are related to a difference in perceived curvature. For example, there are structural limitations on the shapes that big objects can have (Gordan, 1981): Big objects must withstand gravity and tend to have more rectilinear forms, whereas small objects can be either boxy or curvy (e.g., notebooks, basketballs). Further, neural regions involved in processing objects are sensitive to differences in curvature, particularly along a boxy to curvy axis (e.g., Srihasam, Vincent, & Livingstone, 2014; Brincat & Connor, 2004).

Consistent with this idea, participants rated big objects as boxier and small objects as curvier, for all four of our stimulus sets, including both the texforms and the basic textures (see Appendix, Section VII). However, this boxy-curvy dimension is only one possible dimension within a large feature space: because mid-level features represent combinations of simpler features (e.g., a 'corner' is a particular combination of two lines), the possible set of mid-level features is unconstrained. Further research will be required to create a vocabulary for describing mid-level perceptual features, and to parse the space of mid-level features into psychologically meaningful dimensions. Critically, the primary goal of the present work was to demonstrate that mid-level perceptual features differences exist between big and small objects.

Finally, across the experiments we found different patterns of search asymmetries, which may inform our intuitions about the feature spaces of big and small objects. When stimuli were widely sampled (Experiment 1), we found that searching for a small object among big objects was more efficient than searching for a big object among small objects. This suggests that small objects are more different from big objects than big objects are from small objects. At first blush this seems illogical, but such asymmetries in similarity can arise when the features of one category are a partial subset of the features of the other category (Tversky & Gati, 1978). On this account, the features of big objects in Experiment 1 are a subset of the features of small objects, but small objects have some features that are uncommon among big objects (e.g., both small and big objects can be boxy, but more small objects are curvy). However, this particular asymmetric relationship may not be a general property of big and small objects, as there were no reliable asymmetries based on whether the target was a big or a small object when stimuli were tightly controlled (Experiment 2), reduced to texforms (Experiment 3), or in our subsequent replications of these two experiments (see Appendix, Sections IV and V). Future research will help understand the degree to which there may be a true asymmetry in the feature spaces of big versus small objects.

How Do We Develop Sensitivity to These Features?

Although real-world size is a broad distinction that spans many basic-level categories, big and small objects seem to have reliably different mid-level perceptual features. There are two main perspectives for how sensitivity to these perceptual features may arise.

One possibility is that our visual system is innately predisposed to be sensitive to differences in certain perceptual features. For example, recent evidence posits the existence of a protomap of curvature along the ventral visual stream (Srihasam et al., 2014). On this account, our perceptual systems are naturally wired to discriminate the broad categories of big and small objects.

Alternatively, experience-dependent tuning mechanisms may detect perceptual regularities for conceptually relevant dimensions (Kohonen, 1982; Polk & Farah, 1995), including (but not limited to) the dimension of real-world size. Indeed, previous work suggests that the mere act of categorizing objects together may cause them to become perceptually similar (Goldstone, 1994), creates task-specific features (Schyns & Rodet, 1997), and causes neural representations in visual cortex to become less discriminable (Folstein, Palmeri, & Gauthier, 2013). On this account, these perceptual differences could become psychologically salient due to extensive experience perceiving and interacting with objects at different real-world sizes.

Implications for Models of Object Recognition and Categorization

Regardless of the ultimate cause for the visual system's sensitivity to perceptual differences between big and small objects, these findings raise the intriguing possibility that earlier stages of visual processing can inform high-level processes about what broad category an object may belong to, rendering object recognition and categorization more efficient. Here we propose that such mid-level features provide information about the broad superordinate category of the object. We use the term mid-level facilitation to refer to the idea that early sensitivity to these kinds of mid-level features may facilitate downstream, higher-level processes like object recognition and action preparation by constraining the possible basic-level identities considered by the visual system.

