American Psychological Association

Journal of Experimental Psychology: Human Perception and Performance

0096-1523/19/$12.00

2019, Vol. 45, No. 7, 863–876
http://dx.doi.org/10.1037/xhp0000619

Real-World Size Is Automatically Encoded in Preschoolers' Object Representations

Bria Long
Harvard University and Stanford University

Mariko Moher
Williams College

Susan Carey and Talia Konkle
Harvard University

When adults see a picture of an object, they automatically process how big the object typically is in the real world (Konkle & Oliva, 2012a). How much life 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 be either 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 addition, we found that both 4-year-olds and adults showed similar item-pair effects (i.e., showed larger Size-Stroop effects for a given pair of items, relative to other pairs). Furthermore, the magnitude of the item-pair Stroop effects in 4-year-olds did not depend on whether they could recognize the pictured objects, suggesting that the perceptual features of these objects were sufficient to trigger the processing of real-world size information. These results indicate that, by 3–4 years of age, children automatically extract real-world size information from depicted objects.

Public Significance Statement
Real-world size interfered with preschoolers' 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 4-year-olds. Size-Stroop effects were generated by pictured objects that 4-year-olds had trouble naming.

Keywords: object representation, real-world size, Stroop effects, visual development

When we look at the world, we easily recognize objects and perceive their physical size, from small objects like cups and paperclips, to bigger objects like cars and pianos. Indeed, our representations of the typical sizes of objects enter into human mental life in many ways—for example, 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), participating in computations of spatial layout (in their role in specifying how far away objects are), and constraining motor interactions (e.g., we tend to pick up small objects with our hands, and we need to navigate around big objects). Thus, the real-world size of objects plausibly structures our visual experience with objects as we learn about the world.

And indeed, there is evidence that real-world size has an organizing role in both perceptual and neural object representation by adulthood (Henik, Gliksman, Kallai, & Leibovich, 2017; Julian, Ryan, & Epstein, 2017; Konkle & Oliva, 2012a; Konkle & Oliva, 2012b). For example, even though there are many kinds of big and small objects, visual search behavior shows that big objects, as a class, actually *look* different from small objects, as a class (Long, Konkle, Cohen, & Alvarez, 2016). Furthermore, at a neural level, the distinction between small and big objects also organizes responses in occipitotemporal cortex (Cate, Goodale, & Köhler, 2011; Khaligh-Razavi, Cichy, Pantazis, & Oliva, 2018; Konkle & Oliva, 2012b; Julian et al., 2017).

Real-world size information appears to be so ingrained in adults' object representations that when they see an object they not

This article was published Online First April 15, 2019.

Bria Long, Department of Psychology, Harvard University, and Department of Psychology, Stanford University; Mariko Moher, Department of Psychology, Williams College; Susan Carey and Talia Konkle, Department of Psychology, Harvard University.

Mariko Moher is now at Connecticut College.

Correspondence concerning this article should be addressed to Bria Long, Department of Psychology, Stanford University, Jordan Hall, 450 Serra Mall, Stanford, CA 94305. E-mail: bria@stanford.edu

only automatically recognize what it is but automatically activate knowledge about its typical size in the real world (e.g., Chiou & Lambon Ralph, 2016; Gliksman, Itamar, Leibovich, Melman, & Henik, 2016; Konkle & Oliva, 2012b; see also Paivio, 1975). As evidence of this, in a Size-Stroop paradigm, adults were 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 were slower and less accurate at making *visual* size judgments when the images' relative visual sizes were 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 were congruent with their real-world sizes (i.e., a big picture of a gazebo and a small picture of a teapot; Konkle & Oliva, 2012a). Thus, in this task, even though real-world size information was task-irrelevant, it was automatically activated and interfered with adults' ability to make visual size judgments. Further, there is some evidence that real-world size is automatically activated even when the task does not involve making a size-related judgment, such as when judging the animacy of a depicted object (Sellaro, Treccani, Job, & Cubelli, 2015) or making a word–nonword judgment (Sereno, O'Donnell, & Sereno, 2009).

It is intuitive that one must recognize an object before any real-world size related processing can proceed, as in classic accounts of semantic processing (e.g., Collins & Quillian, 1969; Jolicoeur, Gluck, & Kosslyn, 1984). However, using "texform" stimuli, it has been shown that real-world size information can be activated without first recognizing the object. Texform stimuli preserve mid-level features like curvature and texture but lack the high-level features that enable basic-level recognition (see the Appendix, Figure A1, for examples). In a line of behavioral and neural studies with adults, we have found that when images are transformed into texforms, (a) visual search effects between big and small objects persist (Long et al., 2016), as do (b) neural differences between big and small object images (Long, Yu, & Konkle, 2018) as well as (c) the Size-Stroop effect (Long & Konkle, 2017). These studies demonstrated that there are *mid-level perceptual* features that systematically distinguish small objects from big objects as classes that underlie object representations along the ventral stream and that can lead to the automatic computation of real-world size.

The fact that real-world size is such an ingrained and organizing property of our object representations raises several critical developmental questions: How does this real-world size sensitivity emerge over development; that is, what, if any, innate or early developing support could exist for it, and what learning mechanisms are involved? Answering these questions not only is an important project within developmental cognitive neuroscience but would also shed light on *how* the distinction between small and 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, because they are the youngest age group likely capable of performing the same exact tasks used to study object-size processing in adults.

Preschoolers are at an interesting point in the development of their object representations. Early in the preschool years, 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 represent the average sizes of some categories (Ebeling & Gelman, 1988; Gelman & Ebeling, 1989). Furthermore, evidence from visual search has suggested that big objects as a class "look" different from small objects as a class to children as young as age 3 (Long, Moher, Carey, & Konkle, 2019), indicating that real-world object size influences preschoolers' perceptual similarity computations. At the same time, however, considerable evidence has suggested that preschoolers' object representations may be far from mature (for reviews, see Jüttner, Wakui, Petters, & Davidoff, 2016; Nishimura, Scherf, & Behrmann, 2009). For example, children continue to exhibit deficits recognizing objects across wide variations in lighting and pose throughout middle childhood (e.g., Bova et al., 2007), don't integrate haptic and visual information in size discrimination tasks until around 8 years of age (Gori, Del Viva, Sandini, & Burr, 2008), and are less deceived by the Ebbinghaus illusion than adults are until 7 years of age (Doherty, Campbell, Tsuji, & Phillips, 2010). Further, as children's *own* physical size relative to objects changes dramatically over the first few years of their life, one might expect their object size representations to mature gradually throughout childhood.

