**Real-world size is automatically**

**encoded in preschoolers’ object representations**

Bria Long1, Talia Konkle1, Mariko Moher2, & Susan Carey1

1Department of Psychology, Harvard University

2Department of Psychology, Williams College

Paper in *XX*

Word Count: XX, including abstract

Running Head: REAL-WORLD SIZE IN PRESCHOOLERS

Corresponding author: Bria Long, brialong@fas.harvard.edu

***Research Highlights***

* Real-world size interfered with four-year-olds’ ability to make visual size judgments about pictured objects in the Size-Stroop task
* The same pairs of objects generated robust Size-Stroop effects in both adults and four-year-olds
* This was true even when four-year-olds could not name the pictured objects

***Abstract***

When adults see a picture of an object, they automatically process how big it typically is in the real world (Konkle & Oliva, 2012a). How much experience is needed for this automatic size processing to emerge? Here, we ask whether preschoolers show this same signature of automatic size processing. We showed 3- and 4-year-olds displays with two pictures of objects and asked them to touch the picture that was smaller on the screen. Critically, the relative visual sizes of the objects could either be congruent with their relative real-world sizes (e.g., a small picture of a shoe next to a big picture of a car) or incongruent with their relative real-world sizes (e.g., a big picture of a shoe next to a small picture of a car). Across two experiments, we found that preschoolers were worse at making visual size judgments on incongruent trials, suggesting that real-world size was automatically activated and interfered with their performance. In a third experiment, we found that adults and 4-year-olds showed similar Size-Stroop effects across the same pairs of big and small objects. Furthermore, whether 4-year-olds recognized the pictured objects did not influence the magnitude of the Size-Stroop effect the objects generated, suggesting that the perceptual features of these objects were sufficient to trigger the processing of real-world size information. Thus, these results suggest that, by 3–4 years of age, children automatically extract real-world size information from depicted objects.

**Introduction**

The representations of the typical sizes of objects enter into human mental life in many ways—from participating in computations of spatial layout (in their role in specifying objects’ distance from us) to providing the standards for the application of words like “big” and “small” (a small car is smaller than average for cars, but nonetheless much larger than a large cup). Further, in our everyday lives, the physical size of an object constrains how we can interact with it. We tend to pick up small objects with our hands (e.g., cups, keys) but we need to navigate around big objects (e.g., couches, tables). Interestingly, by adulthood, this particular distinction between small manipulable objects (roughly, smaller than a breadbox or a table lamp) and large objects we navigate around (roughly, chair size or larger) as classes plays a systematic role in organizing both our perceptual and neural representations of inanimate objects ( Long, Konkle, Cohen, & Alvarez, 2016; Konkle & Oliva, 2012a).

One marker of perceptual representations is performance on speeded search tasks, where search is slower when distractors are perceptually similar to the target (Duncan & Humphreys, 1989). When all images are the same size on the screen, adults took longer to find targets depicting big objects (in the real world) among distractors that were also big objects versus among distractors that were small objects, and vice versa (Long et al. 2016). Further, these search advantages persist when images are transformed into “texforms,” images that preserve some mid-level form and texture information while rendering the objects unrecognizable at the basic-level (see Appendix, Supplementary Figure 1; Long et al. 2016). Thus, this work suggests that there are systematic mid-level features that distinguish small objects vs. big objects as classes.

At a neural level, this distinction between small vs. big objects as classes also organizes responses in occipito-temporal cortex; large swaths of cortex exhibit preferences for objects that are big in the real world vs. small in the real world on both the lateral and ventral surfaces (Konkle & Caramazza, 2013; Konkle & Oliva, 2012a). As with the visual search experiments, these neural effects persist when images are transformed into texforms, supporting the mid-level nature of these neural representations (Long, Yu, & Konkle, 2017).

Furthermore, real-world size information appears to be so ingrained in our object representations that it is automatically activated when we see depicted objects (Konkle & Oliva, 2012b). In a Size Stroop paradigm, adults are asked to make a *visual* size judgment about which of two images is bigger (or smaller) on the screen, while ignoring the objects’ sizes in the real world. Critically, adults are slower and less accurate at making visual size judgments when the images’ relative visual sizes are incongruent with the relative real-world sizes of the depicted objects (i.e., a big picture of a teapot and a small picture of a gazebo) versus when they are congruent with their real-world sizes (i.e., a big picture of a gazebo and a small picture of a teapot; Konkle & Oliva, 2012b; see Figure 1). Thus, even though real-world size information is task-irrelevant, it is automatically activated and interferes with adults’ ability to make visual size judgments. Similar to the visual search findings, texforms also trigger a Size-Stroop effect in adults (Long & Konkle, 2017), suggesting that mid-level features can lead to the automatic computation of real-world size.

That real-world size representations are such an ingrained and organizing property of our object representations raises a central question in developmental science: how does such organization emerge over development, i.e., what, if any, innate support exists for it, and what learning mechanisms are involved? Answering these questions is not only an important project within developmental cognitive neuroscience, but would shed light on *how* the categorical distinction between small vs. large objects as classes becomes an organizing property of adults’ object representations. Here we take a first step in this developmental project, starting with preschoolers as they are the youngest age group likely capable of performing the same exact tasks used to study object size processing in adults (i.e., visual search and Size-Stroop paradigms). The present experiments seek to establish whether preschoolers, like adults, automatically activate real-world size information when they see pictured objects, even when that information is task-irrelevant, and, more speculatively, whether details of the data implicate mid-level perceptual processing as a locus of the observed effects.

Several lines of evidence suggest that both infants and young children encode the real-world sizes of objects. Even newborns can tell the real-word size of an object they are interacting with, and coordinate their grip appropriately (Slater, Mattock, & Brown, 1990). By around 7 months of age, infants can use the typical sizes of familiar objects (i.e., faces) as a monocular depth cue (Yonas, Pettersen, & Granrud, 1982). Further, 7-month-olds appear to learn and use this information rapidly; after playing with two small novel objects, one of which was bigger than the other, they used the information about which object was bigger as a monocular depth cue (Granrud, Haake, & Yonas, 1985). And by 2.5 years of age, 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 represent the average sizes of some categories (Ebeling & Gelman, 1988; Gelman & Ebeling, 1989).

Yet none of these studies address whether infants make a categorical distinction between small objects vs. big objects as classes. Recently, we adapted the visual search paradigms for preschoolers and found visual search advantages for both animacy and object size, suggesting that preschoolers distinguish small vs. big objects as classes (Long et al., 2016; 2017; Long, Moher, Konkle, & Carey, under review). In other words, real-world size appears to be reflected in perceptual similarity computations by the preschool years, as it is in adulthood (Long et al., 2016).

Nonetheless, even if children have encoded the real-world size of a given category, they may not automatically activate this information when they see a pictured exemplar from this category. In other words, real-world size may not yet be automatically activated when preschoolers see pictured objects. In Experiments 1 & 2, we use the Size-Stroop task to investigate whether 3- and 4-year-old children, like adults, automatically activate the real-world size of pictured objects, even when this information is task-irrelevant.

**Experiment 1: Do preschoolers show the Size-Stroop effect?**

We adapted the Size-Stroop task (Konkle & Oliva, 2012b) for children by converting it to an iPad game. Children were asked to “touch the picture that is smaller on the screen.” If preschoolers automatically activate information about objects’ typical sizes in the real world during this task, then they should be slower and less accurate on incongruent displays, in which the object that is bigger in the real world is smaller on the screen, versus on congruent displays, in which the relative visual sizes of the depicted objects is the same as their relative sizes in the real-world .