Although we focused on real-world size in the present study, it is likely that other broad categories are distinguished by mid-level perceptual features. Broad distinctions that are behaviorally salient and have a plausible basis in evolutionary history may be particularly good candidates, whereas arbitrary distinctions may not. Tools, for example, may share similar mid-level features, which allow them to be easily grasped compared with other nonmanipulable objects. Mid-level features may also distinguish animate entities from inanimate objects (Levin, Takarae, Miner, & Keil, 2001; Long, Störmer, & Alvarez, 2014), another core dimension of object representation. Although it is difficult to make a priori predictions for all possible broad category distinctions, the current study introduces an approach for investigating the perceptual correlates of broad conceptual categories.

Conclusion

Using a visual search task, we found that objects appear more similar to other objects of the same real-world size than objects of a different real-world size (when all objects are the same physical size on the screen). These findings show that the visual system is sensitive to mid-level perceptual features that distinguish big and small objects. Because such features can be extracted by early stages of the visual system, these results suggest that early stages of perceptual processing can facilitate broad-category level processing. We propose that examining the intrinsic, statistical dependency between broad conceptual distinctions and perceptual features will advance a more integrated understanding of how we perceive, recognize, and categorize objects.

References

Alexander, R. G., Schmidt, J., & Zelinsky, G. J. (2014). Are summary statistics enough? Evidence for the importance of shape in guiding visual search. *Visual Cognition, 22,* 595–609. http://dx.doi.org/10.1080/13506285.2014.890989

Baayen, R. H. (2009). languageR: Data sets and functions with "Analyzing Linguistic Data: A practical introduction to statistics". R package version 0.955.

Balas, B. J. (2006). Texture synthesis and perception: Using computational models to study texture representations in the human visual system. *Vision Research, 46,* 299–309. http://dx.doi.org/10.1016/j.visres.2005.04.013

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random-effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68,* 255–278. http://dx.doi.org/10.1016/j.jml.2012.11.001

Bates, D. M., & Maechler, M. (2009). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375–32.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94,* 115–147. http://dx.doi.org/10.1037/0033-295X.94.2.115

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences, 105,* 14325–14329.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 433–436. http://dx.doi.org/10.1163/156856897X00357

Brincat, S. L., & Connor, C. E. (2004). Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nature Neuroscience, 7,* 880–886. http://dx.doi.org/10.1038/nn1278

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review, 62,* 193–217. http://dx.doi.org/10.1037/h0047470

DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in Cognitive Sciences, 11,* 333–341. http://dx.doi.org/10.1016/j.tics.2007.06.010

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96,* 433–458. http://dx.doi.org/10.1037/0033-295X.96.3.433

Folstein, J. R., Palmeri, T. J., & Gauthier, I. (2013). Category learning increases discriminability of relevant object dimensions in visual cortex. *Cerebral Cortex, 23,* 814–823. http://dx.doi.org/10.1093/cercor/bhs067

Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nature Neuroscience, 14,* 1195–1201. http://dx.doi.org/10.1038/nn.2889

Goldstone, R. L. (1994). The role of similarity in categorization: Providing a groundwork. *Cognition, 52,* 125–157. http://dx.doi.org/10.1016/0010-0277(94)90065-5

Gonzalez, R. C., Woods, R. E., & Eddins, S. L. (2009). *Digital image processing using MATLAB* (2nd ed.). Knoxville, TN: Gatesmark Publishing.

Gordan, J. E. (1981). *Structures: Or why things don't fall down*. Cambridge, MA: Da Capo Press.