Thus, the present experiments sought to establish whether preschoolers, like adults, automatically activate real-world size information when they see pictured objects. To do so, in Experiments 1 and 2, we used 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 interferes with the task at hand. Further, and more speculatively, we assessed whether the data support mid-level perceptual processing versus basic-level recognition as a locus of the observed effects. To do so, we explored which item pairs generate the greatest Size-Stroop effects as a first step toward understanding whether similar perceptual mechanisms underlie real-world size representations in preschoolers and adults.

Experiment 1: Do Preschoolers Show the Size-Stroop Effect?

We adapted the Size-Stroop task (Konkle & Oliva, 2012a) for children by converting it to an iPad game. Children were asked to "touch the picture that is smaller *on the screen*." Critically, the identity of the objects and their real-world sizes were completely irrelevant to the task. However, 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, than on congruent displays, in which the object that is bigger in the real world is bigger on the screen (see Figure 1).

Method

Participants. Eighty 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 ($N = 16$ in adults; see Konkle & Oliva, 2012a). One child began the task but did not complete more than

Figure 1. Example stimuli for Experiments 1 and 2. In congruent displays, the relative sizes of the objects were consistent with their sizes in the real world, and in incongruent displays, the relative sizes of the objects were inconsistent with their sizes in the real world. See the online article for the color version of this figure.

two trials. This left us with 79 children in the final sample, with 48 three-year-olds (M = 41.8 months, SD = 3.0) and 31 four-year-olds (M = 53.7 months, SD = 3.4). Post hoc sensitivity analyses indicated that this sample was large enough to detect an effect size with d_z = .28 in all children, d_z = .38 in 3-year-olds, and d_z = .45 in 4-year-olds with 80% power (one-tailed t test, α = .05). Although the original Size-Stroop effect that this study is based on had a relatively large effect size in adults (Cohens d_z = 1.43; Konkle & Oliva, 2012a), we aimed to test for markedly smaller effects given both practical and theoretical considerations with this younger population.

Experimental setup. Children sat at a table across from an experimenter who held an iPad for them (see Figure 2). 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), using custom code was written in Javascript. Reaction time, touch position, accuracy, and experimental details were recorded and saved to an online database after each trial.

Stimuli. Images were identical to those used in Experiment 1B of Konkle and Oliva (2012a); 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 1).

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 smaller than the other, and were asked to "touch the shape that is smaller *on this paper*." However, because several children were distracted by the appearance of the iPad for the test phase, the remaining children completed the practice phase on the iPad. During the practice phase on the iPad, children touched a blue dot to begin each trial, after which they were presented with two different colored shapes, one of which was smaller 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 and by marking a stamp on their stamp sheet after three correct trials. To ensure that children understood the instructions, we required all children in the paper phase complete five correct practice trials before the experimenter started the test phase, and all children in the iPad familiarization phase were required to complete nine correct practice trials before the experimenter started the test phase.[1] Critically, the real-world sizes of the objects were never mentioned during either the practice phase or the test phase. Thus, children were instructed only to pay attention to the visual sizes of the images on the screen.

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 500 ms, after which two images appeared on either side of the screen. Children were reminded to "touch the picture that is smaller *on the screen*,"[2] and these instructions were repeated as needed if the child forgot the task. 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 (e.g., 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 (e.g., 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 three correct trials, and children's progress was marked with a stamp by the experimenter.[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 four 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 dur-

[1] Children who practiced with the iPad went on to perform slightly worse overall than did those who practiced with only the paper version, though this was not significant (average error rates; M_{paper} = 7.25%, SD = 12.5; M_{iPad} = 13.2%, SD = 17.4), $F(1, 74)$ = 3.35, p = .07.

[2] In the adult study on which this study is based (Konkle & Oliva, 2012a), half of the participants were asked to indicate which object was larger on the screen, and half of the participants were asked to indicate which object was 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 asked only the latter question.

[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.

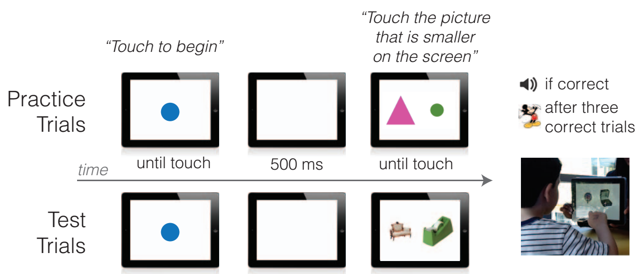

Figure 2. Schematic of practice and test trials used in Experiments 1 and 2. A child performing the task is shown on the right. See the online article for the color version of this figure.

ing the experiment. The image pair that occurred on a given trial was randomized throughout each session for each child.

Data inclusion. We assessed both accuracy and reaction time (RT) as dependent measures, both across all participants and separately for 3- and 4-year-olds. To do so, we adopted the following exclusion criteria and data-trimming methods. First, we excluded all geometric shape practice trials and, a priori, the first 10 trials from the test phase.[4] Additionally, three children (all 3-year-olds) did not complete more than five trials in each condition after these first 10 trials and so were also excluded from all subsequent analyses. Error analyses were thus conducted on the remaining 76 children, with 3-year-olds who contributed on average 54.5 trials to error analyses and 4-year-olds who on average contributed 52.1 trials.

