Cognition 168 (2017) 234–242

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Original Articles

A familiar-size Stroop effect in the absence of basic-level recognition ☆

Bria Long *, Talia Konkle

Department of Psychology, Harvard University, United States

ARTICLE INFO

Article history:
Received 1 March 2017
Revised 23 June 2017
Accepted 23 June 2017

Keywords:
Object recognition
Semantic access
Real-world size
Stroop effect

ABSTRACT

When we view a picture of an object, we automatically recognize what the object is and know how big it typically is in the world (Konkle & Oliva, 2012). Is information about an object's size activated only after we've identified the object, or can this size information be activated before object recognition even occurs? We previously found that big and small objects differ in mid-level perceptual features (Long, Konkle, Cohen, & Alvarez, 2016). Here we asked whether these perceptual features can automatically trigger real-world size processing, bypassing the need for basic-level object recognition. To test this hypothesis, we used an image synthesis algorithm to generate "texform" images, which are unrecognizable versions of big and small objects that still preserve some textural and form information from the original images. Across two experiments, we find that even though these synthesized stimuli cannot be identified, they automatically trigger familiar size processing and give rise to a Size-Stroop effect. Furthermore, we isolate perceived curvature as one feature the visual system uses to infer real-world size. These results suggest that mid-level perceptual features can automatically feed forward to facilitate object processing, and challenge the idea that we must first identify an object before we can access its higher-level properties.

1. Introduction

Our object recognition system runs so smoothly and automatically in the background that we rarely notice it tolling away. This system seems particularly adept at identifying what we see at the *basic level* – for example, if we see a small, smooth object with a handle, we first identify this as "a mug" rather than as something more general ("an inanimate object") or something more specific ("the coffee mug I received from my grandmother", Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). In fact, some work suggests that we can categorize objects at the basic level as quickly as we can detect their presence (Grill-Spector & Kanwisher, 2005). Our automatic and effortless ability to categorize and identify visual objects is often taken as the core goal of the brain's visual recognition system (DiCarlo & Cox, 2007).

However, recently it was also demonstrated that as soon as we see a pictured object, we also automatically activate information about how big or small the object typically is in the world (Chiou & Ralph, 2016; Gliksman, Leibovich, Melman, & Henik, 2016; Konkle & Oliva, 2012; Sellaro, Treccani, Job, & Cubelli, 2015; see also Paivio, 1975; Rubinsten & Henik, 2002). Some evidence for this automatic activation comes from a Size-Stroop paradigm. In this task, participants were asked to compare two objects and decide which one is visually bigger or smaller on the screen, ignoring the real-world size of the objects. The visual sizes of the two depicted objects could either be congruent with their real-world size (e.g. a small cup and a big car), or incongruent (e.g. a big cup and a small car) (see examples in Fig. 1). Critically, the task only required judging which image was bigger or smaller *on the screen*—knowledge about the real-world sizes of the objects was irrelevant to the task. However, participants were faster to make visual size judgments on the congruent trials, indicating that they could not help but automatically process real-world size when presented with pictures of these objects.

Do we need to recognize a pictured object in order to know its size in the real world? Classic models of conceptual representation argue that semantic knowledge about objects is organized as a series of predicates (e.g., "big enough to support a human") that are attached to conceptual nodes, such as "chair" (Collins & Quillian, 1969; Jolicoeur, Gluck, & Kosslyn, 1984). These nodes can be activated by the correct sets of input from the visual processing stream, and in turn, serve as the point from which we access knowledge about objects, such as how big or small they are in the real world, or the context in which they are typically used (i.e., a kitchen). On this account, object recognition precedes our ability to access knowledge about an object. However, recognition need not be the gateway through which we access all kinds of

☆ We have uploaded all data and analysis code to the first author's GitHub account, which is linked to an Open Science Repository for this project (https://osf.io/dt5a6/).

* Corresponding author at: 33 Kirkland Street, Cambridge, MA 02140, United States.
E-mail address: brialorelle@gmail.com (B. Long).

http://dx.doi.org/10.1016/j.cognition.2017.06.025
0010-0277/

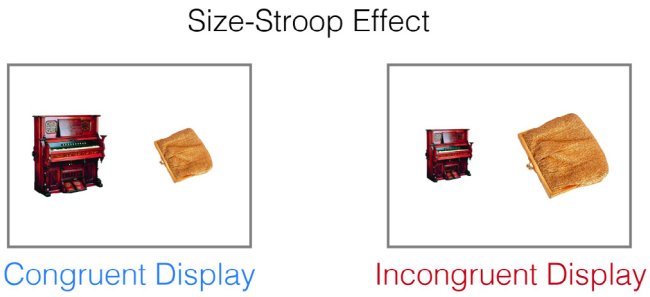

Fig. 1. Example Size-Stroop displays from Konkle and Oliva (2012). Two objects were displayed and the task was to judge which item was bigger on the screen. In congruent trials, the object that was bigger on the screen was also a bigger object in the real world. In incongruent trials, the object that was bigger on the screen was a smaller object in the world. Participants were faster to make visual size judgments when the visual size of the object was congruent with the real-world size of the object, even though the object's real-world size was irrelevant to the task.

object knowledge. On an alternative account, perceptual feature evidence accrued in parallel to the process of object recognition could be used to make inferences about different functional properties of objects, including their size in the real world. Some evidence for this alternative was recently provided by Cheung and Gauthier (2014), who demonstrated that specific perceptual features, like smoothness and symmetry, can automatically activate conceptual information about whether something is animate or inanimate. Thus, an alternative possibility is that perceptual features can automatically activate real-world size information.

In prior work we established that there exist systematic perceptual differences that distinguish big objects from small objects. To do so, we used a visual search task, with the logic that visual search is slower when targets and distractors are perceptually similar (Duncan & Humphreys, 1989; Long, Konkle, Cohen, & Alvarez, 2016). We found that participants searched more efficiently for a small object target (e.g. cup) among big object distractors (e.g. couch, piano, chair), and vice versa. Critically, this visual search advantage persisted even when participants were searching for unrecognizable versions of big and small objects that preserved some texture and form information—"texform" stimuli (Freeman & Simoncelli, 2011; Long et al., 2016). These results indicate that big objects and small objects have systematic perceptual differences that are preserved in "texform" stimuli.

Given this existence proof of feature differences, we can now directly test the deeper question about the role these might play in our cognitive architecture: do these perceptual features directly activate size concepts and automatically trigger real-world size processing, without requiring basic-level object recognition? To do so, we used the Size-Stroop paradigm from Konkle and Oliva (2012), but with unrecognizable texform stimuli. If basic-level recognition is a necessary precursor to real-world size inferences, then these texforms should not trigger any real-world size related processing, and thus should not impact the speed of visual size judgments in the Size-Stroop task. However, if these texform stimuli do trigger real-world size processing, we should see evidence for a Size-Stroop effect.