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

Haldane, J. B. S. (2008). On being the right size. In R. Dawkins, *The Oxford book of modern science writing* (pp. 53–58). Oxford, UK: Oxford University Press. Original published 1928.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40,* 1489–1506. http://dx.doi.org/10.1016/S0042-6989(99)00163-7

Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research, 46,* 1762–1776. http://dx.doi.org/10.1016/j.visres.2005.10.002

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics, 43,* 59–69. http://dx.doi.org/10.1007/BF00337288

Konkle, T., & Caramazza, A. (2013). Tripartite organization of the ventral stream by animacy and object size. *The Journal of Neuroscience, 33,* 10235–10242. http://dx.doi.org/10.1523/JNEUROSCI.0983-13.2013

Konkle, T., & Oliva, A. (2011). Canonical visual size for real-world objects. *Journal of Experimental Psychology: Human Perception and Performance, 37,* 23–37. http://dx.doi.org/10.1037/a0020413

Konkle, T., & Oliva, A. (2012a). A familiar-size Stroop effect: Real-world size is an automatic property of object representation. *Journal of Experimental Psychology: Human Perception and Performance, 38,* 561–569. http://dx.doi.org/10.1037/a0028294

Konkle, T., & Oliva, A. (2012b). A real-world size organization of object responses in occipitotemporal cortex. *Neuron, 74,* 1114–1124. http://dx.doi.org/10.1016/j.neuron.2012.04.036

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems, 25,* 1097–1105. Retrieved from http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

Levin, D. T., Takarae, Y., Miner, A. G., & Keil, F. (2001). Efficient visual search by category: Specifying the features that mark the difference between artifacts and animals in preattentive vision. *Perception & Psychophysics, 63,* 676–697. http://dx.doi.org/10.3758/BF03194429

Long, B. L., Störmer, V. S., & Alvarez, G. A. (2014). Rapid extraction of category-specific shape statistics: Evidence from event-related potentials [Abstract]. *Journal of Vision, 14,* 907. http://dx.doi.org/10.1167/14.10.907

Moores, E., Laiti, L., & Chelazzi, L. (2003). Associative knowledge controls deployment of visual selective attention. *Nature Neuroscience, 6,* 182–189. http://dx.doi.org/10.1038/nn996

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology, 4,* 61–64.

Naber, M., Hilger, M., & Einhäuser, W. (2012). Animal detection and identification in natural scenes: Image statistics and emotional valence. *Journal of Vision, 12,* 1–24. http://dx.doi.org/10.1167/12.1.25

Palmer, S., Rosch, E., & Chase, P. (1981). Canonical perspective and the perception of objects. In J. Long & A. Baddeley (Eds.), *Attention and performance: IX* (pp. 135–151). Hillsdale, NJ: Erlbaum.

Pelli, D. G. (1997). The Video Toolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision 10,* 437–442.

Polk, T. A., & Farah, M. J. (1995). Late experience alters vision. *Nature, 376,* 648–649. http://dx.doi.org/10.1038/376648a0

Portilla, J., & Simoncelli, E. P. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision, 40,* 49–70.

R Development Core Team. (2008). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2,* 1019–1025. http://dx.doi.org/10.1038/14819

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8,* 382–439. http://dx.doi.org/10.1016/0010-0285(76)90013-X

Rosenholtz, R., Huang, J., & Ehinger, K. A. (2012). Rethinking the role of top-down attention in vision: Effects attributable to a lossy representation in peripheral vision. *Frontiers in Psychology, 3,* 13. http://dx.doi.org/10.3389/fpsyg.2012.00013

Rousseeuw, P. J., & Croux, C. (1993). Alternatives to the median absolute deviation. *Journal of the American Statistical Association, 88,* 1273–1283. http://dx.doi.org/10.1080/01621459.1993.10476408

Rubinsten, O., & Henik, A. (2002). Is an ant larger than a lion? *Acta Psychologica, 111,* 141–154. http://dx.doi.org/10.1016/S0001-6918(02)00047-1

Schyns, P. G., & Rodet, L. (1997). Categorization creates functional features. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 681–696. http://dx.doi.org/10.1037/0278-7393.23.3.681

Sereno, S. C., O'Donnell, P. J., & Sereno, M. E. (2009). Size matters: Bigger is faster. *The Quarterly Journal of Experimental Psychology, 62,* 1115–1122. http://dx.doi.org/10.1080/17470210802618900