For RT analyses, we additionally excluded incorrect trials and trials with RTs slower than 4 s, a preset criterion (6.6% of correct trials). This RT cutoff has previously been used as a cutoff when analyzing preschooler's RTs in a touchscreen-based task (Frank, Sugarman, Horowitz, Lewis, & Yurovsky, 2016), to eliminate extralong trials where children are likely off-task. Children were included in RT analyses if, after this RT trimming procedure, they had at least five correct trials per condition (congruent, incongruent), excluding practice trials (i.e., both shape practice trials and the first 10 trials of the test phase). Four additional children, who were all 3-year-olds, were excluded for not meeting these criteria. This left us with 72 children for RT analyses: 41 three-year-olds (M = 42.2 months, SD = 2.9) and all of the 31 four-year-olds. On average, 3-year-olds contributed 47.4 trials to RT analyses, and 4-year-olds contributed 47.8 trials to RT analyses.

Data analysis. We analyzed error patterns and RTs in two ways. The first approach was designed to mirror the analyses used in Konkle and Oliva (2012b). Analyses of variance (ANOVAs) were used to assess the effects of the within-subject variable of trial type (congruent vs. incongruent) and the between-subjects variable of age group (3 vs. 4 years) on error rates (percentage of completed trials that were errors) and average RTs. Planned paired t tests are reported to directly compare congruent and incongruent conditions, where we a priori decided to employ one-sided tests because the results were interpretable only if children performed worse on incongruent relative to congruent displays.[5]

The second analysis approach was employed to validate the robustness of these results using a more appropriate and powerful statistical framework, specifically taking into account the nonnormal distribution of RTs (Whelan, 2008), as well as the variability across both subjects and items. Specifically, we modeled log-transformed RT data using a linear mixed-effects model, and we modeled error patterns using a generalized linear mixed effect model, using the lme4 code package implemented in R (Bates, Maechler, Bolker, & Walker, 2014). In these models, age group, congruency, and their interaction were specified as fixed effects. Random intercepts for subjects and for individual displays were always included, and random slopes were included if the model was able to converge with this more maximal design (Barr, Levy, Scheepers, & Tily, 2013). With these random effects terms, the statistical models are better able to (a) account for the different numbers of trials completed by individual children and also (b) ensure that the results are not strongly driven by particular displays. All data and analysis code are available at the public repository for this article (https://osf.io/wekhy/).

Results

Error results. Children made relatively few errors (M = 10.6%), suggesting they understood the task instructions, though 3-year-olds made more errors than did 4-year-olds (main effect of age: 3-year-olds: M = 14.1%; 4-year-olds: M = 5.6%), $F(1, 74) = 7.08$, $p = .01$, $\eta^2_G = .07$.

Critically, we found that 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 = 8.0%, SD = 12.1; incongruent: M = 13.2%, SD = 18.3), $F(1, 74) = 11.87$, $p < .001$, $\eta^2_G = .03$. The Size-Stroop effect was apparent throughout this age range; there was no interaction between age group and trial type, $F(1, 74) = .31$, $p = .58$, $\eta^2_G < 0$. Further, planned ad hoc comparisons confirmed that the Size-Stroop effect was observed at each age: 3-year-olds (congruent: M = 11.2%; incongruent: M = 17.0%), $t(44) = 2.88$, $p < .01$; 4-year-olds (congruent: M = 3.5%; incongruent: M = 7.7%), $t(30) = 2.2$, $p = .018$ (see Figure 3A). The generalized mixed-effects model confirmed these analyses while accounting for variance across displays and subjects (main effect of congruency; b = .566, SE = .112, Z = 4.896, $p < .001$).

Given that the 3-year-olds showed a fairly high error rate relative to 4-year-olds, it is possible that they could have misunderstood the instructions. On one account, 3-year-olds could have instead considered real-world size the relevant task dimension. If so, 3-year-olds should have consistently chosen the object that is small in the real-world, making few errors on congruent trials and many more errors on incongruent trials. Instead, we found that 3-year-olds made overall more errors relative to 4-year-olds, including on congruent trials. Indeed, the relative difference in error rates across conditions was similar between 3-year-olds and 4-year-olds (3-year-olds: $M_{incong-cong}$ = 5.8%, SD = 13.6;

[4] This cutoff of 10 trials was chosen after piloting the task in a separate group of children and noticing that some children were still responding very slowly during the first few trials. Error rates were still relatively low during these first practice trials, suggesting that children did understand the task instructions (3-year-olds' average error rate: M = 11.6%, SD = 15.2%; 4-year-olds': M = 3.9%, SD = 7.2).

[5] Although we decided a priori to use one-sided significance tests for these comparisons (given that the results are interpretable in only one direction), these results and the following results all still hold if two-sided significance criteria are used.

