VISUAL COGNITION
2019, VOL. 27, NOS. 5–8, 435–451
https://doi.org/10.1080/13506285.2019.1664689

Animacy and object size are reflected in perceptual similarity computations by the preschool years

Bria Long[a,b], Mariko Moher[c], Susan E. Carey[b] and Talia Konkle[b]

[a]Department of Psychology, Stanford University, Stanford, CA, USA; [b]Department of Psychology, Harvard University, Cambridge, MA, USA; [c]Department of Psychology, Williams College, Williamstown, MA, USA

ABSTRACT
By adulthood, animacy and object size jointly structure neural responses in visual cortex and influence perceptual similarity computations. Here, we take a first step in asking about the development of these aspects of cognitive architecture by probing whether animacy and object size are reflected in perceptual similarity computations by the preschool years. We used visual search performance as an index of perceptual similarity, as research with adults suggests search is slower when distractors are perceptually similar to the target. Preschoolers found target pictures more quickly when targets differed from distractor pictures in either animacy (Experiment 1) or in real-world size (Experiment 2; the pictures themselves were all the same size), versus when they do not. Taken together, these results suggest that the visual system has abstracted perceptual features for animates vs. inanimates and big vs. small objects as classes by the preschool years and call for further research exploring the development of these perceptual representations and their consequences for neural organization in childhood.

ARTICLE HISTORY
Received 7 March 2019
Accepted 13 August 2019

KEYWORDS
Object categorization; visual search; animacy; object size

Introduction

By adulthood, our visual system rapidly and efficiently categorizes the objects that we see, allowing us to access a wide array of information about any given object. For example, as soon as we recognize a pictured object, we know both whether it is an animal or an object and how big or small it is in the real world (Grill-Spector & Kanwisher, 2005; Konkle & Oliva, 2012a; Thorpe, Fize, & Marlot, 1996). How does the visual system compute these broad conceptual properties of objects (e.g., Is this alive?)? Intuitively, objects from a particular broad category (e.g., all inanimate objects) can come in so many different shapes that they may not share consistent perceptual features, and thus these properties may be computed solely at a semantic, non-visual level of representation. Contrary to this intuition, our recent work has shown that there are consistent perceptual features related to shape and texture that underlie the dimensions of animacy and real-world size (Long, Konkle, Cohen, & Alvarez, 2016; Long, Störmer, & Alvarez, 2017; see also Levin, Takarae, Miner, & Keil, 2001).

For example, in previous work we have used visual search performance as an index of perceptual similarity (e.g., Long et al., 2016), as search for a target is slower when targets and distractors are perceptually similar, and search for a target is faster when targets and distractors are perceptually dissimilar (Duncan & Humphreys, 1989). Under this logic, if a given animal (e.g., a cat) is perceptually more similar to other animals (e.g., horses, dogs, bees) than inanimate objects (e.g., headphones, cups, staplers) then it should be harder to find an animal among other animals than among objects (and vice versa). In a series of behavioural experiments in adult participants, we found exactly this pattern of results for both the animacy distinctions (Long et al., 2017) and real-world object size distinctions (Long et al., 2016). Using this visual search paradigm allowed us to link these effects largely to perceptual differences (rather than semantic differences), as visual search speeds are well-known to be primarily if not exclusively influenced by mid-level perceptual features (shape, curvature, colour) rather than semantic features (e.g., for reviews, see Rosenholtz, Huang, & Ehinger, 2012; Wolfe, 1994; Wolfe & Horowitz, 2017; but see Telling, Kumar, Meyer, & Humphreys, 2010). We empirically validated this perceptual locus by showing that

CONTACT Bria Long bria@stanford.edu

these visual search advantages persist even when adults are searching for versions of objects that preserve some mid-level perceptual texture and form information yet are unrecognizable at the basic-level (i.e., "texforms", see Long et al., 2016, 2017; Freeman & Simoncelli, 2011, see Appendix, Figure 1). Thus, the adult perceptual systems are readily sensitive to the systematic mid-level perceptual differences that differentiate animates versus inanimates, and big versus small inanimate objects.

In the present paper, we investigate the development of these visual search phenomena, asking whether animacy and object size are reflected in perceptual similarity computations in preschool-aged children. How and when do children come to know, for example, that animals tend to have different textural statistics than inanimate objects (Banno & Saiki, 2015; Long et al., 2017) and that big objects tend to be boxier than small objects (Long et al., 2016; Long & Konkle, 2017)? On one hand, many aspects of visual recognition continue to mature gradually throughout middle childhood (Dekker, Mareschal, Sereno, & Johnson, 2011; Mash, 2006; Bova et al., 2007; for a review see Nishimura, Scherf, & Behrmann, 2009), and so children may not have had sufficient experience with enough animals and objects of different sizes to have extracted the appropriate perceptual representations. On the other hand, even very young infants can readily identify the animacy of entities in their environment (Muentener & Carey, 2010; Saxe, Tenenbaum, & Carey, 2005; Simion, Regolin, & Bulf, 2008) on the basis of perceptual cues (e.g., biological motion, the presence of eyes). Thus, we hypothesized that the visual systems of pre-school aged children have likely had sufficient input to discover (perhaps via statistical learning mechanisms, e.g., Bulf, Johnson, & Valenza, 2011; Kirkham, Slemmer, & Johnson, 2002) the mid-level perceptual features that predict whether something is an animal vs. an object.

With respect to real-world object size, there is some evidence that even newborns can identify the actual size of objects they are attending to (Slater, Mattock, & Brown, 1990) and young infants often try to grasp small manipulable objects well before they can successfully do so. By age 2, children can say when an object is "big" or "little" with respect to other objects of the same kind (e.g., mittens), indicating that they do represent the sizes of some object kinds (Ebeling & Gelman, 1988; Gelman & Ebeling, 1989). Thus, one might expect parallel results for the animacy and object size dimensions in the preschool years, as in adulthood. However, in contrast to animacy, there is no evidence that infants can infer the real-world size of an object based solely on the perceptual information available in a picture. One might also expect these perceptual representations to develop only as children learn to navigate on their own and to mature slowly as children's own body size changes dramatically. Thus, it was an open question whether the visual systems of pre-school age children are already sensitive to the visual features that discriminate small manipulable objects from large objects relevant for navigation.

To test these hypotheses, we adapted the visual search experiments run with adults for use with preschoolers (Long et al., 2016, 2018) by converting them to touch-screen games. If children's perceptual systems are sensitive to the perceptual features that distinguish a given high-level distinction, then we should expect to see speeded visual search on *mixed* displays, when the distractors differ from the target in category membership (e.g., a picture of a cup among five pictures of animals) versus on *uniform* displays, when they do not (e.g., a picture of a cup among five pictures of other inanimate objects, see Figure 1). Across two experiments, we thus compared how quickly children search for targets on mixed versus uniform displays for animacy and real-world size.