**Methods**

***Participants.*** Seventy-nine 3- and 4-year-old children participated, either at the Boston Children’s Museum, the Harvard Lab for Developmental Studies, or the Williams College Children’s Center. A parent gave consent prior to participation, and the Institutional Review Board at Harvard University approved the study. We aimed to recruit enough participants to include approximately double the number of subjects needed to observe the effect in adults in each age group (*N* = 16, Konkle & Oliva, 2012b). One child began the task but did not complete more than two trials. This left us with 78 children in the final sample, with 47 3-year-olds (*M* = 41.87 months, *SD* = 2.99 months) and 31 4-year-olds (*M* = 53.65 months, *SD* = 3.4 months).

***Experimental Set-Up.***  Children sat at a table across from an experimenter who held an iPad for them. The experimenter could not see the images on the screen, and was thus blind to condition. Experiments were run on an iPad in a web-browser (Safari) and custom 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.

***Stimuli.*** Images were identical to those used in Experiment 1B of Konkle & Oliva (2012b); these images of 20 big objects and 20 small objects were matched in terms of their overall area and paired by their vertical height. The same pairs of big and small objects were always presented together on both congruent and incongruent trials (see Figure 2A).

***Procedure.*** There were two phases to the experiment. First, practice trials verified that the child could make visual size judgments about geometric shapes. Next, there was a test phase where children made visual size judgments about two pictured objects.

The first 35 out of 80 children received a paper version of the practice phase. These children were presented with two different colored shapes, one of which was bigger than the other, and were asked to “Touch the shape that is smaller *on this paper*.” All 35 children completed five correct practice trials. However, as several children were distracted by the appearance of the iPad for the test phase, the remaining children completed the practice

phase on the iPad. These 45 children completed nine correct practice trials before the test phase.[[1]](#footnote-1) Here, children touched a blue dot to begin each trial, after which they were presented with two different colored shapes, one of which was bigger than the other. Children were asked to “Touch the shape that is smaller *on the screen*.” These last three words (“on the screen”) were emphasized to clarify any ambiguity in these instructions. When children answered correctly, the iPad played a pleasant sound and advanced to the next practice trial. The experimenter also reinforced on-task performance by saying “good job!” when children selected the correct target.

In the test phase, at the beginning of each trial, all children were asked to “touch the blue dot to begin.” After children touched the blue dot, there was a brief delay of 500ms

after which two images appeared on either side of the screen. Children were asked to “Touch the picture that is smaller *on the screen*.”[[2]](#footnote-2) Critically, there were two different kinds

of trials: congruent trials, when the relative real-world sizes of the pictured objects were congruent with their relative visual sizes on the screen (i.e., a big picture of a car and a small picture of a cup) and incongruent trials, when the relative real-world sizes of the pictured objects were incongruent with their relative visual sizes on the screen (i.e., a small picture of a car and a big picture of a cup). If the child selected the correct image, a pleasant sound was played; if the child selected the incorrect image, no sound was played. In either case, the blue dot then reappeared to signal the start of the next trial. To encourage accuracy, a picture of Mickey Mouse also appeared after every 3 correct trials, and children’s progress was marked with a stamp by the experimenter.[[3]](#footnote-3) The experimenter also periodically gave positive feedback, saying “good job!,” noting how many stamps the child had acquired, and encouraging children to keep playing the game. Children continued until they completed a maximum of 80 trials or wanted to stop the experiment.

***Counterbalancing***. Each pair of big and small objects appeared in both incongruent and congruent configurations. In addition, the visually bigger object appeared on both sides of the screen, creating 4 displays per pair of objects and 80 total possible displays. Every combination of target side (right, left) and trial type (congruent, incongruent) appeared every four trials during the experiment. The image pair that occurred on a given trial was randomized throughout each session for each child.

***Analysis.***  For each child, we excluded the first ten trials from the test phase as practice trials. Children completed an average of 52.32 trials out of a possible 70, excluding these practice trials (range 4 to 70).[[4]](#footnote-4) For error analyses, we analyzed the error patterns of all 78 children. For RT analyses, we excluded incorrect trials and trials with RTs slower than 4 seconds (6.45% of correct trials). This RT cutoffs eliminates trials where children are off-task, and has previously been used as a cutoff when analyzing preschooler’s reaction times in a touchscreen-based task (Frank et al., 2016).[[5]](#footnote-5) Children were included if they had at least 5 correct trials (after the 10 practice trials) per condition (congruent, incongruent) with RTs under 4 seconds. Four children were excluded for not meeting these criteria; all children were 3-year-olds. This left us with 74 children for RT analyses: 43 three-year-olds (M = 42.09 months, SD = 2.89 months) and all of the 31 four-year-olds. On average, 3-year-olds contributed 45.7 correct, speeded trials to RT analyses, and 4-year-olds contributed 47.77 correct, speeded trials to RT analyses.

We analyzed error patterns and reaction times in two ways. First, we performed mixed-effect ANOVAs to assess the effects of the within-subjects factor of trial type (congruent versus incongruent) and the between-subjects factor of age group (3 versus 4 years) on error rates (percentage of completed trials that were errors). Post-hoc tests are reported using one-tailed t-tests, as the results are only interpretable if children performed worse on incongruent relative to congruent displays.

Then, to ensure the robustness of our results, we also ran linear mixed-effect models on log-transformed RT data (as RT data is non-normally distributed) and generalized linear mixed effect models on error patterns. Age group, congruency, and their interaction were modeled as fixed effects, and we included both random intercepts and slopes for the effect of congruency on both subjects and individual items -- the maximum random effects structure justified by our design (Barr et al., 2013). These additional analyses allow us account for the different numbers of trials completed by individual children and to ensure that our results are not biased by particular displays. All analysis code is available at the public repository for this manuscript (www.github.com/brialorelle/kidstroop).

**

*Figure 1: Stimuli and procedure used in Experiments 1–2*. A. *In congruent displays, the relative size of the objects were consistent with their size in the real world, and incongruent displays, the relative size of the two objects were inconsistent with their size in the real world*. B. *Schematic of practice and test trials.*

***Error Results***. Children made relatively few errors (*M* = 10.9%) suggesting they understood the task instructions, though 3-year-olds made more errors than 4-year-olds (main effect of age: 3-year-olds: M = 14.44%, 4-year-olds: M = 5.6%, *F*(1,76) = 6.93*, p* = 0.01, ηG2 = 0.07). In addition, children showed evidence for the Size-Stroop effect in their errors; they made more errors on incongruent than congruent displays (main effect of trial type: congruent *M* = 8.89 (SD = 15.91%), incongruent M = 13.28 (SD = 18.1%), *F*(1,77) = 7.4, p = 0.01, ηG2 = 0.02).

The Size-Stroop effect was apparent throughout this age range; there was no interaction between age group and trial type (F(1,76) = 0.01, p = 0.92, ηG2 < 0). Finally, planned ad-hoc comparisons confirmed that the Size-Stroop effect was observed at *each* age:  3-year-olds: congruent M = 12.47%, incongruent M = 16.98%, *t*(46) = 1.99, *p* = 0.03; 4-year olds: congruent M = 3.47%, incongruent M = 7.67%, *t*(30) = 2.2, *p* = 0.02; see Figure 3A. Our GLMM model confirmed these analyses, finding that this effect generalized across individual subjects (B = 0.55, SE = 0.12, Z = 4.78, p < .001).