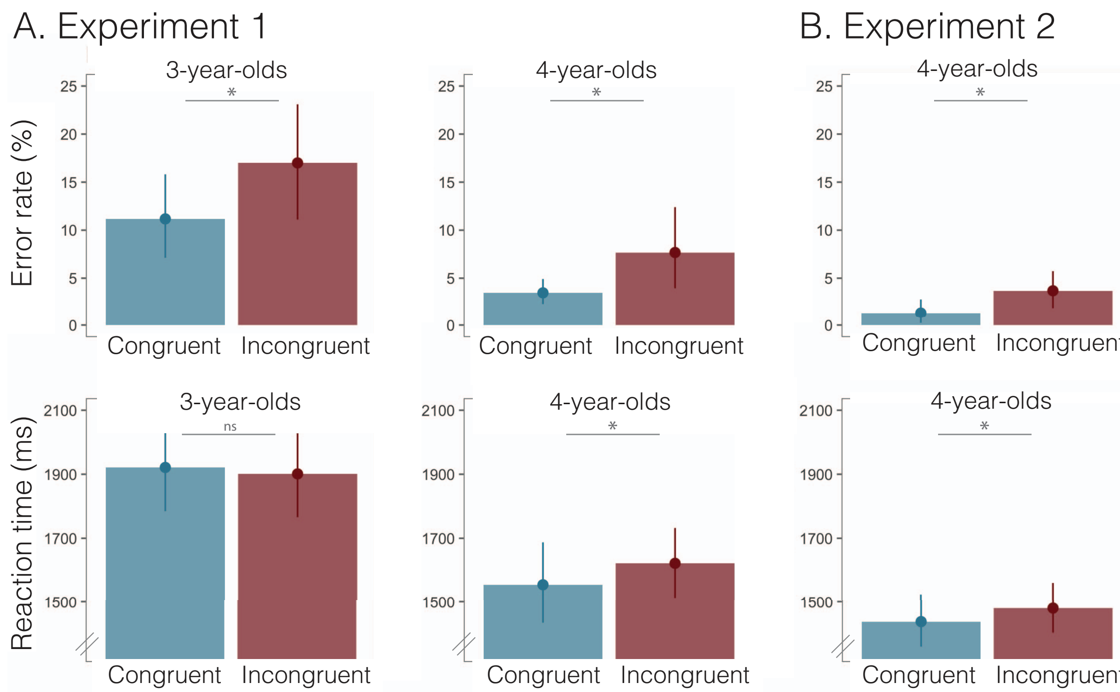

Figure 3. Average error rates (upper panels) and reaction times (lower panels) are shown for congruent (blue [lighter gray]) and incongruent (red [darker]) displays for 3-year-olds and 4-year-olds in Experiment 1 (Panel A) and from 4-year-olds in Experiment 2 (Panel B; replication experiment). Error bars represent bootstrapped 95% confidence intervals. See the online article for the color version of this figure.

4-year-olds: $M_{\text{incong}-\text{cong}} = 4.2\%$, $SD = 10.7$; see all individual data in the Appendix, Figure A2). On another account, 3-year-olds might initially understand the task but then become more confused throughout the session. Contrary to this idea, we found that children's overall error rates were consistent in both age groups across the session (average error rates across quartiles of each child's session; 3-year-olds: $M = 12.7\%, 15.8\%, 14.4\%, 13.6\%$; 4-year-olds: $M = 6.3\%, 5.4\%, 6.8\%, 3.8\%$). Nonetheless, these analyses cannot completely rule out an account where 3-year-olds are somewhat more distracted overall and then become slightly more confused on incongruent trials when there is a conflict between real-world size and visual size.

Reaction time results. Considering both 3- and 4-year-olds together, we found that children did not take longer to make visual size judgments on incongruent versus congruent displays (no main effect of trial type: congruent: $M = 1758$ ms; incongruent $M = 1778$ ms), $F(1, 70) = .9$, $p = .35$, $\eta^2_G < .001$. However, considering 3- and 4-year-olds separately, we found that 4-year-olds showed the Size-Stroop effect in their RTs, whereas the 3-year-olds did not: 4-year-olds (congruent: $M = 1555$ ms, $SD = 359$; incongruent: $M = 1622$ ms, $SD = 319$), $t(30) = 2.37$, $p = .01$, Cohen's $d = .43$; 3-year-olds (congruent: $M = 1921$ ms, $SD = 475$; incongruent: $M = 1901$ ms, $SD = 446$), $t(40) = -.54$, $p = .7$, Cohen's $d = -.08$ (see Figure 3A; see also the Appendix, Figure A2). This same pattern of results was evident in the linear mixed-effects model; that is, when combining across all children, there was no congruency effect in RT ($b = 0.002$, $SE = 0.02$, $t(40.06) = .12$, $p = 0.91$) and a trend toward an interaction between congruency and age ($b = 0.046$, $SE = 0.02$, $t(65.35) = 1.815$, $p = 0.074$); however, congruency was significant when 4-year-olds were considered separately (congruency, $b = 0.047$, $SE = 0.021$, $t(21.32) = 2.21$, $p = .04$).

As is evident in Figure 3A, the 3-year-olds also generally took longer to respond on the iPad than did 4-year-olds. Thus, we conducted an exploratory analysis to examine whether age or overall slowness was more likely to account for the 3-year-olds' lack of the Size-Stroop effect on RTs. First, we analyzed whether children's age (in months) was correlated with the degree to which children made more errors or had slower RTs on the incongruent than the congruent trials. Age was only weakly correlated with the size of the Size-Stroop effect for RTs (RTs: $r = .20$, $p = .09$) and was not positively correlated for the size of the Size-Stroop effect on errors (error rates: $r = -.10$, $p = .38$). Next, we analyzed how overall RT was related to Stroop RT. Children who performed the task slower were marginally more likely to have a smaller Size-Stroop RT effect ($r = -.22$, $p = .06$); if anything, children who took longer to respond seemed to have very positive or very negative Size-Stroop effects (as evidenced by a significantly stronger correlation between overall RT and absolute-valued Stroop effects; $r = .33$, $p < .01$); difference between average and absolute-valued Stroop effect correlations with age was significant, $t(69) = -3.58$, $p < .01$ (see the Appendix, Figure A3).[6] In other words, children who had slower RTs also tended to have

[6] Age was not positively correlated with absolute valued Stroop reaction time effects ($r = -.04$, $p = .72$).

more variance in their RTs, leading to noisier estimates of the Size-Stroop effect.

Experiment 2: Replication

In Experiment 1, we found that both 3- and 4-year-olds showed a Size-Stroop effect in their error patterns, suggesting that preschoolers automatically activated information about the real-world sizes of the depicted objects. Further, we found that the Size-Stroop effect was also evident in 4-year-olds RTs, as it is in adults (Konkle & Oliva, 2012a). However, 3-year-olds also tended to stay less on task and did not consistently make speeded visual size judgments, making it harder to obtain accurate estimates of 3-year-olds' RTs for congruent versus incongruent displays and thus to observe a Size-Stroop effect in their RTs. In Experiment 2 we sought to replicate the results of Experiment 1 in an independent group of 4-year-olds. Given that we observed the RT effect only in the 4-year-olds, the aim of this experiment was to validate the robustness of this RT effect. Further, in Experiment 2, we a priori decided to add an additional exclusion criterion for overall slow responders, because in Experiment 1 we found that children with very slow RTs tended to show highly variable Size-Stroop effects in their RTs.