To our knowledge, this is the first set of experiments to examine these kinds of category-level search benefits in early childhood. In most visual search experiments with preschoolers, the item children are searching for remains constant across the experiment or within blocks (e.g., a red teddy bear) (e.g., Gerhardstein & Rovee-Collier, 2002; Pailian, Libertus, Feigenson, & Halberda, 2016; Vales & Smith, 2015). In contrast, here children were required to search for a different target object on each trial in order to estimate these category-level search benefits. Additionally, prior work has examined how visual search is influenced by verbal categorical labels, which can indeed speed visual search performance in both children and adults (Lupyan & Spivey, 2010; Vales & Smith, 2015). In contrast, here children were simply shown an object and asked to find a match for it among the distractors; the category membership of

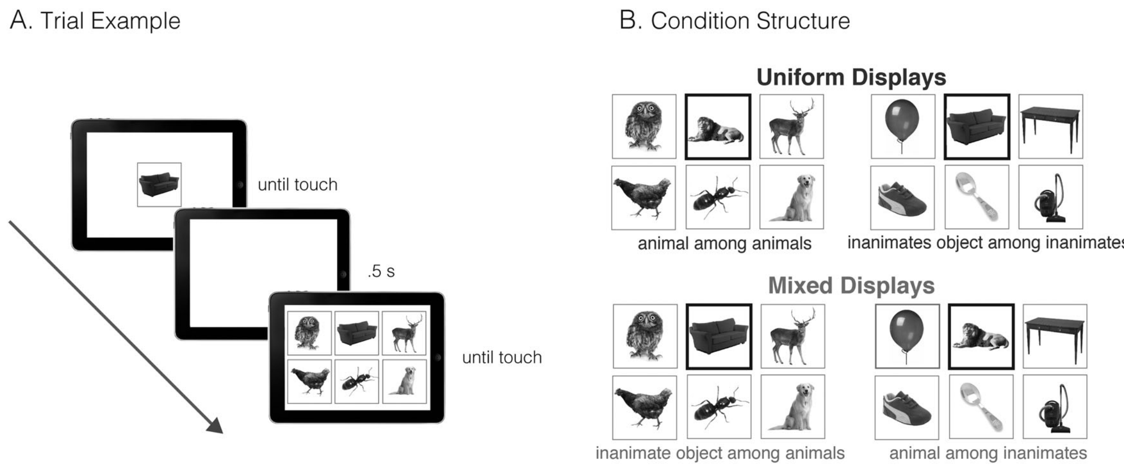


Figure 1. A. An example trial is shown. B. Condition structure for Experiment 1. Uniform displays occurred when target and distractor objects were from the same category; mixed displays occurred when the target differed from the distractors in category membership. Note that stimuli are presented on a white background here for visibility but were presented on a grey background during the experiment (controlled for luminance and contrast, Willenbockel et al., 2010). Targets are outlined in black and distractors are outlined in grey for illustration purposes.

the depicted objects was never mentioned and was completely task-irrelevant.

Experiment 1: Animacy

Experiment 1 explored whether preschool children, like adults, use perceptual features to distinguish animate from inanimate objects in a task where the distinction between these two classes is irrelevant. We investigated whether preschool-aged children are slower to find a target animal among distractor animals than among distractor inanimate objects, and, similarly, whether they are slower to find a target inanimate object among other inanimate object distractors than among distractor animals.

Methods

Participants

Sixteen 3- and 4-year-olds (7 males, $M_{age} = 48.0$ months, $SD_{age} = 8.1$ months) were recruited and participated. This sample size was chosen in advance so we would have approximately twice as many participants as needed to observe the effect in adults (based on pilot data from our laboratory). Parents gave informed consent prior to participation. One additional child started the experiment but did not finish the practice trials. All procedures were approved by the Institutional Review Board at Harvard University, Protocol: 23997, "Development of visual cognitive abilities."

Experiment set-up

The experiment was run in a web-browser (Safari) presented on a touch-screen iPad. All code was written in Javascript using the JQuery toolbox. Reaction time, touch position, accuracy, and experimental details were recorded and saved after each trial to an online database.

Procedure

Children were allowed to hold the iPad, or the experimenter held the iPad. If the children were leaning so that their arm was constrained or they were touching the iPad with both hands, the experimenter ask them to sit up straight and re-demonstrated the proper way to touch the iPad for this task.

Experiment design

On each trial, a target picture was displayed in the center of the screen. Children were instructed to touch the target picture to begin the game. After they touched the target, it disappeared for 500ms and then reappeared among distractors. Children were encouraged to "touch the same one" or to "find the match." When children touched the target object, one of six reward sounds were played, and all of the objects disappeared. When children touched a distractor object, the objects simply disappeared, and no reward sound was played. This trial design is shown in Figure 1A.

The experiment had two phases. During the practice trials, the target item appeared in a display with only two other distractors. These 3 items were randomly positioned in the 6 possible locations on the screen. The child completed practice trials until they had achieved 10 correct practice trials. The purpose of these trials was to familiarize the child with the task procedure in an easy search display. Afterwards, the experimental phase started, in which the target appeared among five other distractors, filling the entire display. The experimenter monitored how engaged children were with the task, periodically saying "good job!" to encourage on-task performance. If the child seemed disinterested and no longer engaged in the task, the experimenter would ask the child if they wanted to continue playing the game. Children continued playing until they expressed a desire to stop or completed the entire possible set of trials (96 trials).

The condition structure of the experiment is shown in Figure 1B. Target items were either animals or objects and were displayed among distractor items that were all either animals or inanimate objects. This yielded two kinds of displays: mixed displays, in which the target and distractors differed in animacy; and uniform displays, in which the target and distractors were all either animate or inanimate. Trials were pseudo-randomized such that each of four conditions (each combination of display type (mixed/uniform) and target category (animate/inanimate)) appeared once every four trials. If participants completed the entire set of possible trials (96 trials), each item appeared equally often as a target in both conditions.

Stimuli

Twenty-four pictures of animals and inanimate objects (48 images total) were selected to broadly span the categories of animate and inanimate items, excluding human faces and bodies. Animal stimuli included mammals, fish, insects, and birds; inanimate stimuli included vehicles, furniture, small manipulable objects such as cups and keys, and food items. Stimuli were selected to be objects that were likely familiar to most 3-year-olds.

All of these stimuli were presented in grey scale and matched along a variety of low-level visual features: contour variance, object extent, image area, and aspect ratio. 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 object's contour, as previous research indicates that contour variance may influence visual search performance (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 pixels within a square frame) and aspect ratio (max height / max width in the picture plane). Stimuli were matched such that animates and inanimate objects did not differ in these features (two-sample t-tests, all $p > .1$). Note that it is likely that these features we controlled for may meaningfully covary with animacy, but by matching them we help situate the effect beyond these more well-established perceptual features that are known to influence visual search times (see Wolfe & Horowitz, 2017). Finally, the two sets of stimuli were selected, they were equalized for average contrast and luminance using the SHINE toolbox (Willenbockel et al., 2010); the luminance of the background was determined by the average luminance of non-background pixels. Example stimuli are shown in Figure 2A; all stimuli are shown in Appendix A and are publicly available on the GitHub repository for this paper.