Setti, A., Caramelli, N., & Borghi, A. M. (2009). Conceptual information about size of objects in nouns. *European Journal of Cognitive Psychology, 21,* 1022–1044. http://dx.doi.org/10.1080/09541440802469499

Simoncelli, E. P., & Freeman, W. T. (1995, October). The steerable pyramid: A flexible architecture for multi-scale derivative computation. *International Conference on Image Processing, 1995,* 3444. http://dx.doi.org/10.1109/ICIP.1995.537667

Srihasam, K., Vincent, J. L., & Livingstone, M. S. (2014). Novel domain formation reveals proto-architecture in inferotemporal cortex. *Nature Neuroscience, 17,* 1776–1783. http://dx.doi.org/10.1038/nn.3855

Telling, A. L., Kumar, S., Meyer, A. S., & Humphreys, G. W. (2010). Electrophysiological evidence of semantic interference in visual search. *Journal of Cognitive Neuroscience, 22,* 2212–2225. http://dx.doi.org/10.1162/jocn.2009.21348

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12,* 97–136. http://dx.doi.org/10.1016/0010-0285(80)90005-5

Turk-Browne, N. B., Jungé, J., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General, 134,* 552–564. http://dx.doi.org/10.1037/0096-3445.134.4.552

Tversky, A., & Gati, I. (1978). Studies of similarity. *Cognition and categorization, 1,* 79–98.

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods, 42,* 671–684. http://dx.doi.org/10.3758/BRM.42.3.671

Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review, 1,* 202–238. http://dx.doi.org/10.3758/BF03200774

Wolfe, J. M. (2001). Asymmetries in visual search: An introduction. *Perception & Psychophysics, 63,* 381–389. http://dx.doi.org/10.3758/BF03194406

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience, 5,* 495–501. http://dx.doi.org/10.1038/nrn1411

Appendix

Stimuli Details, Replications, & Supplemental Analyses

I. Experiment 2 (Controlled Stimuli): Linear Mixed-Effects Modeling

Given that we used a relatively small set of items (60 total) in Experiment 2, it was important to test whether these results generalize across items. We used a linear mixed-effects model to test for fixed effects of set size, display type, and their interaction while simultaneously generalizing across individual subjects and items. Specifically, we modeled log RT as a function of set size and display type, including random effects of set size, display type, and their interaction for both subjects and items on the intercept and the slope terms of the model—the maximal random-effects structure justified by our experimental design (Barr et al., 2013). The models were implemented using R (R Development Core Team, 2008) and the R packages lme4 (Bates & Maechler, 2009) and language R (Baayen, 2009).

We tested for significant effects by performing likelihood-ratio tests, comparing a model with the set size by display type interaction as a fixed effect to another model without it, but which was otherwise identical, including the same exact random-effects structure (Barr et al., 2013). Models were fit using full maximum-likelihood estimation to facilitate comparison between models. Comparing these two models revealed that the RT by set size slope was significantly lower on mixed trials, $\chi^2(1) = 4.95$, $p = .025$. Thus, we can conclude that search was more efficient on mixed trials, and that this effect generalized across participants and items.

II. Experiments 3 and 4: Stimulus Generation Details

The model measures basic features (lines/edges of different orientations and sizes), and correlations between basic features