To anticipate our results, we find that unrecognizable texform stimuli generate a Size-Stroop effect (Experiment 1), and the strength of this effect depends on the degree to which texforms preserve information related to real-world size (Experiment 2). To provide some intuitions about the features preserved in the texforms that underlie these effects, we explored several properties. We found that the perceived curvature of the texforms, but not perceived viewing distance or depicted depth, predicted the magnitude of the Size-Stroop effect for individual displays. Taken together, these results demonstrate that real-world size information is automatically activated by perceptual features, including curvature properties, when observers perform a visual size task. Broadly, these results are consistent with the possibility of a modified cognitive architecture in which early visual processing can directly trigger the processing of higher-level object properties, including real-world size.

2. Experiment 1

Texform images of big and small objects were generated using a computational model of early visual processing (all stimuli in Fig. 2; Freeman & Simoncelli, 2011; Long et al., 2016). In the first experiment, two texforms were presented simultaneously at different visual sizes, and we asked participants to make a visual size judgment about which of two texforms was bigger or smaller on the screen. Unbeknownst to the participants, on some displays, the relative visual sizes of the texforms were congruent with the real-world sizes of their original objects (e.g. a big piano texform and a small key texform). On other displays, this relationship between

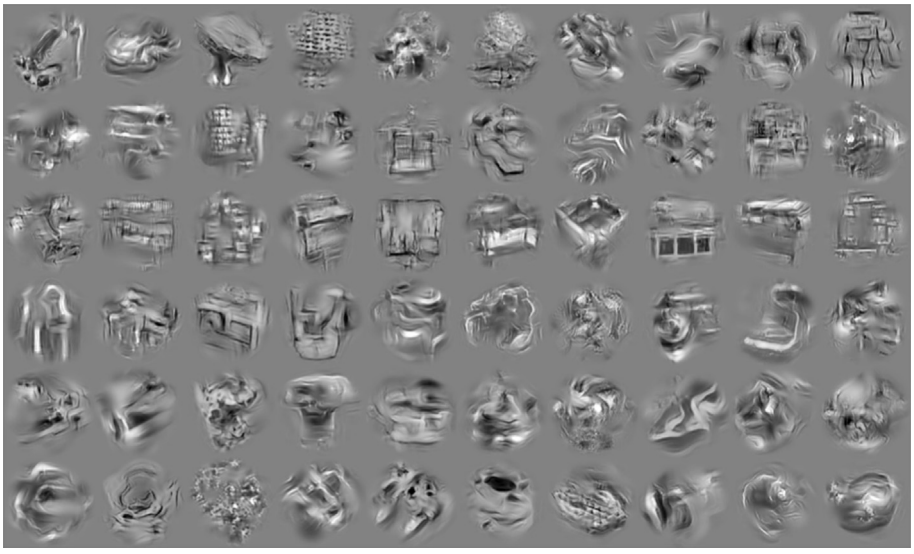

Fig. 2. All 60 texforms used in Experiments 1 and 2. The top three rows correspond to texforms generated from pictures of big objects, and the bottom three rows correspond to texforms generated from pictures of small objects.

visual size and real-world size was reversed. If real-world size information can be triggered from these texforms in the absence of basic-level object recognition, then participants should be faster to make a visual size judgment on congruent displays.

2.1. Methods

2.1.1. Participants

Sixteen Harvard affiliates or students, age 18–35, gave informed consent and participated in the experiment. This sample size was chosen following Konkle and Oliva (2012). Participants had normal or corrected-to-normal vision.

2.1.2. Stimuli

The stimulus set consisted of 60 texform images generated from images of 30 big, inanimate objects and 30 small, inanimate objects (see Fig. 2). Big objects included things like cars and tables and were chair-sized and bigger; Small objects included things like mugs and cameras, and were table-lamp sized and smaller. The texform stimuli were synthesized using an algorithm that preserves mid-level image features from the original images, such as local combinations of orientations (see Long et al., 2016 for more detailed description of the procedure; see also Freeman & Simoncelli, 2011).

To ensure that these texforms were unrecognizable at the basic-level, we asked 30 observers to name a larger set of texforms. We then selected this set of 60 texforms to minimize recognizability, even when we coded generously for basic-level identity (e.g., 'stove' was accepted as a correct response for 'jukebox' because it is the same sized object with a similar shape). In this final subset of 60 items, the average identification accuracy was 2.83%, SD = 4.02% (N = 30). See the Appendix for examples of two items and guesses from 30 observers.

To create the Stroop displays, we required a visually big and a visually small version of each texform. We used the original synthesized texforms as the visually big size (440 × 440 pixels), and then rescaled the image to make a visually smaller size (300 × 300 pixels), and placed it centered in a uniform gray background (440 × 440 pixels); see Fig. 3. Including the backgrounds, visually big and small texforms had the same degree of visual angle (~18.5 deg). Within these backgrounds, visually big stimuli subtended around ~16–18 deg of visual angle, while visually small stimuli subtended around ~11–13 deg of visual angle.

2.1.3. Apparatus

Participants were positioned 57 cm away from an Apple iMac computer (1024 × 768 pixels, 60 Hz), such that 1 cm on the screen