Method

Participants. Thirty-five 4-year-olds were recruited for Experiment 2 so that approximately the same number of 4-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 = 53.4 months, SD = 3.2; 15 boys). Post hoc sensitivity analyses suggested that this sample size would be sufficient to detect an effect size of $d_z = .44$ with 80% power (one-tailed t test; $\alpha = .05$).

Experimental setup, stimuli, and counterbalancing. All aspects of Experiment 2 were identical to those in Experiment 1, except that all children did all practice trials on the iPad, and we encouraged children to obtain 20 stamps (rather than just as many stamps as possible), to help maximize the number of children who could be included in RT analyses.

Data inclusion. As in Experiment 1, the first 10 test trials were discarded for all participants. Children completed an average of 47.0 trials (range = 28–56) out of a possible 70. All children were included in error analyses because they completed five or more trials in each condition after these first 10 trials in the test phase. For RT analyses, we applied the same exclusion criteria as in Experiment 1, excluding trials where children responded incorrectly or took longer than 4 s to respond (M = 1.2% of correct trials). No children were excluded on the basis of not having five or more test trials with correct responses made in less than 4 s. As planned, we then excluded children whose average RTs (across both conditions) were slower than 2 SDs from the average group RT (only two participants were excluded for this reason; mean RTs = 2,603 ms and 2,432; Z scores = 3.1 and 2.5). After applying these inclusion criteria, we analyzed the RTs of 31 children, who completed an average of 45.26 trials after practice.

Analysis. The same analysis plan (ANOVAs, linear mixed-effects modeling) from Experiment 1 was followed in Experiment 2.

Results and Discussion

Error results. Overall, we replicated the main finding from Experiment 1, finding that 4-year-olds made more errors on incongruent displays than congruent displays (incongruent: M = 3.7%; congruent: M = 1.4%), $t(32) = 2.55$, $p = .008$, even though they made fewer errors overall when compared to 4-year-olds in Experiment 1 (see Figure 3B). This result was confirmed in our generalized linear mixed-effects model (main effect of congruency, $b = 1.216$, $SE = .56$, $Z = 2.19$, $p = .028$).

Reaction time results. Further, 4-year-olds also exhibited a Size-Stroop effect in their RTs, taking longer to make visual size judgments on incongruent displays than on congruent displays (congruent: M = 1438 ms; incongruent: M = 1480 ms), $t(30) = 2.30$, $p = .01$, Cohen's $d = .41$ (see Figure 3B). Our linear mixed-effects model on logged RTs revealed the same pattern of results ($b = .04$, $SE = .02$, $t(21.22) = 2.28$, $p = .03$).[7] Thus, these data replicate the pattern of effects seen in Experiment 1; 4-year-olds exhibited a Size-Stroop effect in both their errors and RTs.

Size-Stroop Item-Pair Analyses

Experiments 1 and 2 show that the Size-Stroop effect is observable in error rates by age 3 and in RTs as well by age 4 and thus establish that when preschool-age children see pictured objects, information about the real-world size of these objects is automatically activated. 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'. In the following post hoc analyses, we begin to provide some insight into these questions by taking advantage of the fact that children and adults completed this task using the same images (Experiments 1 and 2 for children; Konkle & Oliva, 2012a, for adults). In both the present experiments and the original experiment with adults, these items 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 (RT to incongruent pair and RT to congruent pair) for each individual pair of objects for both children and adults.

Further, for both adults and children there was variability in the magnitude of the Size-Stroop effect that a given item pair generated; some item pairs generated greater incongruency costs than did others. By comparing item-pair effects between children and adults, we could indirectly examine whether similar mechanisms underlie the Size-Stroop effect for preschoolers and for adults.

One possibility is that Size-Stroop item-pair effects are highly correlated between children and adults, which would provide indirect evidence that the mechanisms underlying these effects are similar. In adults, our prior research has highlighted that item-pair

[7] As an exploratory analysis, we included the two children with slow overall reaction times (RTs). We found that including these children did not change the pattern of effects in the linear mixed-effects model on log RTs ($b = .03$, $SE = .1$, $t = 2.11$, $p = .04$), but did change the pattern of effects in a traditional paired t test, $t(32)= 1.05$, $p = .15$.

effects elicited by texform stimuli are related to differences in mid-level visual features. For example, big objects tend to be boxy and small objects tend to be curvy, and these kinds of differences at this perceptual level are sufficient to trigger a Size-Stroop effect, even when the items on the display are not recognizable at the basic-level (Long & Konkle, 2017). If children are sensitive to the same mid-level visual features—that is, if they have abstracted the same mid-level features that characterize small objects and big objects as classes—then the same item pairs should generate stronger Size-Stroop-effects in both children and adults. Our first analysis thus analyzes the degree to which children's and adults' Stroop item-pair effects are positively correlated.

However, it is also possible that children automatically compute the size of the depicted objects in this paradigm by first recognizing them as a kind (e.g., "cup") and then retrieving information about the average size of that kind (e.g., "cups are small enough to be held with one hand"). To examine this possibility, we took advantage of the fact that the pictures of big and small objects used in this study were drawn from an adult study, and thus not all of the objects were necessarily recognizable by preschool-age children. Specifically, we analyzed whether in Experiments 1 and 2 the more recognizable objects (to preschoolers) give rise to a stronger Size-Stroop effect than did the less recognizable objects. For this analysis, we first asked an independent sample of preschool children to identify the pictured objects. Given the expected variation in recognizability, we next assessed whether the recognizability of the objects affected the magnitude of Stroop effects across item-pairs. If children's Size-Stroop effects rely on them first recognizing the objects, then one would expect to find stronger effects for pairs of items depicting more identifiable objects than for pairs of items depicting less identifiable objects. Our second analysis tested this hypothesis.

Analysis 1: Item-Pair Effects in Adults Versus Children

For our first analysis, we correlated the item-pair effects observed in the studies with preschool children and with adults. We used RT data from the 4-year-olds who contributed RT data in Experiments 1 and 2 to calculate item-pair effects for 4-year-olds and the original data from Experiment 1B of Konkle and Oliva (2012a) to calculate item-pair effects for adults. Because adults made relatively few errors, we restricted our analyses to RT data from adults and children.