*Figure 3*: Error and reaction time analyses from Experiment 1 and Experiment 2, a replication in 4-year-olds. Error analyses included all children; reaction time analysis included children with 5 or more correct responses in each condition after outlier rejection (see text). In Experiment 1, both 3- and 4-year-olds made more errors in the Size-Stroop task on incongruent displays. Only 4-year-olds were also slower on incongruent trials, though 3-year-olds took much longer to make visual size judgments. In Experiment 2, we replicated the same results in an independent group of 4-year-olds (though these children overall responded more quickly and made fewer errors). Error bars represent within-subjects standard error (Morey, 2008).

***Reaction Time Results****.* When we considered both 3- and 4-year-olds together, we found that, overall, children did not take longer to make visual size judgments on incongruent versus congruent displays (no main effect of trial type: congruent *M*= 1763ms, incongruent *M* = 1781ms, *F*(1,72) = 0.16, *p* = 0.69, ηG2 < 0.001)*.*  Three-year-olds took longer to make visual size judgments than four-year-olds (main effect of age group, *F*(1,72) = 11.3, *p* < .001, ηG2 = .12), though the interaction between age group and condition was not significant (*F*(1,72) = 2.87, *p* = .09, ηG2 = 0.004).

However, we planned to examine results for 3- and 4-year-olds separately, as we anticipated that 3-year-olds might not be able to perform the task as well as 4-year-olds. These planned ad-hoc tests revealed that 4-year-olds showed the Size-Stroop effect in their RTs (congruent *M* = 1555ms, *SD* = 359ms, incongruent *M* = 1622 ms, *SD* = 319 ms, *t*(30) = 2.37, *p* = 0.01, Cohen’s *d* = 0.43), while the 3-year-olds did not (congruent *M* = 1928 ms, *SD* = 465, incongruent *M* = 1887 ms, *SD* = 491 ms, *t*(41) = 0.79, *p* = 0.79, Cohen’s d = -0.13, Figure 3B). We found the same pattern of results in our linear mixed effect models, both for all children and for four-year-olds (see Supplementary Materials: analysis code).

A final exploratory analysis examined whether age or overall slowness was more likely to account for the 3-year-olds’ lack of the Size-Stroop effect on RTs. We analyzed whether children’s age (in months) predicted the degree to which children made more errors or had slower RTs on the incongruent than the congruent trials. It did not; age was uncorrelated with the size of the Stroop effect (RTs: *r* = .19, *p* = .10; Error rates: *r* = .04,, *p* =.63). We then asked whether overall RT predicted the variability of the Size-Stroop effect for all children. We found that children who performed the task more slowly were more likely to show either a very positive or a very negative Size Stroop effect (age correlation with absolute valued Stroop effects, *r* = 0.25, *p* = 0.03); in other words, children whose reaction times were longer tended to have more variance in their RTs, leading to noisier estimates of the Size-Stroop effect.

in other words, children whose reaction times were longer tended to have more variance in their RTs, leading to noisier estimates of the Size-Stroop effect.

Thus, though 3-year-olds understood the task instructions, they tended to stay less on task and did not consistently make *speeded* visual size judgments. Both of these factors would make it harder to obtain accurate estimates of 3-year-olds’ reaction times for congruent versus incongruent displays and thus to observe a Size-Stroop effect in their reaction times. Furthermore, as 3-year-olds showed the Size-Stroop effect in their error rates, the different pattern of reaction time effects across age groups is likely driven by the fact that 3-year-olds simply had more variance in their reaction times. We thus conclude that there is unlikely to be a difference in how 3- versus 4-year-olds process real-world object size in this task.

**Experiment 2: Replication**

The results of Experiment 1 suggest that real-world size was automatically activated and interfered with children’s ability to make visual size judgments in the Size-Stroop task. We found that both 3-year-olds and 4-year-olds made more errors on incongruent relative to congruent displays, and that 4-year-olds also tended to take longer to make visual size judgements on incongruent relative to congruent displays. As we did not anticipate that only 4-year-olds would show a Size-Stroop effect in their reaction times, in Experiment 2 we sought to replicate these RT results in an independent group of 4-year-olds. We also planned to exclude 4-year-olds who responded slowly from reaction time analyses, as in Experiment 1 we found that children with very slow RTs tended to show highly variable Size Stroop effects in their RTs.

**Methods**

***Participants****.* Thirty-five four-year-olds were recruited for Experiment 2 so that

approximately the same number of four-year-olds would contribute to RT analyses as in

Experiment 1. Children were recruited and participated at the Boston Children’s Museum or the Harvard Lab for Developmental Studies. One child began the task but did not complete more than two trials and was excluded from analysis. One other child participated but was excluded for parental interference, leaving us with 33 four-year-olds in our final sample (*M* = 53.27 months, *SD* = 3.20 months, 15 males).

***Experimental Setup, Stimuli, & Counterbalancing****.* All aspects of Experiment 2 were identical to those of Experiment 1, except that we encouraged children to obtain 20 stamps (i.e., 60 correct trials). We did this to maximize the number of children who could be included in RT analyses.

**Analysis and Results**

***Error Analysis.*** We analyzed error rates in all 33 children who participated. These children completed an average of 46.96 trials (range=28 to 56) out of a possible 70.

***Error Results.*** As in Experiment 1,children made more errors on incongruent displays (incongruent *M* = 3.69%, congruent *M* = 1.36%, *t*(32) = 2.55, *p* = .008), even though they made fewer errors overall when compared to 4-year-olds in Experiment 1 (see Figure 3A).

***Reaction Time Analysis.*** First, we applied the same exclusion criteria as in Experiment 1. We excluded trials where children responded incorrectly (that is, chose the visually bigger image; *M =* 2.55% of all trials) or took longer than 4 seconds to respond (*M =* 1.23% of correct trials). No children were excluded on the basis of not having 5 or more test trials with correct responses made in less than 4 seconds.

As planned, we excluded children whose average RTs (across both conditions) were slower than 2 standard deviations from the average group RT (only 2 participants; mean RTs=2029ms, 2161ms, z-scores=2.05, 2.55).[[6]](#footnote-7) After applying these inclusion criteria, set in advance, the RTs of 31 children (*M* = 53.39 months, *SD* = 3.21 months, 13 males), who completed an average of 46.87 trials, were analyzed.

***Reaction Time Results.***  As in Experiment 1, four-year-olds took longer to make visual size judgments on incongruent trials (congruent *M* = 1433ms, incongruent *M* = 1476ms, *t*(30) = 2.34, *p* = .013, Cohen’s *d*=.42, Figure 3B).[[7]](#footnote-8) Our linear mixed-effect model on logged RTs revealed the same pattern of results (*B* = .04, *SE* = .01, *df* = 1373.33, *t* =2.71, *p* = .01).[[8]](#footnote-9) Thus, these data replicate the pattern of effects seen in Experiment 1; four-year-olds exhibit a Size-Stroop effect in both their errors and reaction times.

**Experiment 3: Size-Stroop Display Effects**

Experiments 1 and 2 establish that preschool aged children resemble adults in one important aspect: when children see pictured objects, they automatically activate information about the real-world size of these objects, for these results show that the Size-Stroop effect is observable in error rates by age 3 and in RTs by age 4. However, these results leave open the exact representations and computations underlying children’s specification of real-word size of the pictured objects, and whether these are the same as those of adults.