Analysis

For accuracy analyses, all test trials were analyzed. For reaction time analyses, we excluded incorrect trials (16.1% of all trials). We subsequently excluded reaction times slower than 4 seconds to eliminate trials where children were off task (5.1% of correct trials, $M = 1.3$ trials per subject). Two subjects were excluded from the reaction time analysis because they did not have more than one speeded, correct trial in each of the four conditions after outlier rejection, leaving us with a total of 14 subjects for this analysis. On average, these 14 children contributed 27.1 trials ($SD = 12.4$ trials; Uniform trials, $M = 13.5$ trials, Mixed trials, $M = 13.6$ trials) to the analysis. Average percent correct and reaction times were analyzed using 2-way repeated-measures ANOVAs with target category (animals, inanimate objects) and display type (mixed animacy, uniform animacy) as factors. These analyses were confirmed using linear mixed effect models with logged reaction times modelled as a fixed effect of target category and display type with random intercepts for subjects and target items (using lmer package in R, Bates, 2005; significance in these models were calculated using the lmerTest

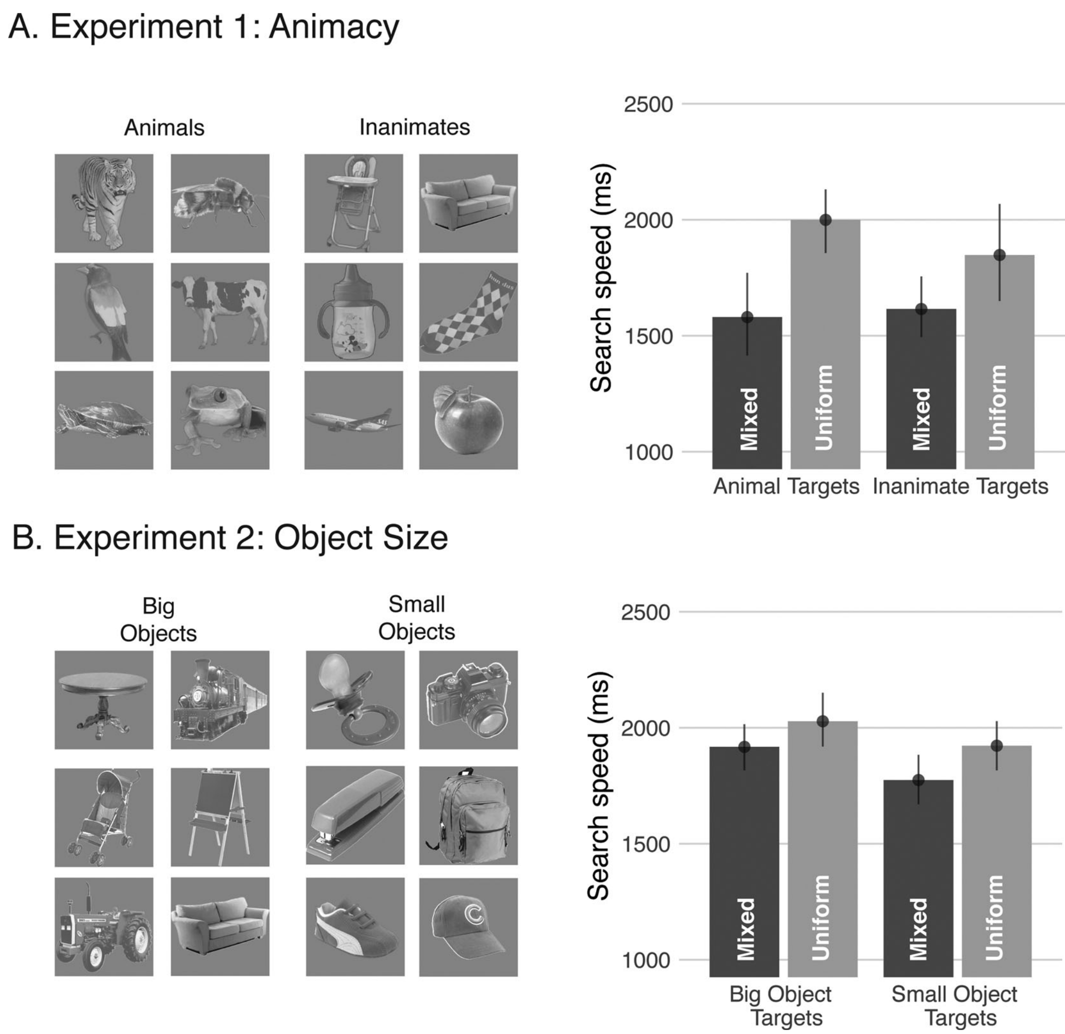


Figure 2. A. Results are shown for both experiments, considering animacy in Experiment 1 (A) and size in Experiment 2 (B). The left column shows example stimuli, which are depicted here at a higher contrast than was used in the experiment for visualization purposes. The right columns show the data graphs, where visual search speed is plotted for uniform displays (black bars) and mixed displays (grey bars). The data are additionally broken down by target-category. Error bars represent bootstrapped 95% confidence intervals.

package in R, Kuznetsova, Brockhoff, & Christensen, 2017; see Supplemental Materials). These supplemental analyses ensure the robustness of these effects after directly accounting for different numbers of trials contributed by different participants and search items. All raw data and analysis code are publicly available on the OSF repository for this manuscript (https://osf.io/d5uzg/).

Results

We first asked whether children understood the task instructions and could successfully perform the task. We found that children performed accurately and equally well on mixed and uniform animacy trials (uniform animacy, $M = 86.3\%$ correct, mixed animacy, $M = 85.0\%$ correct, $F(1, 13) = .21$, $p = .65$, $\eta_p^2 = 0.02$), and also equally well whether the target image depicted an animal or an object (animate target, $M = 83.7\%$ correct, object target, $M = 87.6\%$ correct, $F(1,13) = 2.19$, $p = 0.16$, $\eta_p^2 = 0.14$).

The main question of interest is in children's reaction times: do children find targets more quickly on mixed animacy displays (when distractors differed from the target in animacy) versus on uniform animacy displays (when targets and distractors were from the same animate category). Indeed, search speeds were faster when the distractors differed from the target in animacy (Figure 2A; uniform animacy, $M = 1924$ ms, $SD = 308$ ms, mixed animacy: $M = 1598$ ms, $SD = 363$ ms, main effect of display type; $F(1, 13) = 45.9$, $p < .001$, $\eta_p^2 = 0.78$, Cohen's $dz =$

1.57).[1] No other main effects or interactions reached significance (animal targets, $M = 1790$ ms, object targets, $M = 1731$ ms, no main effect of category; $F(1, 13) = 1.08$, $p = .32$, $\eta_p^2 = 0.08$, no interaction between category and condition; $F(1, 13) = 1.80$ $p = .20$, $\eta_p^2 = 0.12$). These results were verified using linear mixed-effect models, which confirmed that they generalize both over variation in participants and stimulus items (main effect of display type, $B = .22$ $SE = .05$, $t = 4.5$, $p < .001$).

In sum, children found targets more quickly on mixed animacy displays, when the distractors differed from the target object in animacy, than on uniform animacy displays, when the targets and distractors were either all animals or all inanimate objects. Even though the animacy of the targets and distractors was task-irrelevant and children were never asked anything about the depicted items, this factor impacted children's ability to find depicted objects. Given the reliance of this visual search paradigm on mid-level perceptual features in adulthood (Long et al., 2016, 2017) these results suggest that preschoolers visually encode these depicted objects with perceptual features that naturally distinguish animates from inanimate objects.