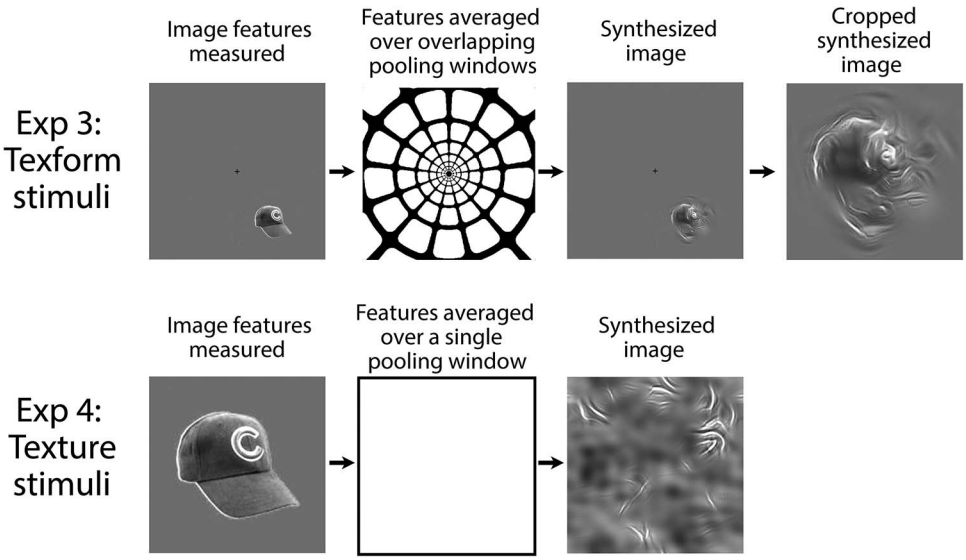

Figure A1. Schematic of stimuli generation procedure for the stimuli used in Experiments 3 and 4. In Experiment 3, we generated texforms by preserving the spatial arrangement of image features using the algorithm developed by Freeman and Simoncelli (2011). Experiment 4 used the same algorithm, except image features were pooled over the entire image.

(*Appendix continues*)

across space and size (useful for detecting corners and parallel lines). First, the model decomposes images using a steerable pyramid in the Fourier domain (Simoncelli & Freeman, 1995). Steerable pyramid models decompose an image using a bank of wavelet filters at multiple scales and orientations. The model first splits the image into different spatial frequency bands. In this implementation, these subbands were scaled to four different sizes, and the degree to which four orientations ranges are present in those scaled images was measured, creating 16 different filters. This created an overcomplete representation of the image that contained information about both the frequency and location of orientation information.

In the midventral model, developed by Freeman and Simoncelli (2011), the responses from these filters are correlated with each other, as well as with responses between different scale filters and between different orientation filters. The mid-level model contains several features: (a) marginal pixel statistics over the entire image and within pooling regions, (b) features analogous to the response of V1 simple cells and V1 complex cells for each combination of spatial frequency and orientation at each location, (c) cross-correlations of these complex cell responses across different scales and orientations, (d) spectral statistics, or features derived from products of V1 simple cells that are sensitive to changes in phase. Coarsely, these correspond to sensitivity to luminance, contrast, spatial frequency, sharp line changes, contours, edges, junctions, corners, and shading.

These feature representations are then down-sampled, that is, averaged across portions of the image dubbed "pooling regions." The size of these pooling regions is extremely important to the resulting synthesized image. These are derived from a model of the receptive fields in V2 (Freeman & Simoncelli, 2011). To create our texform stimuli, we choose pooling regions that were slightly different than those used by Freeman & Simoncelli to create texforms that were unrecognizable at the basic level (parameters: critical spacing = .5, radial/tangential aspect ratio = 1). To create the textures used in the Experiment 4, we used only one pooling window that averaged these features across the entire image (see Figure A1).

Stimuli were placed on a 640 × 640 gray background that had the same average luminance as the image, and stimuli were placed at four different positions within these pooling windows at the same distance from the center of the image (or "fixation"). Lastly, Gaussian white-noise images were adjusted iteratively (using a variant of gradient descent) to conform to these modeled parameters from an original image (Freeman & Simoncelli, 2011) for 50 iterations.

III. Experiment 3 (Texforms): Linear Mixed-Effects Modeling

In Experiment 3, we also conducted linear mixed-effects modeling to ensure the results generalized across items. However, the maximal random-effects model justified by our design without the predicted interaction failed to converge. In this model, random item intercepts tended to be perfectly correlated with the overall intercept, suggesting less variability at the item level and thus an

Figure A2. Schematic of the basic level guessing task and example responses.