Adults and children clearly represent the average size of objects within kinds, for they can ascertain whether a given phone, spoon, car, etc. is big or small, where this means big or small for a phone, spoon, or car (Gelman & Ebeling, 1988, Ebeling & Gelman 1989). Thus, it is possible that both children and adults automatically compute the size of the depicted objects by accessing the kinds of the objects (e.g., “cup”) and retrieving information about the average size of those objects (e.g., “can be held with one hand”). However, two results suggest that this is not how the Size-Stroop effect is generated in adults. First, as mentioned in the introduction, mid-level perceptual features can trigger Size Stroop effect in the absence of basic-level recognition; in other words, the Size-Stroop effect is observed even when the stimuli are texforms (Long & Konkle, 2017). Furthermore, brief training with novel objects that do *not* differ in mid-level features (but do differ in real-world size) does not induce a Size-Stroop effect (Konkle & Oliva, 2012b), further suggesting that mid-level features play an important role in real-world size computations.

Thus, one empirical route for evaluating whether preschoolers’ visual systems have also identified the mid-level perceptual features that distinguish big and small objects (as classes) would be to see if children show the visual search advantage and Size-Stroop effect with unrecognizable texforms. Unfortunately—as might be predicted by the fact that children rarely completed these studies with recognizable objects —pilot studies showed that children would not sit through paradigms when stimuli were meaningless blobs.

Here, we take advantage of display item effects to bring indirect evidence to bear on the question of whether similar computational mechanisms underlie the Size-Stroop effect for preschoolers and for adults. The same pairs of images were used in an adult study (Konkle & Oliva, 2012b) and in Experiments 1 and 2, and there was variability in the magnitude of the Size-Stroop effect across different display pairs – that is, some display pairs elicited much larger Size-Stroop effects than others, for both children and adults. This could be, in part, because certain objects have perceptual features that are more or less typical of big vs. small objects as classes (Long et al., 2016; Long & Konkle, 2017). Thus, in a given pair of big and small objects, one or both of the depicted objects could have perceptual features that are more typical of their size in the real world and thus generate stronger Size-Stroop effects, or they could have atypical features and generate inverted Size-Stroop effects. For example, big objects tend to be boxier to withstand gravitational constrains, whereas small objects tend to be rounder in order to be easily handheld (Long et al., 2016).

We thus analyze these display effects with respect to two distinct hypotheses. First, we know that mid-level perceptual features can drive the Size-Stroop effect in adults (Konkle & Oliva, 2012b; Long & Konkle, 2017). Thus, if this is true for children as well—and if children have abstracted the same mid-level features that characterize each class (as have adults) —then the same pairs of big and small objects should generate stronger Size-Stroop-effects in both children and adults. The first analysis thus assessed the degree to which children’s and adults’ display effects are positively correlated.

Of course, even if this is the case, it is possible that children may *also* use information about the average sizes of object kinds in their automatic computations of the real-world size of the pictured objects. Our second analysis took advantage of the fact that because the items were drawn from an adult study, not all of the objects were actually recognizable by preschool aged children. We thus asked an independent sample of preschool children to identify the pictured objects, and we then examined whether recognizability of the objects affected the item pair differences in Experiments 1 and 2.

**Methods**

***Size-Stroop Item Effects.***  In Experiments 1 and 2, we used the same stimuli used in Experiment 1b of Konkle and Oliva (2012b), the study that initially demonstrated the Size-Stroop effect with adults. In both the present experiment and the original experiment with adults, stimuli were presented in consistent pairs; for example, a picture of a grill was always paired with a picture of a die on both incongruent and congruent trials. Thus, we could obtain measures of the Size-Stroop effect for each individual pair of objects for both children and adults. We used the original data from Experiment 1B of Konkle & Oliva (2012b) to calculate item effects for adults, and we used RT data from the 4-year-olds who contributed RT data in Experiment 1 and 2 to calculate item effects for 4-year-olds.

To calculate Stroop RT item effects for children, we averaged RTs for each congruent and incongruent display separately in each 4-year-old who contributed to RT analyses in either Experiments 1 or 2(62 children).Then, we averaged across participants to calculate group average congruent and incongruent RTs for all 20 pairs. Stroop item effects were then calculated by subtracting the average congruent from incongruent RT. To calculate RT item effects in adults, we performed the same analysis using data from the paired trials of Experiment 1B of Konkle & Oliva (2012b). It is important to note that the procedure for this adult experiment is nearly exactly the same as that of the present Experiments 1 and 2. All of the changes that were made to this task (i.e., converting the task for use on a touchscreen interface, the appearance of Mickey Mouse every 3 correct trials) were for the purposes of adapting this task for use with children. Thus, the item effects from this adult experiment are directly comparable with the item effects obtained from Experiments 1 and 2. See Supplemental Figure 2 for a side-by-side comparison of the average RT data for adults and children.

For our first analysis, we correlated the Stroop item effects observed in the studies with preschool children and with adults. If both children and adults are relying on the same mid-level perceptual features in their automatic computation of real-world size, the same pairs of items should generate stronger or weaker Size-Stroop effects in both populations.

**Results: Adults vs. Children**

../../PaperFigures/Figure4-5_Exp3DisplayLevel/v12-Redo/Appendix_Adults_withDisplayEffects.aiWe found that item effects for preschoolers and adults were highly correlated (*r* = .69, *p* = .001; Figure 4); the same pairs of objects generated stronger Stroop item effects in both adults and children. Thus, these results support the hypothesis that children may compute real-world size from the same mid-level perceptual features adults rely upon to distinguish big from small objects as classes.

*Figure 4.*  Size-Stroop effects (Incongruent – Congruent RT) are plotted for each pair of objects for all 4-year-olds in Experiment 1 and 2 as a function of adult’s Stroop effects for the same pairs of items.

We now turn to analyses that explore whether preschoolers might also draw on knowledge of object size derived from object kind recognition. We first assessed the degree to which preschoolers could recognize the kinds of the depicted objects, and if variability in recognizability predicted the magnitude of the Size-Stroop effects in Experiments 1 and 2.

**Methods: Object Identification**

***Participants****.* Four-year-olds (*N* = 24) participated in a basic-level recognition task. Two additional children participated but were excluded because of (1) a speech articulation difficulty or (2) difficulty speaking English.