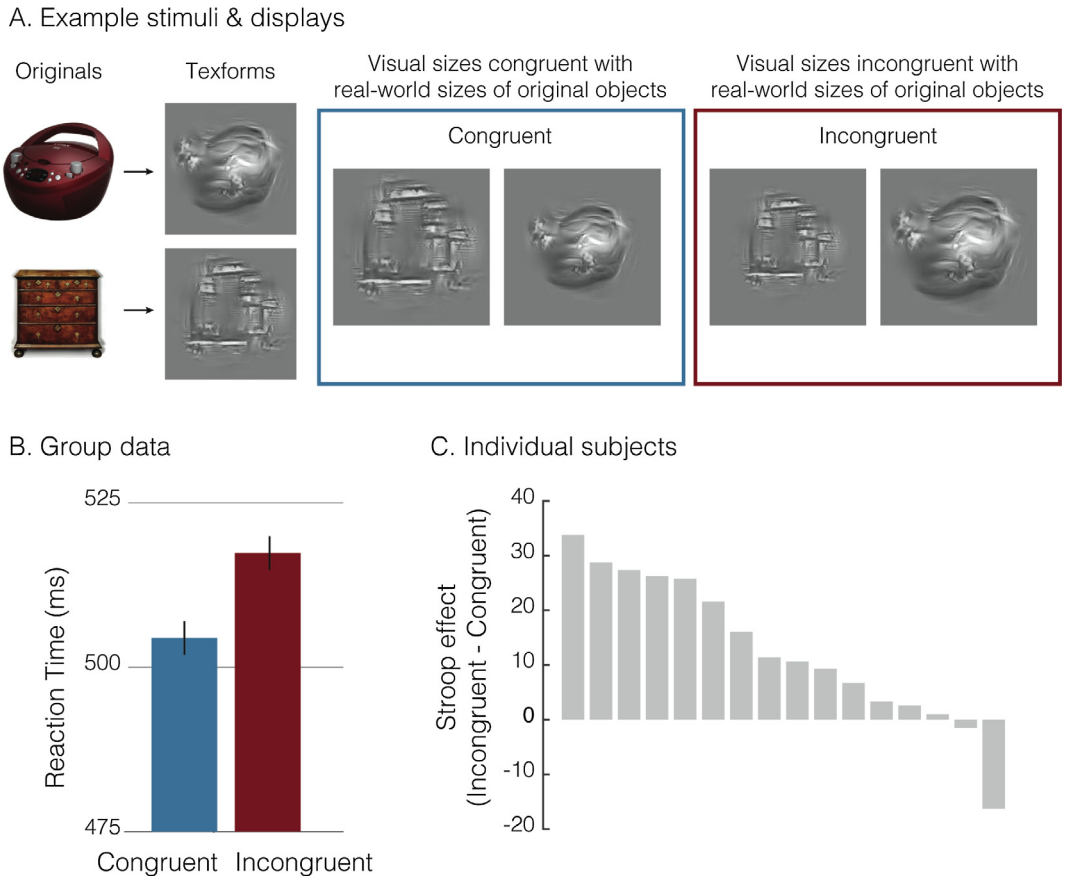

Fig. 3. Example displays and stimuli used in Experiment 1. (A) (left) Examples of original objects and texforms. Texform stimuli were generated from pictures of big and small objects using a texture-synthesis model (Freeman & Simoncelli, 2011). (right) Example Size-Stroop displays with texforms. Participant's task was to make a visual size judgment about which object was smaller or bigger on the screen. On congruent displays, the visual sizes of the texforms were congruent with the real-world sizes of their original objects. For example, a texform of a dresser would be presented at a visually big size, and a texform of a boombox would be presented at a visually small size. On incongruent displays, the visual sizes of the texforms were incongruent with the real-world sizes of their original objects. Here, the texform of a dresser was presented at a visually small size, whereas the texform of a boombox was presented at a visually large size. (B) Average reaction times from Experiment 1 are plotted for congruent and incongruent trials. Error bars represent within-subjects standard error (Morey, 2008). (C) The average Size–Stroop effect is plotted for each subject, measured by taking the difference in reaction times between incongruent and congruent trials.

was approximately equal to 1 deg of visual angle. Experiments were run using Psychtoolbox (Brainard, 1997; Pelli, 1997) in Matlab 2010a.

2.1.4. Design

The design of this study was identical to Konkle and Oliva (2012), except that stimuli were grayscale *texforms* instead of color images of recognizable objects.

On every trial, a fixation cross first appeared for 700 ms. Afterwards, two grayscale texforms appeared on either side of fixation on a white background. On half the trials, participants were asked to judge which texform was visually *bigger* on the screen as fast as possible. On the other half of the trials, participants were asked to judge which texform was visually *smaller* on the screen as fast as possible. Participants indicated which side of the screen corresponded to the visually bigger or visually smaller image by pressing either the *m* key or the *c* key. The images remained present on the display until the participant responded. High accuracy was encouraged as incorrect responses resulted in feedback and a 5s interval before the next trial began. After a correct response, there was a 900 ms interval before the next trial.

Trials were blocked into 4 sets, where the task switched after each set. Half of the participants started with the "visually bigger" task, and half of the participants started with the "visually smaller" task. To orient people to the tasks, all participants first saw example trials and read instructions, and then completed 24 practice trials with both task instructions (12 in each task) in the same counterbalanced order as the experiment.

The critical manipulation was that the two texforms on each display were presented at visual sizes that were either congruent or incongruent with the real-world size of the original objects. For example, in a congruent display, a shoe texform would be presented at a visually small size and a couch texform would be presented at a visually big size, as typically shoes are small and couches are big in the world. On incongruent trials, this was reversed: a shoe texform would be presented at a visually big size, and a couch texform would be presented at a visually small size. The fact the texforms were generated from objects of different real-world sizes was not mentioned at any time during the experiment. Furthermore, participants never saw a version of the Size-Stroop task with recognizable objects.

At an item level, each big object texform and small object texform were counterbalanced such that they appeared equally in both congruent/incongruent configurations, with the correct answer on the left/right side of the screen, and across both visual size tasks. Big and small object texforms were pseudo-randomly paired, such that the same random pairs of big and small texforms occurred together in the first half of the experiment for each participant. In the second half of the experiment, big and small object texforms were randomly paired together; this procedure was used in Konkle and Oliva (2012) to take into account pictorial issues related to recognizable objects, and for consistency we followed the exact procedure here. Overall, there were 480 trials (30 pairs of objects $\times$ 2 congruent/incongruent conditions $\times$ 2 left/right sides of screen $\times$ 2 bigger/smaller tasks $\times$ 2 different pairings of texforms; yielding 240 congruent/240 incongruent trials).

2.1.5. Analysis

Incorrect trials and trials for which reaction times (RT) were shorter than 200 ms or longer than 1500 ms were excluded, following Konkle and Oliva (2012) (2.55%). Trimmed reaction times were analyzed using a 2 $\times$ 2 repeated-measures ANOVA, with congruency (congruent/incongruent) and task (bigger/smaller on the screen) as factors.

2.2. Results

Our main question of interest was whether we would observe a Size-Stroop effect without basic-level object recognition. Fig. 3B shows the reaction time for the congruent and incongruent trials. Overall, we found evidence for a Size-Stroop effect: on incongruent trials, participants were slower to make visual size judgments when the real-world size of original objects were incongruent with their sizes on the screen (M_{diff} = 12.92 ms, SD_{diff} = 13.62 ms, main effect of congruency, $F(1,15) = 14.3$, $p = 0.002$, $\eta_p^2 = 0.489$, Cohen's $d = 0.95$, Fig. 2B). Furthermore, 14 out of 16 participants showed the effect in the predicted direction (Fig. 3C).