(*Appendix continues*)

overly complicated model. When random intercepts for items were removed (but random slopes for items retained), both models with and without the predicted interaction converged. Comparing these two models (as in Experiment 2) revealed that search was more efficient on mixed versus uniform displays, even when generalizing across participants and items, $\chi^2(1) = 6.05$, $p = .01$.

IV. Replication of Experiment 2 (Controlled Stimuli)

Experiment 2 (with controlled stimuli) showed a trend for a greater search advantage with small object targets, whereas Experiment 3 (with texforms) showed the opposite trend for a greater search advantage with big object targets. This difference is likely driven by subtle feature differences between the controlled stimuli and texforms. However, there were also minor methodological changes between these experiments that could contribute to these opposing trends. To examine this possibility, we conducted a replication study of Experiment 2, with two changes. First, items were presented in a circular display (as in Experiment 3), and second, the stimulus set was comprised of the original big and small objects that were used to create the texforms of Experiment 3.

Overall, we found that these differences in stimuli and display configuration did not influence the results (i.e., the pattern was the same as the original Experiment 2)—visual search was more efficient in mixed displays than in uniform displays: uniform slope, $M = 65.30$, $SD = 21.86$; mixed slope, $M = 55.41$, $SD =$

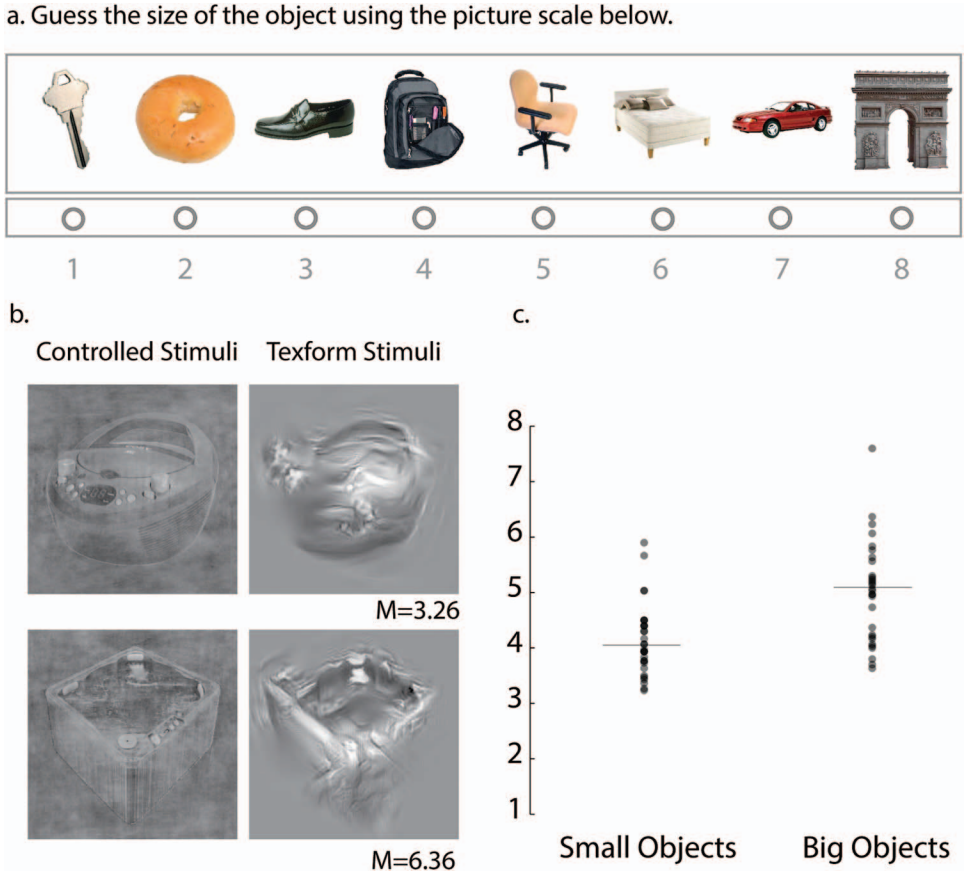

Figure A3. Task and results from the texform norming experiment. (a) Participants ($N = 30$) were asked to judge the size of the texform stimuli using a picture scale. (b) Example-controlled stimuli (used in Experiment 2) and texform stimuli (used in Experiment 3) are shown side by side. Below each texform is the average size ranking using the scale in the panel above. (c) Average size ranking values for all texforms used in Experiment 3. Each semitransparent dot represents one image; the lines represent the average of the size rankings for each object size. See the online article for the color version of this figure.

(*Appendix continues*)