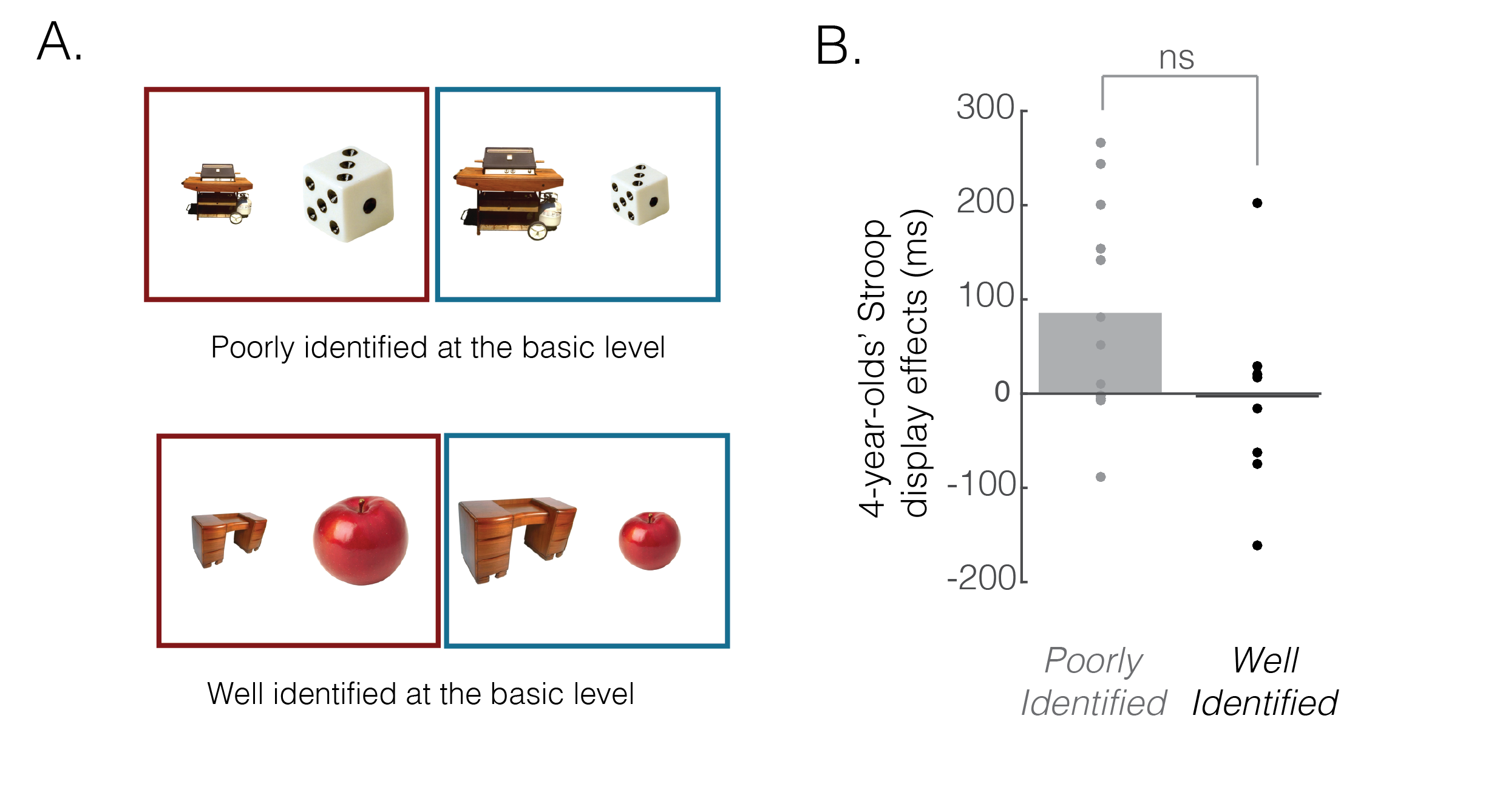
***Procedure:*** Each child saw all 40 objects from Experiments 1 & 2 and was asked: “What does this look like?” If children did not provide a response, they were prompted with a broader question, “Does it remind you of anything?” and encouraged to guess. This second question was designed to elicit descriptions from children that could indicate whether they recognized the pictured object (even if they could not name the object). Images were ordered such that no more than two items from the same size category appeared back to back.

***Analysis: Basic-level Identification:*** First, we coded all responses for *any* evidence of basic-level kind recognition. Most often this was conveyed by the basic-level label (e.g. “apple”, “desk”). However, as we were interested in whether children could recognize the pictured objects (even if they could not name them), we also accepted responses that described the kind. For example, if a child described what an object was used for (i.e., for a die, “you roll it and it gives you a number for a game”), this was counted as a correct recognition of the relevant basic-level kind (see Supplemental Figure 3 for example responses from children and how they were coded).

Overall, children identified the correct basic-level category of the objects 76.1% of the time, gave an incorrect answer 16.8% of the time, and did not give a response 6.9% of the time. Some items were always identified correctly (i.e., apple, 100% identification rate), while others were rarely identified correctly (i.e., perfume bottle, 33.3% identification rate). We then grouped together pairs where the basic-level identities of *both* the big and smallobjects were well identified (greater than 75%, 8/20 pairs, *M* = 95.0% across all 16 items) and pairs where one or more items were poorly identified (75% or less; 12/20 pairs, *M* = 63.5% across all 24 items). Most pairs contained only one item that was poorly identified (8/12 pairs) and four pairs contained two items that were both poorly identified. See Figure 5A for an example of a pair of objects where both items were poorly identified by 4-year-olds (the barbeque, 54.1% identification rate, the die, 62.5% identification rate) and a pair where both items were well identified by 4-year-olds (the desk, 87.5% identification rate; the apple, 100% identification rate).

***Analysis: Real-World Size Identification:*** In a second analysis, we counted as correct any identification of the target as an object in the same real-world size category, as most often, children’s *misidentifications* were of objects from the same real-world size category as the target (though rarely from the same taxonomic superordinate category; 75.1% of misidentifications,t-test against 50%, *t*(23) = 5.51, *p* < .01). As examples, even though very few children identified the pencil sharpener as a pencil sharpener, many children said that it looked like another small object (i.e., binoculars, camera), and two children misidentified the grill as a desk. Here, we separated pairs where children identified any object within the correct size-category at a rate above the median across all items, as size-identification was relatively high (both items >87.5% correct, 8/20 pairs, *M* = 97.7% across items) and pairs where children identified either object within the correct size-category at a rate below the median (one or both items <87.5% correct, 12/20 pairs, *M*=80.6% across items).

**Results: Object Identification**

If anything, pairs of objects that were well-identified at the basic level generated *smaller* Size-Stroop effects in RTs (*M* = -5.4ms) than pairs of objects that were not both well-identified (*M* = 87.17ms; unpaired two-sample t-test, *t*(18) = -1.85, *p*= 0.082; Figure 5B).[[9]](#footnote-11) For example, the Size-Stroop RT effect for the poorly recognized barbecue/die pair was 51.3ms , whereas the Size-Stroop RT effect for the well-recognized desk/apple pair was -161.1ms (Figure 5A). 

However, this strict measure of basic-level identification may not predict Stroop item effects because it underestimates children’s ability to identify these objects. If 4-year-olds’ activation of real-world size information derives from basic-level identification,

then the mistaken categorizations that are still consistent with the objects’ actual real-world sizes might contribute to a Size-Stroop effect, just as correct identification would.

We then examined if Size-Stroop effects were stronger when children could reliably identify any object within the correct size category. As before, we found that pairs of objects whose sizes were poorly identified generated equivalent Size-Stroop effects than pairs with objects whose sizes were well identified (*t*(18)=-1.72, *p* = 0.103).

Taken together, these two analyses suggest that adults and children show remarkable consistency in the pairs of big and small objects that generate Size-Stroop effects, and that explicit recognition of these objects is not a major mediating factor in the Size-Stroop effect for children, as is the case for adults (Konkle & Oliva, 2012a; Long & Konkle, 2017). Taken with the recent visual search findings in preschoolers (Long et al, under review), these results thus support the hypothesis that preschoolers have abstracted the mid-level perceptual features that distinguish big from small objects, as classes, and that preschoolers automatically use these features to compute the real-world size of pictured objects.