Across the two tasks ("which is bigger" vs. "which is smaller"), participants were equally fast (no main effect of task, $F(1,15) = 0.38$, $p = 0.548$, $\eta_p^2 = .025$); task did not interact with the magnitude of the Stroop effect ($F(1,15) = 1.33$, $p = 0.266$, $\eta_p^2 = 0.082$). Consistent with this result, targeted *t*-tests revealed a Size-Stroop effect both when participants reported which item was bigger (M_{diff} = 8.18 ms, SD_{diff} = 18.03 ms, $t(15) = 1.82$, $p = 0.089$) and when participants reported which item was smaller (M_{diff} = 17.68 ms, SD_{diff} = 24.29 ms, $t(15) = 2.91$, $p = 0.011$). Numerically, the effect was stronger when observers were judging which texform was smaller on the screen, which coincides with previous findings of Konkle and Oliva (2012). No differences were observed in error rates (all $p > 0.2$).

Experiment 1 demonstrated that participants were faster at judging the visual sizes of the texforms when their original real-world sizes were congruent with their visual sizes. Thus, even though these texform stimuli were not identifiable, their original real-world sizes impacted how quickly participants made visual size judgments in the Size-Stroop paradigm. These results suggest that real-world size information can be activated from mid-level feature processing alone, even when basic-level recognition is impaired.

3. Experiment 2

The results of Experiment 1 rely on the fact some texture and form features are preserved in the texform stimuli which still enable participants to reliably process real-world size information, even though they cannot recognize the original objects. However, not all of the texform stimuli preserve real-world size information equally well—although none of these texforms can be recognized at the basic-level, some texforms can be reliably classified as big or small objects, while other texforms cannot (Long et al., 2016). Thus, we reasoned that texforms that are well classified by their real-world size should do a better job of activating real-world size information, and thus should also generate the largest Size-Stroop effects. In Experiment 2, we systematically paired texforms according to how classifiable they were by their real-world size, allowing us to estimate Size-Stroop effects for individual displays. We expect that displays with highly classifiable texforms should generate larger Size-Stroop effects.

3.1. Methods

3.1.1. Participants

Twenty-four Harvard affiliates or students were recruited, gave informed consent, and participated in this study. The sample size was larger than in Experiment 1 to have added power for display-level effects. Participants were between 18 and 35 years of age and had normal or corrected-to-normal vision.

3.1.2. Stimuli

Stimuli were the same as in Experiment 1, but were paired on each display by their real-world size classifiability. To measure this for each texform, an Amazon Mechanical Turk study was run in which participants (N = 30) guessed the real-world size of each texform using a Likert scale (1: small as a key, 8: big as an arch; these data were also reported in Long et al. (2016); see Konkle and Oliva (2011), for more extensive characterization of the 1–8 size scale

and it's relationship to actual physical size). Responses were counted as correct if they fell within any response in the correct size category (Small: key-sized through backpack-sized, Big: chair-sized through arch-sized), and averaged across subjects to create a size classifiability score for each texform. Then, 30 big object texforms and 30 small object texforms were ordered as a function of how well they were classified as big versus small objects and then paired, creating 30 pairs of big and small objects. Importantly, these 30 pairs spanned nearly the entire range of size classification accuracy: some pairs of objects were very well classified as big or small objects, some were near chance classification accuracy (50%), while others were systematically misclassified as big or small objects, leading to performance well below chance (see Fig. 4A, range 16.7–95.0%, *SD* = 21.0%).

In addition, we made two changes to how we created visually small versions of the texforms. First, we ensured that the transition between the texforms and their backgrounds was gradual. To do so, we gradually faded each texform into the background by first overlaying a semi-transparent circle on each texform (using a Gaussian window) before embedding them on gray backgrounds. This blurring was done to remove a few edge artifacts introduced by bounding box of the texforms. Secondly, we resized the images to only 80% of their original size (352/440 pixels, ~12–14 deg of visual angle) to make the task slightly more difficult, thereby increasing our chance of finding differences among individual displays. All other procedures were identical to Experiment 1.

3.1.3. Analysis

First, we removed outliers (4.0% of trials) and analyzed our data in the same way that we did in Experiment 1. We also calculated display-level Size-Stroop effects by subtracting the difference between incongruent and congruent reaction times for each display after averaging across all subjects and both tasks.