Method. To calculate Stroop item-pair effects for children, we computed the average RTs for each congruent and incongruent display in each 4-year-old who contributed to RT analyses in either Experiment 1 or 2 (62 children). Next, we computed the average congruent and incongruent RTs for all 20 displays at the group level. Finally, Size-Stroop item-pair effects were calculated by subtracting the average congruent from incongruent RT at the group level for each item pair. The same calculation was used for the data from Experiment 1B of Konkle and Oliva (2012a). Given that the adult data were collected using a difference interface (i.e., keypresses on a computer), we caution against comparing the absolute magnitude of the effects and here focus on only the correlation across item-pair effects between children and adults.

Results. Item-pair effects for preschoolers and adults were highly correlated ($r = .64$, $p = .001$; see Figure 4); the same pairs

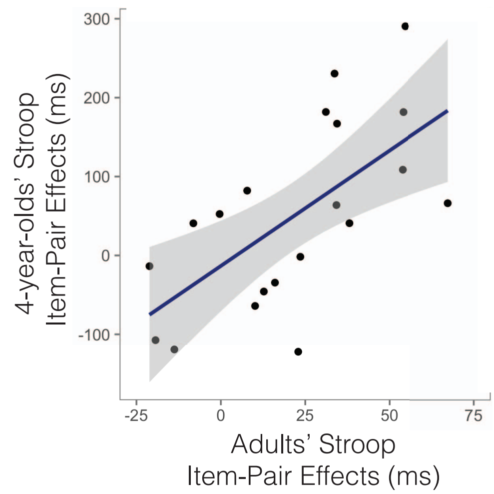

Figure 4. Size-Stroop item-pair effects (incongruent–congruent reaction time) are plotted for each pair of objects for all 4-year-olds in Experiments 1 and 2 as a function of adults' Stroop item-pair effects for the same pairs of displays. Shaded regions represent 95% confidence intervals around the mean. See the online article for the color version of this figure.

of objects generated greater RT differences between congruent and incongruent item pairs in both adults and children. Thus, these results provide indirect support that the same mechanisms underlie the Size-Stroop effect in adults and preschoolers and are consistent with the hypothesis that children may automatically compute real-world size from the same mid-level perceptual features adults rely upon to distinguish big from small objects as classes.

Analysis 2: Object Identification

We now turn to an analysis of the item-pair effects that explores whether preschoolers might also draw on knowledge of object size derived from basic-level kind recognition. We first established the degree to which preschoolers could recognize the basic-level kind of each depicted object. Next, we assessed the degree to which variability in recognizability predicted the magnitude of the Size-Stroop effects.

Method. Four-year-olds ($N = 24$) participated in the basic-level recognition task. Two additional children participated but were excluded because of (a) a speech articulation difficulty or (b) difficulty speaking English. Each child saw all 40 objects from Experiments 1 and 2 and was asked "What does this look like?" If children did not provide a response, they were prompted with the 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.

Next, we coded all of these responses for any evidence of basic-level kind recognition. Some responses were straight forwardly correct (e.g., "apple," "desk"). However, we also counted descriptions of the object kind as correct (e.g., for a die, "you roll it and it gives you a number for a game"; see the Appendix, Figure A4; e.g., responses from children and how they were coded). This more liberal method was followed because the aim here was separating objects that were identifiable to children from those objects that were not, regardless of whether the children knew the exact label. 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 7.1% of the time. Some items were always identified correctly (e.g., apple; 100% identification rate), whereas others were rarely identified correctly (i.e., perfume bottle; 33.3% identification rate).

To assess the reliability of these item-level identification rates, we split our 24 participants into two halves and computed the correlation over the identification scores for the 40 items between each random half of participants, then repeated this procedure for 100 random splits of the participants. Averaging over these iterations, the average item-level identification split-half reliability was $r = .87$. Thus, this group of 4-year-olds were relatively consistent in their basic-level identification of these objects.

Results. For the critical analysis, we separated displays into two groups: (a) displays where the basic-level identities of both the big and small depicted objects were well identified (greater than 75% [8/20 pairs]; $M = 95.0\%$ across all 16 items) and (b) displays where one or more of the depicted objects was poorly identified (with recognition rates at 75% or less [12/20 pairs]; $M = 63.54\%$ across all 24 items). We found that pairs of objects that were well identified at the basic-level did not generate larger Size-Stroop effects in RTs ($M = -5.4$ ms) than did pairs of objects that were not both well identified ($M = 85.2$ ms; see the Appendix, Figure A5); if anything, we found a trend in the opposite direction: unpaired two-sample t test, $t(18) = -1.77$, $p = .09$.[8] For example, the Size-Stroop RT effect for the poorly recognized barbecue–die pair was 66 ms, whereas the Size-Stroop RT effect for the well-recognized desk–apple pair was -120 ms. Finally, we additionally examined whether 4-year-olds' Stroop error rates followed these RT effects. We found that children did not show greater Stroop error effects when the item pairs contained well-identified items versus when they contained poorly identified items (well-identified pairs of objects: $M = 2.72\%$ Stroop error rates; poorly identified pairs of objects: $M = 3.28\%$ Stroop error rate), $t(18) = -.40$, $p = .69$.

These results on both RTs and error rates provide some evidence that children's Stroop effects are not driven by first recognizing the object and then accessing the real-world size. However, the strength of this evidence is tempered by the fact that there were only a small number of item pairs ($N = 20$), and these were not specifically designed to probe for the role of explicit object recognition on the Size-Stroop effect.

In the next analysis, we considered the possibility that children might not recognize an object or describe its function but might systematically confuse it as some other object of a similar size. As examples, even though 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. So, we recoded the basic-level guesses not for correctness but for real-world size. As before, these ratings were relatively consistent across children (split-half reliability, average $r = .64$). Next the displays were again divided into two groups: (a) displays in which either object's size was correctly guessed at a rate above the median across all items (both items >87.5% correct [8/20 pairs]; $M = 97.7\%$ across items) and (b) displays that fell below the median (one or both items <87.5% correct [12/20 pairs]; $M = 80.6\%$ across items). This grouping method was used because size identification was relatively high overall. As before, we found that pairs of objects whose sizes were poorly identified generated Size-Stroop RT effects that were equivalent to those of pairs of objects whose sizes were well identified, $t(18) = -1.42$, $p = .17$. Further, Size-Stroop error rates also did not depend on whether displays contained items whose size was more or less identifiable ($M = 1.79\%$ vs. $M = 3.90\%$, respectively), $t(18) = 1.6$, $p = .13$.