**General Discussion**

Across two experiments, we found that preschoolers were impaired at making visual size judgments about pictured objects when the relative sizes of the images was incongruent with their relative sizes in the real-world. This effect was evident in preschoolers’ error patterns and reaction times: 3- and 4-year-olds made more errors on incongruent displays, and four-year-olds also took longer to make visual size judgments on incongruent displays. In Experiment 3, we found that the same pairs of big and small objects generated stronger Size-Stroop effects in children and adults, regardless of how well children could identify these depicted objects. Thus, these results suggest preschoolers automatically activate real-world size information when they see pictured objects: Even though real-world size was task-irrelevant, it was automatically activated and interfered with preschoolers’ ability to make visual size judgments.

*How do children compute real-world size information?*

It could have been the case that preschoolers showed the Size Stroop effect, but that different pairs of big and small objects generated stronger Size Stroop effects in adults and children. Instead, we found a convergence in Stroop item effects across adults and children. Thus, these results provide indirect evidence that children, just like adults, use mid-level perceptual features to directly infer the real-world sizes of objects (Long et al, 2016; Long & Konkle, 2017).

There are two ways that this argument could be confirmed with respect to the present findings. The first way would be to specify the mid-level features that are reliable cues to real-world size, and to show that the presence or absence of these features explains the item effects we see both with children and adults. Delineating the key perceptual features that distinguish big objects from small objects is still an area of active research (Long et al., 2016).

A second prediction of this account is that we should be able to find other evidence that children can infer real-world object size in the absence of basic-level recognition. Across other stimuli sets, the degree to which preschoolers can recognize these objects should not impact the Size-Stroop effects they generate. Furthermore, this account implies that children may be able to infer the real-world size of an object they cannot recognize at a basic or superordinate level. In other words, children should consistently be able to guess a depicted something is big in the real world, even if they cannot tell what it is. Future work that examines the boundary conditions of object size representations in early childhood will bear on this hypothesis.

*Does real-world size organize neural responses to objects in preschoolers?*

In adults, large swaths of object-selective cortex respond more strongly to pictures of small objects than big objects, and other regions show the opposite preference (Konkle & Oliva, 2012a; Konkle & Caramazza, 2013; Julian, Ryan, & Epstein, 2016); these preferences are stable across changes in the retinal and imagined sizes of objects (Konkle & Oliva, 2012b).Recent work implicates these perceptual processing regions as causally implicated in the Size-Stroop effect (Chiou & Lambon Ralph, 2016). In particular, the Size-Stroop effect is reduced when transcranial magnetic stimulation (TMS) is applied over perceptual processing regions (i.e., the lateral occipital complex) but *not* over regions invoked in semantic processing (i.e., the anterior temporal lobe; ATL). Conversely, conceptual judgments about object size are disrupted by TMS to the ATL (Chiou & Lambon Ralph, 2016). In other words, the perceptual representations housed in these regions appear to be causally implicated in the Size-Stroop effect in adults.

Thus, as preschoolers show two behavioral signatures of real-world size representation, this large-scale neural organization could also be evident by this age.Recent work on neural responses to objects in children suggest that this is plausible: adults and children exhibit similar patterns of neural responses to objects, bodies, faces, and scenes by 5–7 years of age (Cohen et al., 2016). Future studies could explore whether a large-scale organization of object-selective cortex by real-world object size is already in place by the preschool years.

*Might younger children also show the Size-Stroop effect?*

When and how do children begin to automatically process the real-world sizes of pictured objects? On the one hand, younger infants and children may activate real-world size when they see pictured objects, but may first need to access basic-level representations (e.g., “bottle”) before they can access size representations. In other words, it could take many years before mid-level features are implicated in the processing of real-world size (Long & Konkle, 2017). If this is the case, we would not expect to observe a Size-Stroop effect until children these younger children, given that there is little evidence that basic-level recognition mediates the Size-Stroop effect. On the other hand, mid-level perceptual representations may become linked to real-world size processing relatively early in life. Infants could acquire the perceptual representations that characterize big vs. small objects as classes without the need for basic-level kind representations, possibly as a byproduct of visual and haptic experience with objects of different sizes. If so, then the perceptual features of unfamiliar objects could already activate real-world size information in young infants.

As this Size-Stroop paradigm was already difficult to run with 3-year-olds, future research will need to develop new methods to examine if and how younger children activate real-world size information when they see pictured objects. We suspect that using a combination of both looking time and reaching measures, as have previously been used to study depth perception in infancy (i.e., Yonas et al., 1982) may help us gain traction on these questions. An understanding of the mechanisms that lead to adult-like real-world size representations is not only important from a developmental perspective, but will inform theories of why and how real-world size organizes our cognitive and neural representations of objects in adulthood.

­

**References**

Bulf, H., Johnson, S. P., & Valenza, E. (2011). Visual statistical learning in the newborn

infant. *Cognition*, 121(1), 127-132.

Chiou, R., & Lamdon Ralph, M. A. (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.

Cohen, M., Dilks, D., Feather, J., Koldewyn, K., Weigelt, S., & Kanwisher, N. (2016). Common

representational structures across the ventral visual pathway of children and adults. *Journal of Vision*, *16*(12), 776-776

Ebeling, K. S., & Gelman, S. A. (1988). Coordination of size standards by young

children. *Child Development*, 888-896.

Frank, M. C., Sugarman, E., Horowitz, A. C., Lewis, M. L., & Yurovsky, D. (2016). Using tablets

to collect data from young children. *Journal of Cognition and Development*, *17*(1), 1-17.

Julian, J. B., Ryan, J., & Epstein, R. A. (2016). Coding of Object Size and Object Category in

Human Visual Cortex. *Cerebral Cortex.*

Gelman, S. A., & Ebeling, K. S. (1989). Children's use of nonegocentric standards in

judgments of functional size. *Child Development*, 920-932.

Golarai, G., Liberman, A., & Grill-Spector, K. (2015). Experience Shapes the Development of

Neural Substrates of Face Processing in Human Ventral Temporal Cortex. *Cerebral Cortex*, bhv314.

Granrud, C. E., Haake, R. J., & Yonas, A. (1985). Infants’ sensitivity to familiar size: The effect

of memory on spatial perception. *Perception & psychophysics*, *37*(5), 459-466.

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.

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.

Konkle, T., & Oliva, A. (2011). Canonical visual size for real-world objects. *Journal of*

*Experimental Psychology: Human Perception and Performance*, *37*(1), 23.

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.

Konkle, T., & Oliva, A. (2012b). A real-world size organization of object responses in

occipitotemporal cortex. *Neuron*, *74*(6), 1114-1124.

Konkle, T., & Caramazza, A. (2013). Tripartite organization of the ventral stream by

animacy and object size. *The Journal of Neuroscience*, *33*(25), 10235-10242.

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.

Long, B., Konkle, T., & Alvarez, G. (2015). Real-world object size is automatically activated

by mid-level shape features. *Journal of Vision*, *15*(12), 7-7.

Long, B., Moher, M., Konkle, T., Alvarez, G.A., & Carey, S. (2015). Broad category

membership guides visual attention in young children.  *Poster presented at the Society for Research in Child Development,* March 19-21, Philadelphia, PA.

Marchette, S. A., Vass, L. K., Ryan, J., & Epstein, R. A. (2015). Outside looking in: landmark

generalization in the human navigational system. *The Journal of Neuroscience*, *35*(44), 14896-14908.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau

(2005). *Reason*, *4*(2), 61-64.