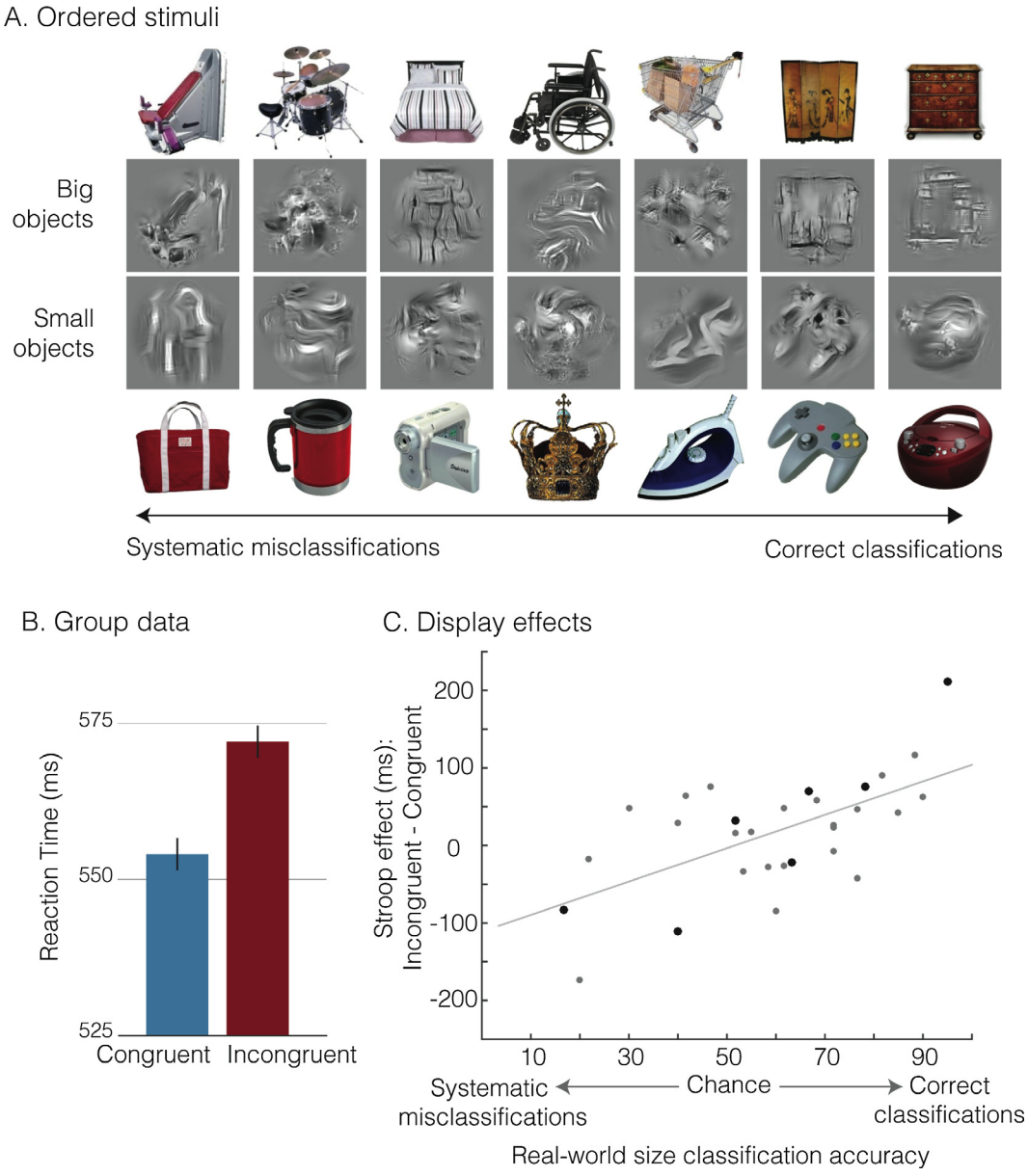

Fig. 4. (A) Texforms are ordered by how classifiable they are as big or small in the real-world. Original objects are shown adjacent to their texforms separately for big objects (top two rows) and small objects (bottom two rows). (B) Group average across all Stroop displays, replicating the main effects from Experiment 1. Error-bars represent within-subjects standard error of the mean (Morey, 2008). (C) Display-Analyses. The strength of the Stroop effect for any given pair of texforms (y-axis) was predicted by how well these texforms were classified as big versus small objects (x-axis). Black dots represent the Stroop effects for each pair of texforms depicted in Panel A.

3.2. Results

Overall, we replicated the same pattern of effects we found in Experiment 1 (Fig. 4B). Participants were faster at making visual size judgments when the original sizes of the texforms were congruent with their sizes on the screen (M_{diff} = 18.05 ms, SD_{diff} = 17.02 ms, $F(1,23) = 30.9$, $p < 0.0001$, $\eta_p^2 = 0.573$, see Fig. 4B). In addition, we found an effect of task: participants were generally slower when judging which texform was smaller on the screen ($F(1,23) = 8.8$, $p < 0.007$, $\eta_p^2 = 0.277$) but also showed a stronger Stroop effect (M_{diff} = 28.34 ms, SD_{diff} = 23.29 ms) than when they judged which texform was bigger on the screen (M_{diff} = 8.26 ms, SD_{diff} = 28.17 ms, congruency by task interaction, $F(1,23) = 5.94$, $p = 0.023$, $\eta_p^2 = .205$). This was the same trend we observed in Experiment 1, and that was found by Konkle and Oliva (2012) with pictures of recognizable objects. Thus, we again found that real-world size information is automatically activated by mid-level features when observers make visual size judgments.

Our critical question for Experiment 2 was whether displays with texforms that are well classified as big or small objects are also the displays that generate the largest Size-Stroop effects. Consistent with this prediction, the degree to which pairs of texforms were classified as big versus small objects predicted the magnitude of their Size-Stroop effect ($r = 0.61$, $p < 0.001$, see Fig. 4C). Furthermore, when we performed this same correlation in every subject, we found a positive correlation in each case (average correlation across subjects, $r = 0.42$, $SD = 0.13$). This result was also confirmed with a linear regression analysis, where average size classification accuracy significantly predicted display-by-display Stroop effects (B = 215 ms, $t(28) = 4.08$, $p = 0.0003$, adjusted $R^2 = 0.35$).

Overall, these results confirm and extend the results from Experiment 1, demonstrating that the degree to which real-world size information is present in the texforms also predicts the strength at which we see automatic real-world size interference in the Size-Stroop task.

4. Which mid-level features activate size information?

In Experiments 1 and 2, we found that the mid-level features preserved in texforms activated real-world size information in the Size-Stroop paradigm. What could these mid-level features be? To provide an intuitive sense of what kind of information is captured and could be playing a role in this Size-Stroop task, we examined three candidate perceptual properties: perceived curvature, perceived viewing distance, and depicted depth.

Intuitively, man-made objects that are big in the real world may tend to be boxier in order to withstand gravitational and physical constraints. Conversely, small, graspable objects can have almost any given shape. Consistent with this idea, in prior work we found that big and small objects tend to differ in perceived curvature; big objects tend to be boxier than small objects (Long et al., 2016). To ask whether observers use this to infer real-world size, we examined if the perceived curvature of a texform predicted it's perceived real-world size. If so, the curvature information preserved in the texforms could trigger the automatic processing of real-world size.

In addition, it is possible that texforms preserve information that conveys distance information. For example, it is possible that big texforms appear further away than small texforms. If this were the case, observers could infer real-world size from the perceived distance of a texform (Amit, Algom, & Trope, 2009; see also Paivio, 1975). That is, texform features might activate distance information, and distance might trigger size representations. On this account, mid-level features would not directly activate size representations. To address this possibility, we had observers rate the perceived distance of big and small texforms.

Finally, we examined whether texforms of big and small objects differ in how much depth they depict. This dimension is related to perceived distance, but measures not how far or close the object is to the viewer, but how far the object itself extends in depth. For example, a picture of a table that is rotated to show all four legs may extend further in depth than when it is not rotated. If there are consistent differences in depicted depth across big and small objects that are also preserved the texform algorithm, observers could be using this information to infer real-world size.

To explore these properties, we obtained behavioral ratings of both texforms and their recognizable counterparts on these three perceptual properties. Then, we examined if texforms' values on these properties predicted their perceived size in the real world. Finally, we asked if any differences between big and small object texforms on these three properties predict the Size-Stroop display effects we observed in Experiment 2.

4.1. Methods

4.1.1. Stimuli

The texforms used in Experiments 1–2 and their recognizable counterparts were divided into two counterbalanced sets. This ensured that participants would never see the original objects from which the texforms were generated. Thus, each counterbalanced set contained 30 texforms and 30 original objects.

4.1.2. Participants

108 participants participated on Amazon Mechanical Turk for the following rating studies. Overall, we collected ratings from 16 participants on both sets of 60 images (30 texforms and 30 recognizable objects) for each property.