18.21, $t(12) = 3.01$, $p = .01$. This result was confirmed by a three-way repeated-measures ANOVA on RT, with set size (3, 9), real-world target size (big, small), and display type (uniform, mixed) as factors. Participants responded faster when the target was small, $F(1, 12) = 38$, $p < .001$, $\eta_p^2 = 0.76$, and when the displays were mixed, $F(1, 12) = 20$, $p < .01$, $\eta_p^2 = 0.625$. Most important, the interaction between set size and display type was significant, $F(1, 12) = 8.3$, $p = .01$, $\eta_p^2 = 0.41$, indicating that the increase in RT with additional distractors was reliably lower for mixed trials than for uniform trials. This effect was again not modulated by the real-world size of the target, $F(1, 12) = 1.44$, $p = .253$, $\eta_p^2 = 0.11$.

V. Direct Replication of Experiment 3 (Texforms)

To ensure that we had reliable results, we conducted a direct replication of our study with another group of 13 participants. Overall, we found the same pattern of results: visual search was more efficient in mixed displays than in uniform displays (uniform slope: $M = 78.28$, $SD = 16.89$, mixed slope: $M = 64.44$, $SD = 13.84$, $t(12) = 4.08$, $p < .01$). This result was confirmed by a 3-way repeated measures ANOVA on RT, with set size (3,9), real-world target size (big, small), and display type (uniform, mixed) as factors. Participants responded faster when the displays were mixed, $F(1, 12) = 22$, $p < .001$, $\eta_p^2 = 0.647$). Most important, the interaction between set size and display type was significant, $F(1, 12) = 15.6$, $p < .01$, $\eta_p^2 = 0.57$). As before, this effect was not modulated by the real-world size of the target, $F(1, 12) = 2.2$, $p = .164$, $\eta_p^2 = 0.16$).

VI. Experiment 3: Texform Norming

Consistency of guesses. In our texform norming task, participants were informed that the texforms were "scrambled objects". Even so, participants were not very accurate in identifying the basic-level category of the texforms ($M = 2.8\%$, $SD = 4.03\%$). Not only were they inaccurate, but they were also inconsistent with each other. To show this inconsistency, we grouped responses by basic-level category and counted the number of unique responses to a given texform. Responses were grouped relatively generously; similar subordinate categories were grouped together (i.e., high-heeled shoe, boot, and shoe). When an observer failed to give a response ("I don't know"), this was not counted as a unique response. Unlike participants in the search task (Experiment 3), participants in the norming task rarely responded with "I don't know" ($M = .5\%$ of all responses; 10 responses across all participants). Unique responses to a given texform accounted for 74.8% of the reported object identities ($M = 22.3$, $SD = 3.4$ unique identities for 30 participants, 60 items). Thus, the texforms do not appear to look like any particular object (see Figure A2).