Nasr, S., Echavarria, C. E., & Tootell, R. B. (2014). Thinking outside the box: rectilinear

shapes selectively activate scene-selective cortex. *The Journal of Neuroscience*, *34*(20), 6721-6735.

Rousseeuw, P. J., & Croux, C. (1993). Alternatives to the median absolute deviation. *Journal*

*of the American Statistical association*, *88*(424), 1273-1283.

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.

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.

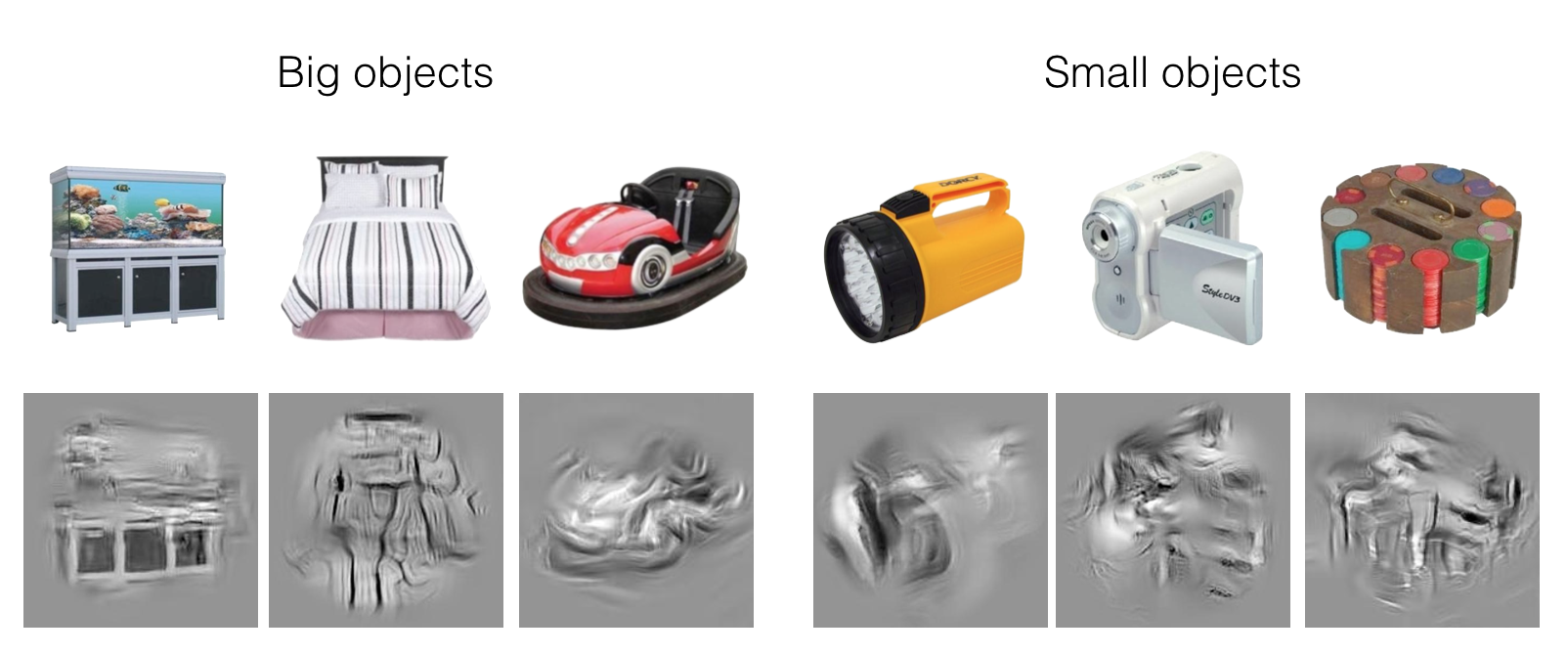
Yarden Gliksman, S. I., Leibovich, T., Melman, Y., & Henik, A. (2016). Automaticity of

Conceptual Magnitude. *Scientific Reports*, *6*.

Yonas, A., Pettersen, L., & Granrud, C. E. (1982). Infants' sensitivity to familiar size as

information for distance. *Child Development*, 1285-1290.

**Appendix**

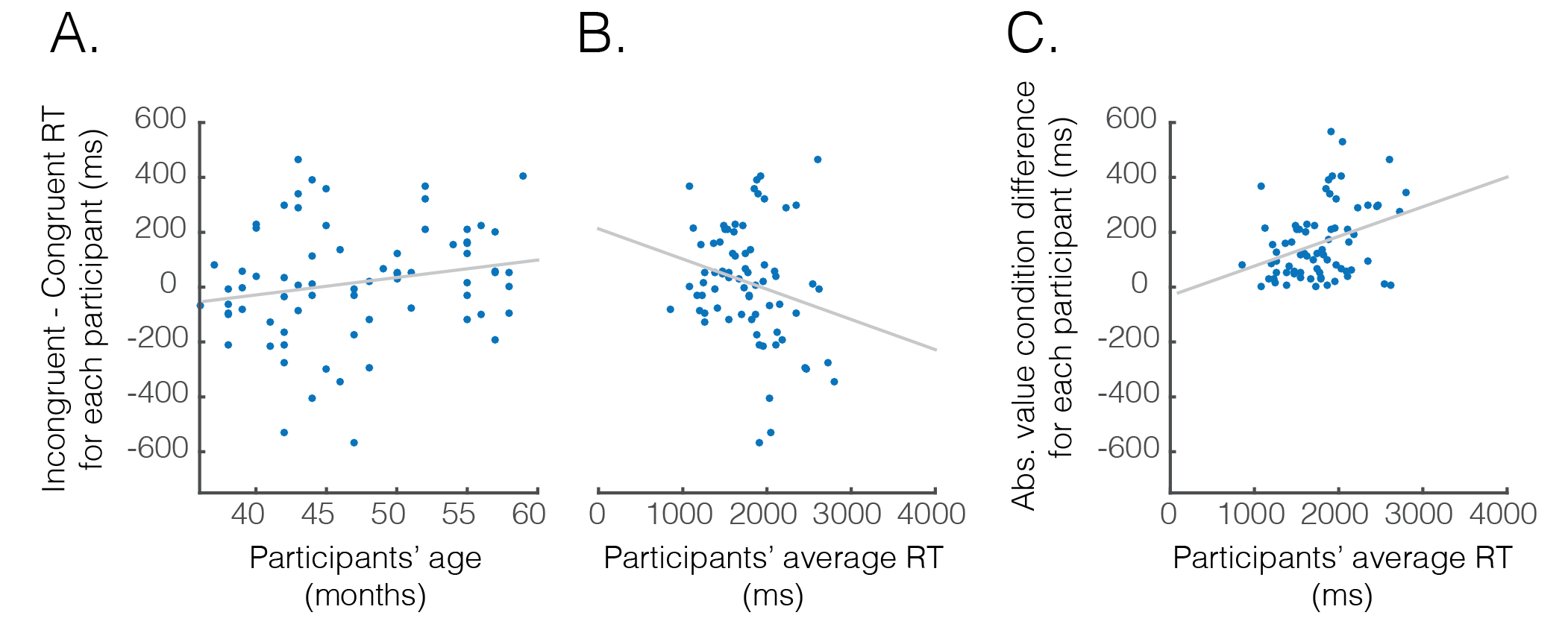
****

*Supplementary Figure 1.* Examples of recognizable images and their corresponding texforms, for a group of three big objects (left) and three small objects (right) (Long et al., 2016; Freeman & Simoncelli, 2011).