Experiment 2: Real-world object size

Experiment 2 asks whether preschoolers also visually encode depicted objects using perceptual features that naturally distinguish big objects and small objects. Here, we test whether preschool children, like adults, also find depicted objects more quickly when they are looking for a small object target (e.g., a picture of a cup) surrounded by big object distractors (e.g., pictures of desks, couches, and cars) and vice versa, than when targets and distractors are all of the same size in the real world (e.g., all pictures of big objects or pictures of small objects). Note that while we vary the *real-world* sizes of the depicted objects, all objects are presented at the same *visual* size on the screen, and the real-world size of the depicted objects is completely task-irrelevant.

It might be expected that the effect in Experiment 2 will be smaller than that observed for the animate/animate distinction in Experiment 1, given the evidence that infants already use certain perceptual features to pick out novel animates, construed as causal, intentional, and communicative agents (e.g., Luo, Kaufman, & Baillargeon, 2009; Muentener & Carey, 2010; Schlottmann & Ray, 2010; Simion et al., 2008). In contrast, there is no such evidence that infants use perceptual features to identify whether a novel object depicted in a picture or on a screen is smaller or larger than a breadbox. We therefore aimed to include approximately three times as many participants in Experiment 2 as in Experiment 1, which then should provide us with the power to detect whether the effect is already present in preschoolers.

Methods

Participants

Sixty 3- and 4-year-olds ($M_{age} = 45.5$ months, $SD_{age} = 6.6$ months, 18 males) were recruited at a Children's Museum and participated in this experiment; another eight children started the task but did not finish the practice trials. Using the same exclusion criterion as Experiment 1 (see Data Analysis) resulted in a final sample size of 46 children, which post-hoc sensitivity analyses suggest should be enough power to detect an effect with $dz = .46$ (90% power, one-tailed t-test, Faul, Erdfelder, Lang, & Buchner, 2007). All parents gave informed consent prior to the experiment.

Experimental set-up

The coding environment and iPad were the same as for Experiment 1. For Experiment 2, we also standardized the way children interacted with the iPad by having children sit at a small table across from an experimenter who held the iPad for them.

Procedure

The trial design and procedure were the same as in Experiment 1, except that only one reward sound was played in order to reduce potential confusion about the purpose of the sound. The experimental design was the same as in Experiment 1, except that we reduced the number of stimuli to 20 images per category to potentially increase the number of children that would complete the full counterbalanced set of trials (80 trials). Target items were pictures of small objects or big objects and were displayed among distractor items that were all either pictures of small objects or pictures of big objects. All stimuli were presented at the same size on the screen. This yielded two kinds of displays: mixed-size displays, in which the target and distractors differed in real-world size; and uniform-size displays, in which the target and distractors were all small objects or all big objects. Trials were pseudo-randomized such that each of four conditions (each combination of display type (mixed/uniform) and target category (small objects/ big objects)) appeared once every four trials. Items appeared equally often as a target in both conditions across all trials.

Stimuli

Twenty small objects and twenty big objects that were familiar to children (40 images total) were selected to broadly span each category (see examples in Figure 2B; all stimuli are in Appendix A). Small objects were typically table-lamp sized and smaller; big objects were chair-sized or larger. Small objects were chosen to have a canonical orientation (Palmer, Rosch, & Chase, 1981), and buildings were excluded from the set of big objects. All stimuli were equalized across luminance and contrast (SHINE toolbox, Willenbockel et al., 2010) and matched such that they did not differ in the same low-level visual features controlled for in Experiment 1 (area, aspect ratio, object extent, and contour variance; two-sample t-tests, all $p > .1$).

Data analysis

On average, children completed 38.0 test trials (range: 2 to 69). As in Experiment 1, we excluded incorrect trials (22.0% of test trials) and trials with RTs slower than 4 seconds (10.5% of correct test trials). We excluded 14 children who did not complete more than one speeded, correct trial in each of the four counterbalanced conditions; the exclusion rate here (23%) was not markedly greater than that of Experiment 1 (19%), suggesting that running this experiment in a Children's Museum (rather than in the lab) did not introduce more noise into the results. This left us with 46 children who contributed, on average, 34.3 trials to analysis ($SD = 16.5$ trials). As in Experiment 1, we analyzed these data using a repeated-measures ANOVA with target category (big objects, small objects) and display type (mixed size, uniform size) as factors, and these effects were confirmed using a linear mixed-effect model with the same specifications as in Experiment 1.

Results

As in Experiment 1, children were just as accurate at finding targets on mixed and uniform displays (uniform size, $M = 80.9\%$, mixed size, $M = 82.6\%$, $F(1,45) = 0.88$, $p = 0.35$, $\eta_p^2 = 0.02$). While children were slightly more accurate at finding small objects relative to big objects (small objects, $M = 83.2\%$, big objects, $M = 80.2\%$, $F(1,45) = 4.01$, $p = 0.05$, $\eta_p^2 = 0.08$), there was no interaction with the display type (uniform size vs. mixed size; $F(1,45) = 0.49$, $p = 0.49$, $\eta_p^2 = 0.01$).

The main question of interest is whether children would find the target object more quickly on mixed size displays (when distractors differed from the target in real-world size), than on uniform size displays (when distractors did not differ from the target in real-world size). They did: children's search speeds were faster on mixed size displays than on uniform size displays (uniform size displays: $M = 1975$ ms, mixed size displays: $M = 1846$ ms, main effect of display type; $F(1, 45) = 9.08$, $p = 0.004$, $\eta_p^2 = 0.17$, Cohens $dz = .46$, Figure 2B).[2] Children found target pictures that depicted small objects more quickly than those that depicted big objects (small objects, $M = 1849$ ms, big objects, $M = 1973$ ms, main effect of target category; $F(1, 45) = 7.04$, $p = 0.01$, $\eta_p^2 = 0.14$). However, this did not interact with display type ($F(1, 45) = 0.18$, $p = 0.67$, $\eta_p^2 = 0.004$). We confirmed these results using linear mixed effects models (main effect of display type, $B = 0.06$, $SE = 0.03$, $t = 2.38$, $p = 0.017$).

As anticipated, we also observed that the effect size for the object size distinction (Cohens $dz = .46$) was smaller than for the animacy distinction (Cohens $dz = 1.57$) in Experiment 1. This was also true when we equalized the number of subjects in Experiment 2: across 1000 bootstrapped samples with the same sample size as Experiment 1 ($N = 16$), we found that the average effect size was comparable to the full sample (average Cohens $dz = .49$, $sd = .26$). However, future work that directly compares these two dimensions in a within-subjects experiment should seek to confirm this observation.

In sum, preschoolers found depicted objects more quickly when the distractors differed from the target in real-world size versus when they did not. This was true even though even though the depicted objects were all the same size on the screen and their real-world sizes were completely irrelevant to the task. Thus, these results suggest that preschoolers visually encode these depicted objects with perceptual features that naturally distinguish big objects from small objects.