4.1.3. Procedure

Participants rated images in a random order on one of three properties using a 5-point Likert scale. To rate perceived curvature, the instructions were: "How curvy or boxy is this object?" To rate perceived viewing distance, the instructions were: "How far away is the object depicted in this image?" To rate depicted depth, the instructions were: "How much depth is depicted in the picture of this object?"

4.1.4. Analysis

For each image, scores for each of these three properties were calculated by averaging across the 16 raters. We also computed property difference scores for all Stroop displays used in Experiment 2. Specifically, for each property and pair of big and small texforms, we subtracted the property score for the small object texform from the property score for the big object texform. This was done because we expected big object texforms to have higher values on each of these property (i.e., to be boxier, farther away, and depicting more depth). Then, we correlated these difference scores with the Stroop display effects (Incongruent RT – Congruent RT) from Experiment 2. Finally, we also used a linear regression, entering display-by-display differences in curvature, perceived viewing distance, and depicted depth as predictors, and display-by-display Stroop effects as the dependent variable.

4.2. Results

4.2.1. Curvature

Consistent with our previous findings (Long et al., 2016), we found that recognizable big objects were perceived as boxier than recognizable small objects ($t(58) = 3.68$, $p < 0.001$). This relationship also held for texform images ($t(58) = 4.07$, $p < 0.001$). In addition, the perceived curvature of the texform stimuli correlated with their perceived size in the real-world, when considering size as a continuous dimension ($r = 0.75$, $p < 0.001$; Fig. 5A). Finally, the Size-Stroop display effects seen in Experiment 2 were also

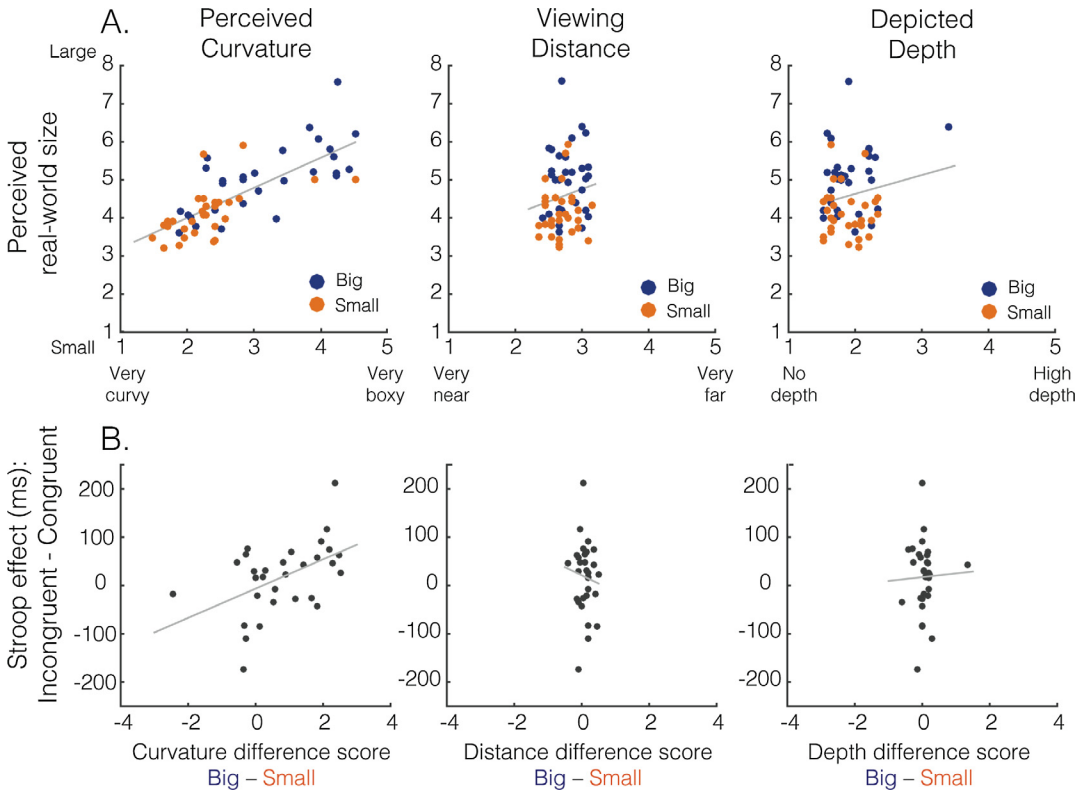

Fig. 5. (A) The perceived real-world size of the texforms (y-axis scale: 1 = small as a key, 8 = big as an arch) is plotted as a function of their perceived curvature (x-axis, left panel), their perceived distance from the viewer (x-axis, middle panel), and the amount of depth depicted in each image (x-axis, right panel). Texforms generated from pictures of small objects are colored in orange (gray); texforms generated from pictures of big objects are colored in blue (dark gray). (B) Each dot represents an individual Stroop display (pair of texforms). The strength of the Stroop effect for each display (y-axis) is plotted as a function of how *different* the two items on the display were in terms of their perceived curvature (big object texform – small object texform), perceived viewing distance, and depicted depth. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

predicted by the perceived curvature differences of the big and small object texforms on each display ($r = 0.48$, $p < 0.01$; B = 31.97 ms, $t(26) = 2.77$, $p = 0.010$; Fig. 5B). In sum, texforms that were perceived as boxier were also perceived as bigger in the real world. And, when a big object texform was boxy and a small object texform was curvy, this pair of texforms tended to generate a robust Size-Stroop effect. Thus, these results provide reasonable evidence that *perceived curvature* is one property of the mid-level feature information in texforms that is used to infer real-world object size.

4.2.2. Viewing distance

We next examined whether we would see this same pattern of effects with perceived viewing distance. Recognizable pictures of big objects were perceived as farther away than small objects ($t(58) = 12.63$, $p < 0.001$). However, big object texforms were not perceived as farther away than small object texforms ($t(58) = 1.85$, $p = 0.07$). A texform's perceived viewing distance also was not correlated with its perceived size in the real world ($r = 0.15$, $p = 0.25$; Fig. 5A). Finally, display-level differences between the perceived viewing distance of big and small object texforms did not predict the Size-Stroop effects they generated ($r = -0.09$, $p = 0.63$; B = 25.75 ms, $t(26) = 0.37$, $p = 0.72$; Fig. 5B). Thus, it is unlikely that the Size-Stroop effects found in the first two experiments are mediated through the automatic processing of viewing distance. Further, the mid-level features preserved by the texform algorithm do not preserve differences in perceived viewing distance, so any effects found with texform stimuli are unlikely to be driven by this factor.