Taken all together, these analyses show that there is remarkable consistency in the specific pairs of big and small objects that generated Size-Stroop effects in adults and children and suggest that explicit recognition of these objects was not a major mediating factor in the Size-Stroop effect for children. These analyses begin to shed light on the mechanisms supporting the Size-Stroop effect in children, indicating that similar mechanisms support real-world object size representations in adults and preschoolers. Broadly speaking, they are consistent with the hypothesis that preschoolers have abstracted the mid-level perceptual features that distinguish big from small objects, as classes, and that preschoolers' visual systems automatically use these features to compute the real-world size of pictured objects.

General Discussion

Across two experiments, we found evidence to suggest that preschoolers automatically activate real-world size information when they see pictured objects. Preschoolers were impaired at making visual size judgments about pictured objects when the relative sizes of the images were incongruent with their relative sizes in the real-world, even though the real-world size was not relevant to the task. This effect was evident in preschoolers' error patterns and RTs: 3- and 4-year-olds made more errors on incongruent displays, and 4-year-olds also took longer to make visual size judgments on incongruent displays. Further, item-pair analyses showed that the same pairs of big and small objects generated stronger Size-Stroop effects in 4-year-olds and adults, regardless of how well 4-year-olds could identify these depicted objects. Taken together, these results suggest that object size is automatically encoded in preschoolers' object representations and points toward the idea that similar mechanisms may underlie object size representations in both preschoolers and adults.

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-pair effects across adults and 4-year-old children. These results provide indirect evidence that 4-year-olds, just like adults, use mid-level perceptual features to directly infer the real-world size of objects (Long &

[8] Because 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 than did well-identified pairs of objects.

Konkle, 2017; Long et al., 2016). Further, this conclusion is also consistent with recent evidence from visual search paradigms that preschoolers are sensitive to the visual features that distinguish big objects from small objects (Long et al., 2019). However, it's also possible that the item-pair similarities between children and adults could be driven by properties of this specific stimulus set that are not related to mid-level feature information per se.

One obvious empirical route for confirming whether preschoolers' visual systems can also use mid-level perceptual features to infer real-world size would be to see whether children show the 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.

We see two ways of exploring this hypothesis. 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 with both children and adults. Delineating the key perceptual features that distinguish big objects from small objects, however, is still an area of active research (Long & Konkle, 2017; Long et al., 2016), so pursuing this empirical approach awaits further progress on that front.

The second way would be to follow up the suggestive data that children's recognition of objects in terms of basic-level or superordinate-level kinds is irrelevant to the effects of item pairs. Given that we did not explicitly manipulate how recognizable the items in the displays would be to children, it is possible that a greater disparity in recognition across Stroop displays could reveal a different pattern of results. In other words, the relative contributions of information derived from perceptual features versus object kind recognition to automatic size processing still remains an open question. Nonetheless, a straightforward prediction of the hypothesis that preschool children's Size-Stroop effects (like that of adults') can be driven by mid-level perceptual features alone would be that children can infer the real-world size of an object they cannot recognize at a basic or superordinate level.

Does Real-World Size Organize Neural Responses to Objects in Preschoolers?

Given that preschoolers show the same behavioral signatures of real-world size representation as do adults, one possibility is that preschoolers also show the same neural signatures of object size representation. 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 (Julian et al., 2017; Konkle & Caramazza, 2013; Konkle & Oliva, 2012b); these preferences are stable across changes in the retinal sizes of objects (Konkle & Oliva, 2012b) and are also elicited by unrecognizable versions of big and small objects (Long et al., 2018).

Thus, future neuroimaging 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. The few neuroimaging studies in infants and 5- to 7-year-old children have suggested that this may be plausible. For example, early regional preferences for faces versus houses are evident in infants (Deen et al., 2017), but the degree of selectivity of these regions develops through childhood and into adolescence (Golarai, Liberman, & Grill-Spector, 2017; Gomez, Natu, Jeska, Barnett, & Grill-Spector, 2018; Natu et al., 2016). One interesting possibility is that the broader division of big and small objects emerges early and helps to scaffold further category-selective responses. Consistent with this idea, the general overall similarity structure in neural responses to faces, objects, bodies, and scenes is similar between adults and children 5–7 years of age (Cohen et al., 2016), despite the further refinements to come in adolescence related to facial recognition and reading–writing abilities (e.g., Carey & Diamond, 1994; James, 2017).

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? One possibility is that younger infants and toddlers may first need to access basic-level representations (e.g., "bottle") before they can access size representations. In doing so over the first few years of life, they may then learn the mid-level features that are implicated in the processing of real-world size. However, another possibility is that 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 versus 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 (Granrud, Haake, & Yonas, 1985). If so, then the perceptual features of unfamiliar objects could already activate real-world size information in young infants.

Because this Size-Stroop paradigm was already difficult to run with 3-year-olds, future research will need to develop new methods to examine whether and how younger children activate real-world size information when they see pictured objects. An understanding of the mechanisms that lead to adultlike real-world size representations not only is an important question in developmental cognitive neuroscience but will inform theories of why and how real-world size organizes individuals' cognitive and neural representations of objects in adulthood.