General discussion

To assess whether animacy and object size are already reflected in perceptual similarity computations by the preschool years, as they are in adulthood, we examined if these distinctions impact how quickly preschoolers find depicted objects in a visual search paradigm. In Experiment 1, preschoolers found depicted objects more quickly when distractors differed from the target in animacy (e.g., a picture of a rabbit among pictures of a cup, a shoe, a desk, etc.) vs. when distractors depicted objects of the same animacy as the target (e.g., a picture of a rabbit among pictures of ants, horses, fishes, etc.). In Experiment 2, preschoolers found depicted objects more quickly when distractors depicted objects that differed from the target in object size (e.g., a picture of a cup among pictures of desks, cars, fridges, etc.) versus when distractors depicted objects of the same size in the real-world as the target (e.g., a picture of a cup among pictures of pens, bottles, plates, etc.). These effects emerged even though the category membership of the depicted objects was totally task-irrelevant, and children were never asked about the properties of the depicted objects.

Taken together, these results demonstrate that when preschoolers perceive depicted objects, their perceptual systems are sensitive to visual features that broadly distinguish the classes of animates and inanimates, and the classes of big objects and small objects. Below, we discuss the possible features that are enabling these visual search benefits, the implications of these results for theories of object recognition in childhood and neural organization in the preschool years. Finally, we discuss different developmental mechanisms that could lead to the emergence of these perceptual representations.

What features guide children's search behaviour?

Here, we used a variant of the visual search paradigm specifically designed to tap perceptual processing (Levin et al., 2001; Long et al., 2016, 2017), where the exact target item is previewed before the search display is presented. For this kind of paradigm, the overarching consensus is that the semantic attributes of the stimuli do *not* guide visual attention (for a comprehensive review, see Wolfe & Horowitz, 2017). Indeed, there is both psychophysical and electrophysiological evidence in adults suggest that these animacy and object size search benefits arise early in perceptual processing (Long et al., 2017) and persist when the objects themselves are unrecognizable at the basic-level (Long et al., 2016; 2017). Thus, children are likely relying predominantly on perceptual processing mechanisms during this task; this could be further examined by conducting the visual search task using stimuli which preserve the mid-level features but where the depicted objects themselves are unrecognizable.

Accepting that mid-level perceptual features drive these search benefits in children, what *are* these perceptual features? Delineating the visual features that characterize category distinctions and relating them to intuitive visual concepts is still very much an active line of inquiry. Some work suggests that a substantial portion of the perceptual differences between animates/inanimate as well as big/small objects seem to be captured by a single boxy-to-curvy axis of *perceived curvature* (Levin et al., 2001; Long & Konkle, 2017; Long, Yu, & Konkle, 2018). That is, animates tend to be curvier than inanimates, and big objects tend to be boxier than small objects (Long et al., 2016; 2017; Long & Konkle, 2017). Consistent with this idea, basic curvature differences drive visual search performance (e.g., Wolfe, Yee, & Friedman-Hill, 1992), and explain part of the mixed-animacy search benefit in adults (Long et al., 2017; Levin et al., 2001). Of course, however, curvature is not the only perceptual feature important for these distinctions. For example, for the animacy distinction both lower-level, overall textural differences (Banno & Saiki, 2015) as well as higher-level features such as animal part-typicality (Levin et al., 2001) seem to contribute to this mixed-category search benefit.

It is also worth noting that visual search performance can also be modulated by higher-level cognitive factors, including object labels. For example, when children were given an object picture cue accompanied by its label (e.g., "bed"), children found the target faster on subsequent trials than when no label was present (Vales & Smith, 2015); and this pattern of results is also evident in adults (Lupyan & Spivey, 2010). Thus, specific variants of visual search paradigms can allow one to measure both visual feature differences as well as the influence of other higher-level mechanisms and may be useful to explore how these two factors may jointly guide visual attention throughout childhood.

Implications for object recognition in childhood

Why might it be useful for the visual system to abstract these mid-level perceptual feature representations early in development? One possibility is that having these more abstract representations could simply allow the visual system to operate more efficiently. For example, the visual system could instead classify entities at a coarser level (e.g., "animal", "big object"), providing additional information without categorizing every object at a fine-grained level (e.g., "couch", "cat", "fork", e.g, Hochstein & Ahissar, 2002). This strategy could also be particularly useful if, for example, a scene contained many objects that a child did not yet have robust basic-level representations for.

A second possibility is that these perceptual representations are useful when children are learning new object categories, particularly when learning them from depicted examples in textbooks or picture books. For example, one strategy when encountering a new category (e.g., an anteater) is to encode all the salient perceptual features of this object – it's part-structure, curvature, colour, size, and so forth. However, by initially relying on critical perceptual features that diagnose it as an animal, big object, or small object, learners may need to encode fewer specific features about this object category.

Finally, another possibility is that distinctions in the earliest perceptual processing stages might facilitate high-level or semantic processing, speeding access to animacy and size properties. Related to this possibility, some of our recent work has shown that the mid-level perceptual features that distinguish objects of different sizes can automatically feed forward to activate real-world size information, without even requiring basic-level recognition. For example, using a Size-Stroop paradigm, Konkle & Oliva (2012a) found that adults were better at making a visual size judgment (e.g., which of two objects is bigger, *on the screen*) when their relative real-world sizes were congruent with their relative pictured sizes (e.g., a small picture of a cup and a big picture of a car), and this effect persisted when the object pictures were texturized in such a way that preserved mid-level features but rendered them unrecognizable (Long & Konkle, 2017).

Consistent with adults, preschool-aged children also show the Size-Stroop effect with intact pictures of objects (Long, Moher, Carey, & Konkle, 2019); it is an open empirical question whether preschoolers would also show this effect with the unrecognizable texforms. However, some aspects of these children and adult datasets hint at similar underlying perceptual mechanisms in both groups: first, the magnitude of the Size-Stroop effect did not vary according how well preschoolers could recognize the depicted objects, and second, adults and children showed similar Size-Stroop effects across different pairs of big and small objects (Long et al., 2019). Together, these results point towards the idea that early access to these mid-level representations may render the process of visual recognition more efficient throughout development.

Implications for neural organization in early childhood

Recent research has established that while objects have many different properties, it is their animacy and real-world size that structure the large-scale organization of object-selective cortex (e.g., Chao, Martin, & Haxby, 1999; Julian, Ryan, & Epstein, 2016; Konkle & Caramazza, 2013; Konkle & Oliva, 2012b). That is, there are large zones of the cortical sheet that respond more to pictures of animals (regardless of their real-world size, e.g., fish, gorillas), big inanimate objects (e.g., cars, couches), and small inanimate objects (e.g., cups, pens). Further, this topographic organization by high-level animacy and size properties are strongly related perceptual differences in their mid-level features, rather than semantic differences per se (Long et al., 2018). Finally, recent work in adults suggests that the neural similarity of the category representations in object-selective cortex are actually highly correlated with the visual search speeds between categories (Cohen, Alvarez, Nakayama, & Konkle, 2017). Thus, the fact that preschoolers show visual search advantages for both the animacy and object size distinctions, suggests that animacy and object size may already jointly structure the large-scale organization of object-selective cortex by the preschool years.