4.2.3. Depicted depth

Recognizable big object images had slightly more depicted depth than recognizable small objects images ($t(58) = 2.51$, $p = 0.02$). However, this was not true of texforms ($t(58) = 0.64$, $p = 0.527$). Further, the depicted depth of the texform images did not correlate with their perceived size in the real world ($r = 0.17$, $p = 0.18$; Fig. 5A). Finally, display-level differences between the depicted depth of big and small object texforms also did not predict the Size-Stroop effects they generated ($r = 0.03$, $p = 0.86$; B = −11.89 ms, $t(26) = -0.27$, $p = 0.79$; Fig. 5B). Thus, as with viewing distance, is it unlikely that depicted depth information is triggering size processing in the Size-Stroop effect, nor is depicted depth information a part of the mid-level features preserved by the texform algorithm.

5. General discussion

Overall, we found that real-world size information was automatically activated when observers made visual size judgments, even though basic-level recognition was impaired. In Experiment 1, we found that visual size judgments took longer when the retinal sizes of unrecognizable texforms were incongruent with their familiar, real-world sizes. In Experiment 2, we validated this result, and further demonstrated that texforms that were well classified as big versus small objects (while still remaining unrecognizable) generated larger Size-Stroop effects. We then explored three possible perceptual properties that might be preserved in texforms and underlie this Size-Stroop effect: curvature, viewing distance, and depicted depth. Only perceived curvature information was reliably retained in the texforms, and this feature predicted both the perceived real-world size of texforms and the display-by-display Stroop effects. Taken together, these results demonstrate that intact basic-level recognition is not necessary for the visual system to activate real-world size information. Furthermore, the presence

of size-related perceptual features, including curvature, is sufficient to automatically trigger real-world size processing in the Size-Stroop paradigm.

5.1. Sufficient vs. necessary features of big and small objects

While the texture synthesis algorithm that we use to generate these stimuli works by preserving mid-level feature information from the original images, one drawback is that it does not provide an intuitive explanation of what the critical features are that distinguish big from small objects. To this end, we explored a few candidates: perceived curvature, viewing distance, and depicted depth. Of these, only the perceived curvature of the stimuli had predictive power. Recognizable big objects tend to be boxier, recognizable small objects tend to be curvier, and the same is true of texforms. This suggests that *perceived curvature* is one reliable cue to real-world size. It is likely that curvature features are computed relatively early in visual processing and that these features can be used to trigger real-world size processing.

Why might this be the case? Boxier objects tend to do a better job of withstanding gravitational constraints: for example, buildings, bookshelves, desks, and tables may all have boxier shape features simply in order to support themselves (Long et al., 2016). We rarely observe large, man-made structures that are very curvy and do not have stable, boxy bases. In contrast, small objects can often be hand-held and have rounder shapes that enable grasping. These biases in shape features may cash out in systematic differences in perceived curvature between big and small objects. If the visual system is tuned to these natural statistics (i.e., Simoncelli & Olshausen, 2001), the visual system may learn, over time, that objects that tend to be big in the real world tend to have more rectilinear features.

However, it is important to note that we do not think that this single curvy-boxy axis reflects the totality of the features that distinguish big objects from small objects. Indeed, there are likely other mid-level features that contribute to our perception of an objects' real-world size, and are capable of triggering real-world size processing (e.g. information capturing graspable or structural parts). Further, while the texform algorithm does preserve some mid-level features that are cues to real-world size, it likely eliminates—or greatly reduces—others. This was the case with both viewing distance and depicted depth information: While recognizable pictures of big objects were perceived as farther away and extending farther in depth than small objects, this was not evident in judgments of big and small object texforms. Overall, the present work highlights curvature as a sufficient and reliable cue that activates real-world object size, and opens up future avenues for quantifying the other mid-level cues that may trigger real-world size processing.

5.2. Implications for cognitive architecture

Within a classic framework of object processing, the visual system extracts feature information leading to basic-level recognition, and these basic-level object representations then serve as pointers to more general knowledge about those objects (Jolicoeur et al., 1984; Rosch et al., 1976; but see Fabre-Thorpe, 2011; Macé, Joubert, Nespoulous, & Fabre-Thorpe, 2009). While this framework has intuitive appeal, the present results challenge a straightforward version of this model in which observers first explicitly recognize an object, and only then are able access knowledge about that object. Instead, we find that knowledge about an object's size in the real world can be activated in the absence of explicit access to a basic-level representation.

To accommodate these results, there are at least two possible accounts with distinct implications for the underlying cognitive architecture of object processing. First, within a classic hierarchy, mid-level features could be implicitly activating many basic-level object representations below the threshold for recognition, and this activation is then spreading to activate higher-level knowledge associated with those object representations. Note that this is still a substantial departure from the idea that explicit access to the basic-level is needed to access higher-level object knowledge. Alternatively, these findings are also consistent with a modified architecture, in which mid-level perceptual systems have parallel pathways to basic-level object representations as well as to broader category representations. On this account of the data, mid-level representations are directly activating information about higher-level object properties, including information about their size in the real world, bypassing a basic-level object representation.

Given this modified architecture, one interesting possibility is that the connection between mid-level features and real-world size might actually facilitate the process of object recognition. Specifically, mid-level features could automatically feed-forward to activate broad category information, such as the fact that an object is likely big or small in the real world. This activation could in turn constrain the space of possible basic-level identities considered by the visual system for basic-level recognition. The idea that directly activated higher-level knowledge can constrain object recognition is analogous to the framework proposed by Bar et al. (2006), which suggests that the context in which objects appear is processed prior to and informs basic-level recognition. However, unlike Bar's proposal in which the surrounding scene informs basic-level recognition, the mid-level perceptual features of the object itself could activate knowledge that informs the process of basic-level recognition.

This proposal also raises new questions about how various sources of object information combine. For example, mid-level feature information may not always perfectly determine the perceived size of an object—indeed, some big objects are rounder (hot air balloons) while some small objects are squarer (picture frames). When mid-level features activate higher-level knowledge about objects, to what degree do these facilitate, interfere with, or are overridden by basic-level recognition processes? One possibility is that the influence of mid-level predictions may play a more substantial role when objects are occluded or obscured (e.g. as in the case with texforms), compared to cases of clear central presentations when objects can be quickly identified.

Finally, it is important to note that here, mid-level features triggered real-world size processing specifically when participants performed visual size judgments. It thus remains an open question as to whether mid-level features will always activate real-world size information, or if they will only do so when the task involves a size-related component. Ultimately, future work is required to explore the boundary conditions of this process and how it interacts with other components of cognition (e.g., numerical cognition; see Gabay, Leibovich, Henik, & Gronau, 2013; Henik, Gliksman, Kallai, & Leibovich, 2017).