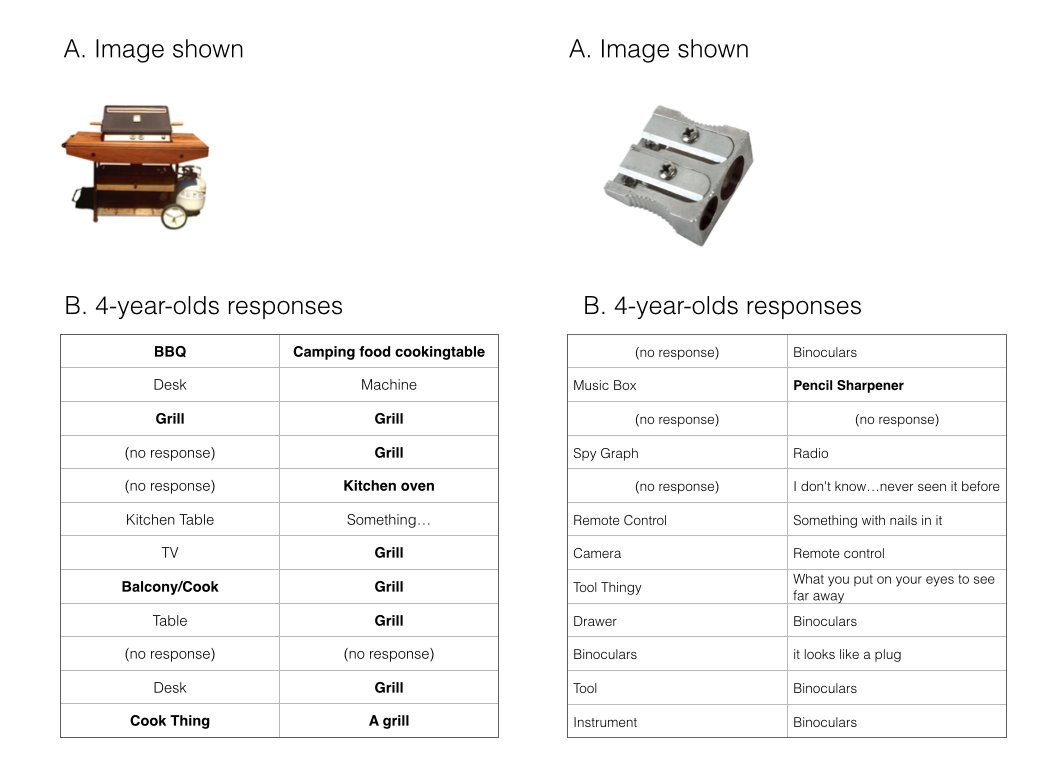
***Supplemental Analyses: Reaction Time Cutoffs***

*Experiment 1:* To ensure that the 4-second global cutoff did not bias our RT results, we reran our analyses only using RT cutoffs defined for each participant. Specifically, we eliminated trials slower or faster than 2 SDs from the mean RT for each participant and condition (incongruent, congruent). We found the same pattern of results. Four-year-olds again tended to show the Stroop effect in their reaction times (Congruent *M* = 1420ms, S*D* = 379ms, incongruent: *M* = 1508ms, *SD* = 314, *t*(30)=2.48, *p* = 0.01, Cohen’s *d*=.45), while three-year-olds did not (congruent, *M* = 1893ms, *SD* = 586ms, Incongruent: *M* = 1945ms, SD= 651ms; *t*(42) = -0.98, *p* = 0.166).

*Experiment 2:* We also reran the analyses of Experiment 2 with these 2SD cutoffs, and found the same pattern of results (Congruent: M=1335.5ms, SD=228.4ms, Incongruent: M=1402.6ms, SD=236.2ms, *t*(31) = -3.37, *p* = 0.001, Cohen’s d = .59).

*Supplementary Figure 2.* Size-Stroop effects for individual children are shown as a function of children’s average age in months (A) and as a function of children’s average speed on the task across both conditions (B). The absolute difference between incongruent and congruent conditions (i.e., the absolute value of the Size-Stroop effect) is plotted for each participant as a function of their average speed on the task (C).

*Untitled:Users:Bria:Dropbox (Personal):Projects:KidStroop:Outputs:PaperFigures:Figure8-Appendix-Adults:Appendix_Adults.aiSupplemental Figure 2* (A). Average congruent and incongruent reaction times for paired trials in Experiment 1B of Konkle & Oliva (2012a). Adults responded with keypresses instead of touching an iPad. (B) The same reaction time results plotted in Figure 3B are shown here for comparison.

**** *Supplemental Figure 3.* Twenty-four four-year-olds were asked “what does this look like?” about the depicted object shown in (A). Their responses are shown in (B). Responses that were counted as correct recognitions are bolded. Note that responses were coded liberally; for example, “balcony/cook” was accepted as a correct answer for the grill (A, left panel).

1. Children in these two familiarization versions did not perform more or less accurately on test trials (no main effect of familiarization version on error rates; *F*(1,77) = .43, *p* = .51) and or on congruent versus incongruent displays (no interaction of familiarization version with trial type on error rates; *F*(1,77) = 1.84, *p*= .18). [↑](#footnote-ref-1)
2. In the adult study on which this study is based (Konkle & Oliva, 2012a), half of the participants were asked to indicate which object is larger on the screen, and half of the participants were asked to indicate which object is smaller on the screen. Size-Stroop effects were observed in reaction times and error rates for both tasks. However, the *indicate-smaller* task produced a slightly bigger effect size, and thus, to maximize power, children were only asked the latter question. [↑](#footnote-ref-2)
3. In a pilot study, we found that marking children’s progress on the stamp sheet dramatically increased the number of trials children were willing to complete, suggesting that children were very sensitive to this feedback. [↑](#footnote-ref-3)
4. Including these practice trials or excluding children who completed less than 10 trials after practice trials (3 children) did not change the pattern of results. [↑](#footnote-ref-4)
5. The same pattern of results was found when, instead of a global cutoff set at 4 seconds, we trimmed reaction times on an individual basis, eliminating any trials slower or faster than 2SDs from the mean of each child’s RT for each condition. See Appendix: Supplemental Analyses for details. [↑](#footnote-ref-5)
6. [↑](#footnote-ref-7)
7. We also confirmed that these children made more errors on incongruent displays, suggesting that they were not engaging in a speed-accuracy tradeoff (incongruent *M* = 3.93%, congruent *M* = 1.08%, *t*(30) = 3.35, *p* = .001, Cohen’s *d* = .60). [↑](#footnote-ref-8)
8. As an exploratory analysis, we included the two children with slow overall RTs. We found that including these children did not change the pattern of effects in the linear mixed-effect model on logRTs (*B* = .03, *SE* = 0.1, *df* = 1452.15, *t* = 2.18, *p* = .03) but did the pattern of effects in a traditional t-test analyses (*t*(32)= 1.05, *df* = 32, *p* = 0.15). [↑](#footnote-ref-9)
9. As 75% was a relatively arbitrary cutoff, we examined whether a range of other cutoffs generated the same patterns of effects. Regardless of the cutoff we used, poorly- identified pairs of objects generated equivalent or larger Size-Stroop effects that well-identified pairs of objects. [↑](#footnote-ref-11)