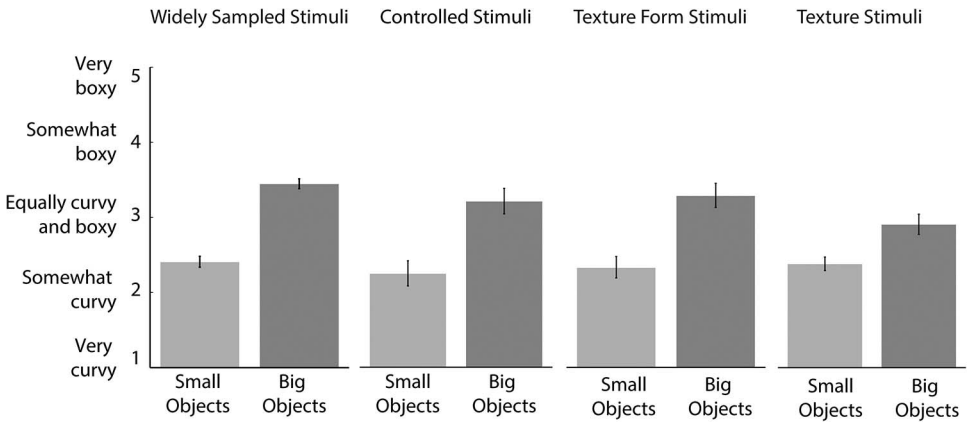

Figure A4. Small objects were judged to be curvier than big objects across our four different stimuli sets. Error bars represent the standard error of the mean.

(*Appendix continues*)

Real-world size classification. Participant's were unable to reliably guess the basic-level identities of the texforms. However, could these same participants guess the size of the original images used to generate the texforms? Participants ($N = 30$) were also asked to "guess the size of the object using the picture scale below" (see Figure A3) for each texform in a random order.

We first binarized all size rankings and compared them to the original size category of each texform (see Figure A3). Small: key (1) through backpack (4); big: chair (5) through arch (8). This resulted in a binary accuracy score for each item and participant. We then averaged across all participants to generate a size classification score for each item.

To ask if participants were able to classify the items at a rate above chance, we compared average size classification scores across items to a bootstrapped distribution for chance performance on this task. Specifically, we simulated chance performance for 1000 experiments for 30 observers rating 30 items. Both the average size classification score for both big objects ($M = 56.44\%$, $SD = 14.06\%$) and small objects ($M = 63.22\%$, $SD = 15.02\%$) fell above the highest value obtained on this distribution, indicating that they were classified at a rate above chance ($p < .0001$). Figure A3 contains a plot that shows the average size rankings of these 60 texforms.

VII. Curvature Ratings (Experiments 1–4)

One possible mid-level perceptual difference between big and small objects is the degree of curvilinearity versus rectinlinearity. Intuitively, small objects may be curvier than big objects, as they are often made to be hand-held, whereas big objects may be more rectlinear, as they are structures that must withstand gravity and provide surfaces. We explored the relationship between our big and small object stimulus sets and curvature judgments in several online behavioral experiments. Four sets of 20 observers on Amazon Mechanical Turk (mTurk) rated each item from our different stimulus sets (in a random order) according to the following scale: 1 (*very curvy*), 2 (*somewhat curvy*), 3 (*equally boxy and curvy*), 4 (*somewhat boxy*), 5 (*very boxy*).

Ratings were averaged for big and small categories (see Figure A4). Overall, small objects were consistently judged to be curvier than big objects in the widely sampled stimuli used in Experiment 1: big objects, $M = 3.45$; small objects, $M = 2.41$, $t(398) = -10.68$, $p < .0001$; the controlled stimuli used in Experiment 2: big objects, $M = 3.29$; small objects, $M = 2.29$, $t(58) = -4.03$, $p < .001$; the texforms used in Experiment 3: big objects, $M = 3.20$; small objects, $M = 2.33$, $t(58) = -4.46$, $p < .001$; and the textures used in Experiment 4: big objects, $M = 3.02$; small objects, $M = 2.46$, $t(58) = -3.25$, $p < .01$. These data suggest that differences in curvature may be one important cue for characterizing the features of big versus small objects.

Received September 10, 2014
Revision received June 26, 2015
Accepted October 9, 2015 ■