In fact, these large-scale organizations could be in place much earlier in development. Recent work demonstrates that infants already show differential responses to faces and scenes in occipito-temporal cortex by 4–6 months of age (Deen et al., 2017), suggesting that basic mid-level feature maps could already be established in this cortex. However, these face-preferring regions in infants do not yet differentiate faces from objects, as in the adult brain, which points towards a substantial role for neural maturation, perceptual experience, or some combination of both (Deen et al., 2017; see also Gomez, Natu, Jeska, Barnett, & Grill-Spector, 2018). Understanding when and how these response selectivities emerge and mature, their relationship to perceptual behaviours, and the degree to which they are shaped by viewing, manipulating, and interacting with objects remain rich topics for future developmental neuroimaging work.

One interesting possibility is that we may be able to exploit the link between perceptual similarity among categories as measured through visual search behaviour, and neural pattern similarity among categories as measured with fMRI (e.g., Cohen et al., 2017). For example, objects that differ in animacy and object size drive substantially different responses across object-selective cortex, even when they are unrecognizable at the basic-level (Long et al., 2018). This raises the interesting possibility that assessing what distinctions are evident in visual search in kids at different ages could be a proxy into how their visual system is developing.

How do children acquire these perceptual feature mappings?

What mechanisms could lead to the development of these perceptual representations? We see three, non-mutually exclusive possibilities. First, innate templates might exist which specifically pick out some of the relevant features. For example, there could be schemas that specify some set of features for animals (e.g, biological motion; Bardi, Regolin, & Simion, 2011; Simion et al., 2008) navigationally relevant information (e.g., long, extended surfaces; Lee & Spelke, 2010) and perhaps even for manipulable entities (e.g., elongated, smooth objects; Almeida et al., 2014). Second, these features could be learned by statistical learning mechanisms that operate over children's visual and haptic experience with animals and with inanimate objects of different sizes. Third, these features could also be derived from statistical learning mechanisms that operate over kind-based object representations (e.g., "dog," "cat," "fish"). For example, the handle on a cup may be more salient within an object kind representation, where it has a distinct functional role, compared to when a cup is processed simply as a spatiotemporally contiguous object. All three of these mechanisms could combine to build these perceptual representations and could do so in very similar or rather different ways for the animacy and object size distinctions.

These perceptual representations may also continue to be refined well beyond the preschool years. Indeed, some work suggests that high-level object recognition abilities appear to mature gradually throughout middle-childhood (for reviews, see Jüttner, Wakui, Petters, & Davidoff, 2016; Nishimura et al., 2009), raising the possibility that children's perceptual representations for the animacy/size distinctions may also be refined with continued experience. For example, children may learn which kinds of shapes are most typical of certain animals, and this information may be integrated into their animacy representations. An open question is the degree to which these kinds of refinements may be related to the category-search benefits observed here, which appear dependent on more basic, mid-level perceptual processes (Long et al., 2016; 2017). Thus, research that examines these perceptual representations both in the present paradigm as well as across different kinds of tasks (e.g, animacy/size categorization) would paint a more complete picture of how these representations emerge and mature. Uncovering this developmental trajectory is clearly important from a purely developmental perspective but would also elucidate why these particular distinctions—animacy and object size—impact our perceptual systems in such similar ways.

Notes

1. We found the same pattern of effects when we changed the RT trimming procedure to only exclude trials with RTs>10 s: main effect of condition, $F(1,13) = 27.77$, $p < .001$).

2. We found the same pattern of effects when changed the reaction time trimming analysis to only exclude trials with RTs>10 s: main effect of condition, $F(1,48) = 5.64$, $p = 0.02$).

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Almeida, J., Mahon, B. Z., Zapater-Raberov, V., Dziuba, A., Cabaço, T., Marques, J. F., & Caramazza, A. (2014). Grasping with the eyes: The role of elongation in visual recognition of manipulable objects. *Cognitive, Affective, & Behavioral Neuroscience*, *14*(1), 319–335. doi:10.3758/s13415-013-0208-0

Banno, H., & Saiki, J. (2015). The use of higher-order statistics in rapid object categorization in natural scenes. *Journal of Vision*, *15*(2), 1–20.

Bardi, L., Regolin, L., & Simion, F. (2011). Biological motion preference in humans at birth: Role of dynamic and configural properties. *Developmental Science*, *14*(2), 353–359. doi:10.1111/j.1467-7687.2010.00985.x

Bates, D. (2005). Fitting linear mixed models in R. *R News*, *5*(1), 27–30.

Bova, S. M., Fazzi, E., Giovenzana, A., Montomoli, C., Signorini, S. G., Zoppello, M., & Lanzi, G. (2007). The development of visual object recognition in school-age children. *Developmental Neuropsychology*, *31*(1), 79–102.

Bulf, H., Johnson, S. P., & Valenza, E. (2011). Visual statistical learning in the newborn infant. *Cognition*, *121*(1), 127–132. doi:10.1016/j.cognition.2011.06.010

Chao, L. L., Martin, A., & Haxby, J. V. (1999). Are face-responsive regions selective only for faces? *Neuroreport*, *10*(14), 2945–2950. doi:10.1097/00001756-199909290-00013

Cohen, M. A., Alvarez, G. A., Nakayama, K., & Konkle, T. (2017). Visual search for object categories is predicted by the representational architecture of high-level visual cortex. *Journal of Neurophysiology*, *117*(1), 388–402. doi:10.1152/jn.00569.2016

Deen, B., Richardson, H., Dilks, D. D., Takahashi, A., Keil, B., Wald, L. L., … Saxe, R. (2017). Organization of high-level visual cortex in human infants. *Nature Communications*, *8*, 13995. doi:10.1038/ncomms13995

Dekker, T., Mareschal, D., Sereno, M. I., & Johnson, M. H. (2011). Dorsal and ventral stream activation and object recognition performance in school-age children. *NeuroImage*, *57*(3), 659–670.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*(3), 433. doi:10.1037/0033-295X.96.3.433

Ebeling, K. S., & Gelman, S. A. (1988). Coordination of size standards by young children. *Child Development*, 888–896. doi:10.2307/1130256

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. doi:10.3758/BF03193146

Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nature Neuroscience*, *14*(9), 1195–1201. doi:10.1038/nn.2889

Gelman, S. A., & Ebeling, K. S. (1989). Children's use of nonegocentric standards in judgments of functional size. *Child Development*, 920–932. doi:10.2307/1131033

Gerhardstein, P., & Rovee-Collier, C. (2002). The development of visual search in infants and very young children. *Journal of Experimental Child Psychology*, *81*(2), 194–215.

Gomez, J., Natu, V., Jeska, B., Barnett, M., & Grill-Spector, K. (2018). Development differentially sculpts receptive fields across early and high-level human visual cortex. *Nature Communications*, *9*(1), 788.

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

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. doi:10.1111/j.0956-7976.2005.00796.x

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, *36*(5), 791–804.