5.3. Conclusion

Overall, we find that real-world size information can be automatically activated in the absence of basic-level recognition. These results challenge the necessity of explicit basic-level recognition for semantic access, and suggest that mid-level features may contain rich information about broad category membership. We propose that examining how mid-level perceptual features activate high-level semantic knowledge is a promising avenue towards understanding how visual input rapidly contacts our conceptual representations, and the architecture underlying visual cognition.

Acknowledgements

We would like to acknowledge J. Freeman for providing the code used to generate the texform stimuli.

Appendix A

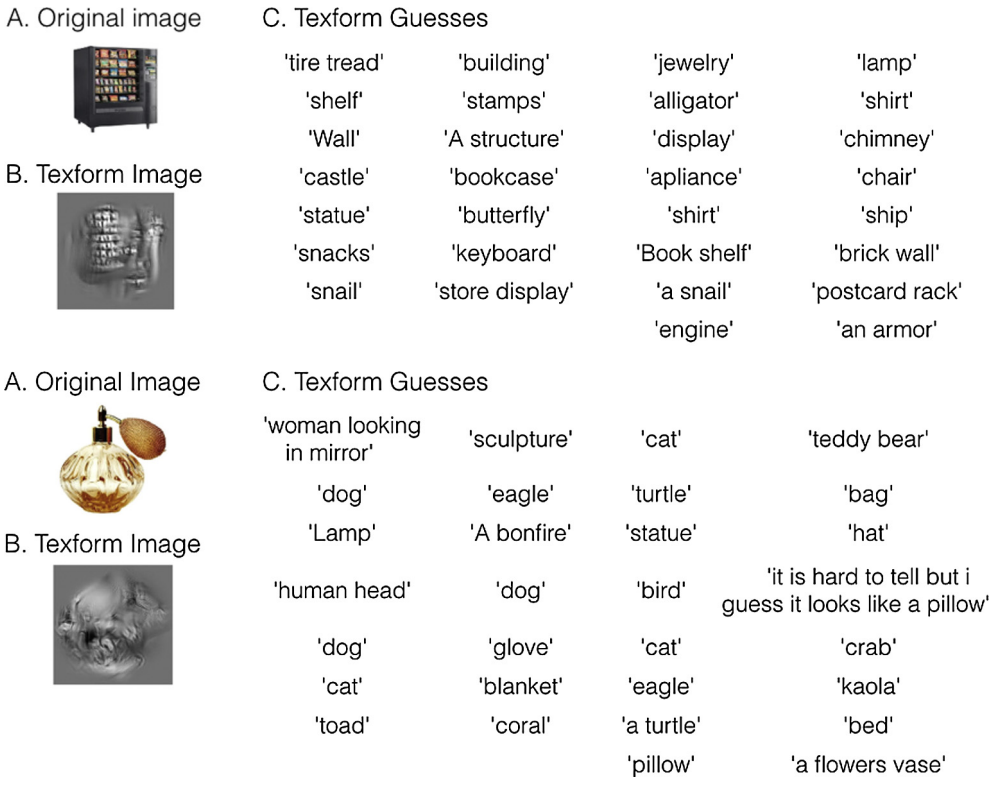

Fig. A1. Examples from the texform norming experiment for two images (upper panel, lower panel). Original images are depicted in (A) and were not shown during the norming experiment. Instead, 30 observers were shown their corresponding texforms (B), and asked to "guess what this could be"; their responses for each texform are shown in (C). Responses were coded liberally; for example, both "store display" and "bookcase" were counted as correct responses for the texform shown in B (upper panel). Average identification accuracy across all 60 texforms was 2.83%.

References

Amit, E., Algom, D., & Trope, Y. (2009). Distance-dependent processing of pictures and words. *Journal of Experimental Psychology: General, 138*(3), 400–415.

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., ... Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the United States of America, 103*(2), 449–454.

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

Cheung, O. S., & Gauthier, I. (2014). Visual appearance interacts with conceptual knowledge in object recognition. *Frontiers in Psychology, 5*.

Chiou, R., & Ralph, M. A. L. (2016). Task-related dynamic division of labor between anterior temporal and lateral occipital cortices in representing object size. *Journal of Neuroscience, 36*(17), 4662–4668.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior, 8*(2), 240–247.

DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in Cognitive Sciences, 11*(8), 333–341.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96*(3), 433–458.

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

Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nature Neuroscience, 14*(9), 1195–1201.

Gabay, S., Leibovich, T., Henik, A., & Gronau, N. (2013). Size before numbers: Conceptual size primes numerical value. *Cognition, 129*(1), 18–23.

Gliksman, S. I., Leibovich, T., Melman, Y., & Henik, A. (2016). Automaticity of conceptual magnitude. *Scientific Reports, 6*.

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

Henik, A., Gliksman, Y., Kallai, A., & Leibovich, T. (2017). Size perception and the foundation of numerical processing. *Current Directions in Psychological Science, 26*(1), 45–51.

Jolicoeur, P., Gluck, M. A., & Kosslyn, S. M. (1984). Pictures and names: Making the connection. *Cognitive Psychology, 16*(2), 243–275.

Konkle, T., & Oliva, A. (2011). Canonical visual size for real-world objects. *Journal of Experimental Psychology: Human Perception and Performance, 37*(1), 23–37.

Konkle, T., & Oliva, A. (2012). A familiar-size Stroop effect: Real-world size is an automatic property of object representation. *Journal of Experimental Psychology: Human Perception and Performance, 38*(3), 561–569.

Long, B., Konkle, T., Cohen, M. A., & Alvarez, G. A. (2016). Mid-level perceptual features distinguish objects of different real-world sizes. *Journal of Experimental Psychology: General, 145*(1), 95–109.

Macé, M. J. M., Joubert, O. R., Nespoulous, J. L., & Fabre-Thorpe, M. (2009). The time-course of visual categorizations: You spot the animal faster than the bird. *PLoS ONE, 4*(6), e5927.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Reason, 4*(2), 61–64.

Paivio, A. (1975). Perceptual comparisons through the mind's eye. *Memory & Cognition, 3*(6), 635–647.

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

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*(3), 382–439.

Rubinsten, O., & Henik, A. (2002). Is an ant larger than a lion? *Acta Psychologica, 111*(1), 141–154.

Sellaro, R., Treccani, B., Job, R., & Cubelli, R. (2015). Spatial coding of object typical size: Evidence for a SNARC-like effect. *Psychological Research Psychologische Forschung, 79*(6), 950–962.

Simoncelli, E. P., & Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annual Review of Neuroscience, 24*(1), 1193–1216.