Julian, J. B., Ryan, J., & Epstein, R. A. (2016). Coding of object size and object category in human visual cortex. *Cerebral Cortex*, *27*(6), 3095–3109. doi:10.1093/cercor/bhw150

Jüttner, M., Wakui, E., Petters, D., & Davidoff, J. (2016). Developmental commonalities between object and face recognition in Adolescence. *Frontiers in Psychology*, *7*, 385.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35–B42. doi:10.1016/S0010-0277(02)00004-5

Konkle, T., & Caramazza, A. (2013). Tripartite organization of the ventral stream by animacy and object size. *Journal of Neuroscience*, *33*(25), 10235–10242. doi:10.1523/JNEUROSCI.0983-13.2013

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*(3), 561–569. doi:10.1037/a0028294

Konkle, T., & Oliva, A. (2012b). A real-world size organization of object responses in Occipitotemporal cortex. *Neuron*, *74*(6), 1114–1124. doi:10.1016/j.neuron.2012.04.036

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). Lmertest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. doi:10.18637/jss.v082.i13

Lee, S. A., & Spelke, E. S. (2010). Two systems of spatial representation underlying navigation. *Experimental Brain Research*, *206*(2), 179–188. doi:10.1007/s00221-010-2349-5

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*(4), 676–697.

Long, B., & Konkle, T. (2017). A familiar-size Stroop effect in the absence of basic-level recognition. *Cognition*, *168*, 234–242. doi:10.1016/j.cognition.2017.06.025

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*, *45*(1), 95–105. doi:10.1037/xge0000130

Long, B., Moher, M., Carey, S., & Konkle, T. (2019). Real-world size is automatically encoded in preschoolers' object representations. *JEP:HPP*.

Long, B., Störmer, V. S., & Alvarez, G. A. (2017). Mid-level perceptual features contain early cues to animacy. *Journal of Vision*, *17*(6), 20. doi:10.1167/17.6.20

Long, B., Yu, C. P., & Konkle, T. (2018). Mid-level visual features underlie the high-level categorical organization of the ventral stream. *Proceedings of the National Academy of Sciences*, *115*(38), E9015–E9024.

Luo, Y., Kaufman, L., & Baillargeon, R. (2009). Young infants' reasoning about physical events involving inert and self-propelled objects. *Cognitive Psychology*, *58*(4), 441–486. doi:10.1016/j.cogpsych.2008.11.001

Lupyan, G., & Spivey, M. J. (2010). Redundant spoken labels facilitate perception of multiple items. *Attention, Perception, & Psychophysics*, *72*(8), 2236–2253.

Mash, C. (2006). Multidimensional shape similarity in the development of visual object classification. *Journal of Experimental Child Psychology*, *95*(2), 128–152.

Muentener, P., & Carey, S. (2010). Infants' causal representations of state change events. *Cognitive Psychology*, *61*(2), 63–86. doi:10.1016/j.cogpsych.2010.02.001

Naber, M., Hilger, M., & Einhäuser, W. (2012). Animal detection and identification in natural scenes: Image statistics and emotional valence. *Journal of Vision*, *12*(1), 25–25. doi:10.1167/12.1.25

Nishimura, M., Scherf, S., & Behrmann, M. (2009). Development of object recognition in humans. *Biology Reports*, *1*, 56. doi:10.3410/B1-56

Pailian, H., Libertus, M. E., Feigenson, L., & Halberda, J. (2016). Visual working memory capacity increases between ages 3 and 8 years, controlling for gains in attention, perception, and executive control. *Attention, Perception, & Psychophysics*, *78*(6), 1556–1573.

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

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*. doi:10.3389/fpsyg.2012.00013

Saxe, R., Tenenbaum, J. B., & Carey, S. (2005). Secret agents: Inferences about hidden causes by 10-and 12-month-old infants. *Psychological Science*, *16*(12), 995–1001. doi:10.1111/j.1467-9280.2005.01649.x

Schlottmann, A., & Ray, E. (2010). Goal attribution to schematic animals: Do 6-month-olds perceive biological motion as animate? *Developmental Science*, *13*(1), 1–10. doi:10.1111/j.1467-7687.2009.00854.x

Simion, F., Regolin, L., & Bulf, H. (2008). A predisposition for biological motion in the newborn baby. *Proceedings of the National Academy of Sciences*, *105*(2), 809–813.

Slater, A., Mattock, A., & Brown, E. (1990). Size constancy at birth: Newborn infants' responses to retinal and real size. *Journal of Experimental Child Psychology*, *49*(2), 314–322. doi:10.1016/0022-0965(90)90061-C

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*(10), 2212–2225. doi:10.1162/jocn.2009.21348

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*(6582), 520–522. doi:10.1038/381520a0

Vales, C., & Smith, L. B. (2015). Words, shape, visual search and visual working memory in 3-year-old children. *Developmental Science*, *18*(1), 65–79.

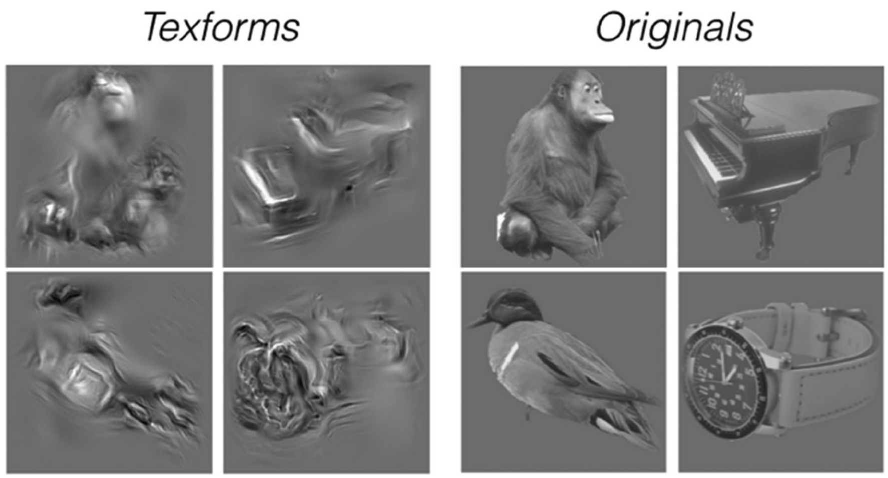
Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods*, *42*(3), 671–684. doi:10.3758/BRM.42.3.671

Wolfe, J. M. (1994). Guided search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, *1*(2), 202–238. doi:10.3758/BF03200774

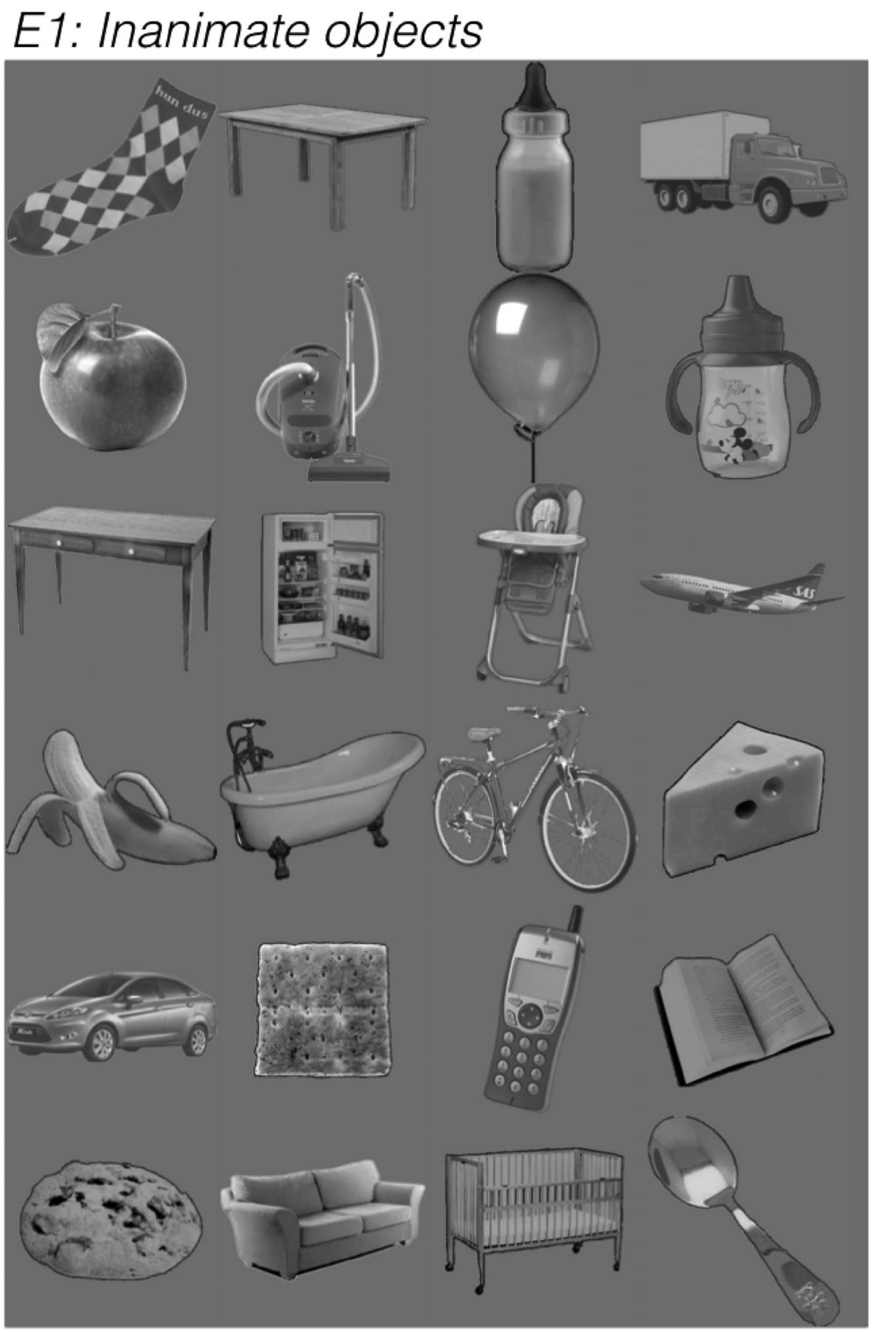
Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour*, *1*(3), 0058.

Wolfe, J. M., Yee, S., & Friedman-Hill, A. R. (1992). Curvature is a basic feature for visual search tasks. *Perception*, *21*(4), 465–480.

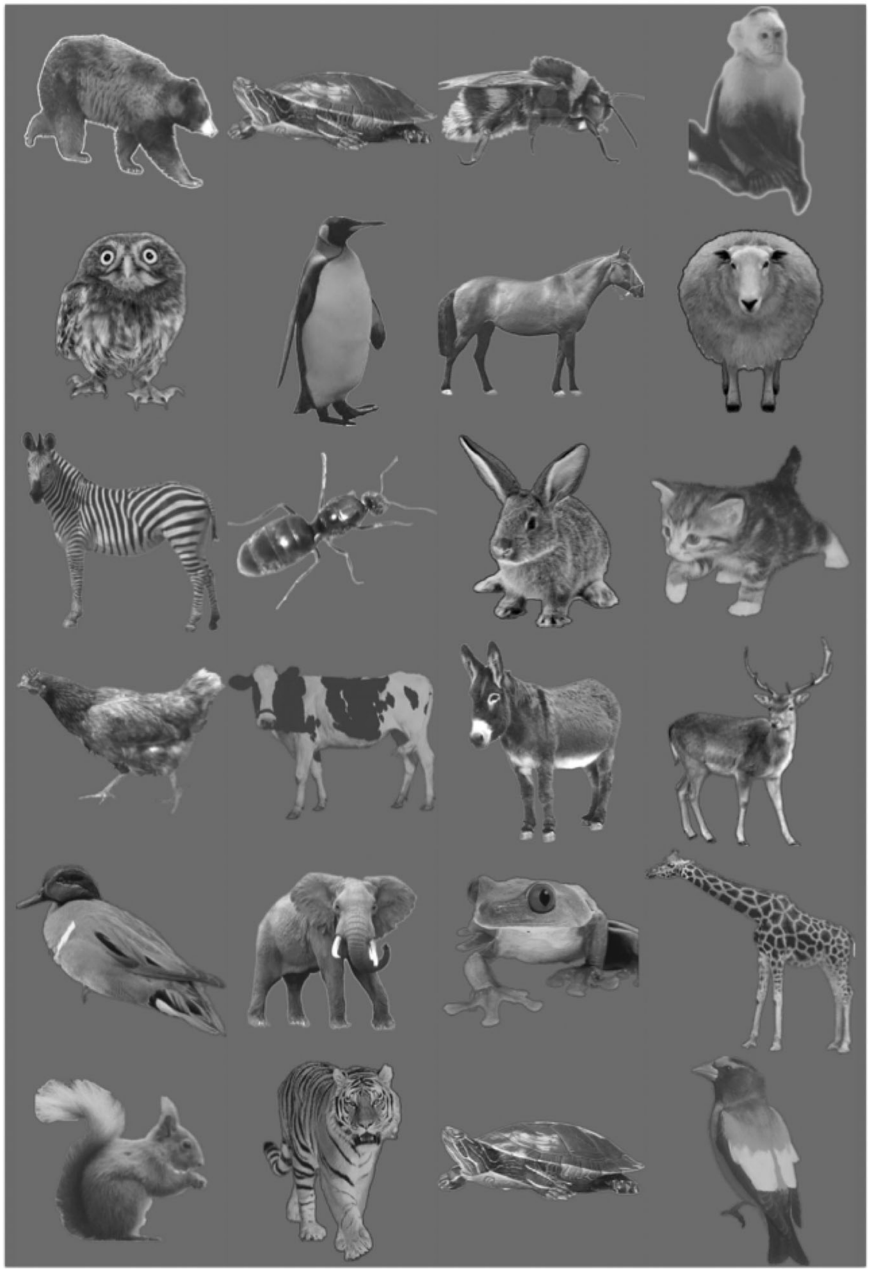
Appendix



Appendix Figure 1. Examples of texforms (left) and the original images from which they were generated. Texforms preserve the mid-level features that distinguish animates from inanimates as well as big objects from small objects (Figure adapted with permission from Long, Yu, & Knkle, 2018)


E1: Inanimate objects

E1: Animals



E2: Small objects

E2: Big objects