References

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

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4 [Computer software]. Retrieved from https://www.fwspubs.org/doi/suppl/10.3996/092015-JFWM-087/suppl_file/10.3996_092015-jfwm-087.s3.pdf

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,* 79–102. http://dx.doi.org/10.1207/s15326942dn3101_5

Carey, S., & Diamond, R. (1994). Are faces perceived as configurations more by adults than by children? *Visual Cognition, 1,* 253–274. http://dx.doi.org/10.1080/13506289408402302

Cate, A. D., Goodale, M. A., & Köhler, S. (2011). The role of apparent size in building- and object-specific regions of ventral visual cortex. *Brain Research, 1388,* 109–122. http://dx.doi.org/10.1016/j.brainres.2011.02.022

Chiou, R., & Lambon 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,* 4662–4668. http://dx.doi.org/10.1523/JNEUROSCI.2829-15.2016

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. http://dx.doi.org/10.1167/16.12.776

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior, 8,* 240–247. http://dx.doi.org/10.1016/S0022-5371(69)80069-1

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.

Doherty, M. J., Campbell, N. M., Tsuji, H., & Phillips, W. A. (2010). The Ebbinghaus illusion deceives adults but not young children. *Developmental Science, 13,* 714–721. http://dx.doi.org/10.1111/j.1467-7687.2009.00931.x

Ebeling, K. S., & Gelman, S. A. (1988). Coordination of size standards by young children. *Child Development, 59,* 888–896. http://dx.doi.org/10.2307/1130256

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,* 1–17. http://dx.doi.org/10.1080/15248372.2015.1061528

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

Gelman, S. A., & Ebeling, K. S. (1989). Children's use of nonegocentric standards in judgments of functional size. *Child Development, 60,* 920–932. http://dx.doi.org/10.2307/1131033

Gliksman, Y., Itamar, S., Leibovich, T., Melman, Y., & Henik, A. (2016). Automaticity of conceptual magnitude. *Scientific Reports, 6,* 21446. http://dx.doi.org/10.1038/srep21446

Golarai, G., Liberman, A., & Grill-Spector, K. (2017). Experience shapes the development of neural substrates of face processing in human ventral temporal cortex. *Cerebral Cortex, 27,* bhv314. http://dx.doi.org/10.1093/cercor/bhv314

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,* 788. http://dx.doi.org/10.1038/s41467-018-03166-3

Gori, M., Del Viva, M., Sandini, G., & Burr, D. C. (2008). Young children do not integrate visual and haptic form information. *Current Biology, 18,* 694–698. http://dx.doi.org/10.1016/j.cub.2008.04.036

Granrud, C. E., Haake, R. J., & Yonas, A. (1985). Infants' sensitivity to familiar size: The effect of memory on spatial perception. *Perception & Psychophysics, 37,* 459–466. http://dx.doi.org/10.3758/BF03202878

Henik, A., Gliksman, Y., Kallai, A., & Leibovich, T. (2017). Size perception and the foundation of numerical processing. *Current Directions in Psychological Science, 26,* 45–51. http://dx.doi.org/10.1177/0963721416671323

James, K. H. (2017). The importance of handwriting experience on the development of the literate brain. *Current Directions in Psychological Science, 26,* 502–508. http://dx.doi.org/10.1177/0963721417709821

Jolicoeur, P., Gluck, M. A., & Kosslyn, S. M. (1984). Pictures and names: Making the connection. *Cognitive Psychology, 16,* 243–275. http://dx.doi.org/10.1016/0010-0285(84)90009-4

Julian, J. B., Ryan, J., & Epstein, R. A. (2017). Coding of object size and object category in human visual cortex. *Cerebral Cortex, 27,* 3095–3109.

Jüttner, M., Wakui, E., Petters, D., & Davidoff, J. (2016). Developmental commonalities between object and face recognition in adolescence. *Frontiers in Psychology, 7,* 385. http://dx.doi.org/10.3389/fpsyg.2016.00385

Khaligh-Razavi, S. M., Cichy, R. M., Pantazis, D., & Oliva, A. (2018). Tracking the spatiotemporal neural dynamics of real-world object size and animacy in the human brain. *Journal of Cognitive Neuroscience, 30,* 1559–1576. http://dx.doi.org/10.1162/jocn_a_01290

Konkle, T., & Caramazza, A. (2013). Tripartite organization of the ventral stream by animacy and object size. *Journal of Neuroscience, 33,* 10235–10242. http://dx.doi.org/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,* 561–569. http://dx.doi.org/10.1037/a0028294

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

Long, B., & Konkle, T. (2017). A familiar-size Stroop effect in the absence of basic-level recognition. *Cognition, 168,* 234–242. http://dx.doi.org/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, 145,* 95–109. http://dx.doi.org/10.1037/xge0000130

Long, B., Moher, M., Carey, S., & Konkle, T. (2019). *Animacy and real-world size are reflected in perceptual similarity computations by the preschool years*. Manuscript submitted for publication.

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 of the United States of America, 115,* E9015–E9024. http://dx.doi.org/10.1073/pnas.1719616115

Natu, V. S., Barnett, M. A., Hartley, J., Gomez, J., Stigliani, A., & Grill-Spector, K. (2016). Development of neural sensitivity to face identity correlates with perceptual discriminability. *Journal of Neuroscience, 36,* 10893–10907. http://dx.doi.org/10.1523/JNEUROSCI.1886-16.2016

Nishimura, M., Scherf, S., & Behrmann, M. (2009). Development of object recognition in humans. *F1000 Biology Reports, 1,* 56. http://dx.doi.org/10.3410/B1-56

Paivio, A. (1975). Perceptual comparisons through the mind's eye. *Memory & Cognition, 3,* 635–647. http://dx.doi.org/10.3758/BF03198229

Sellaro, R., Treccani, B., Job, R., & Cubelli, R. (2015). Spatial coding of object typical size: Evidence for a SNARC-like effect. *Psychological Research, 79,* 950–962. http://dx.doi.org/10.1007/s00426-014-0636-7

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

Whelan, R. (2008). Effective analysis of reaction time data. *Psychological Record, 58,* 475–482. http://dx.doi.org/10.1007/BF03395630

Appendix

Supplemental Figures and Analyses

Figure A1. Examples of recognizable images and their corresponding texforms, for a group of three big objects (left) and three small objects (right; Freeman & Simoncelli, 2011; Long, Konkle, Cohen, & Alvarez, 2016). See the online article for the color version of this figure.

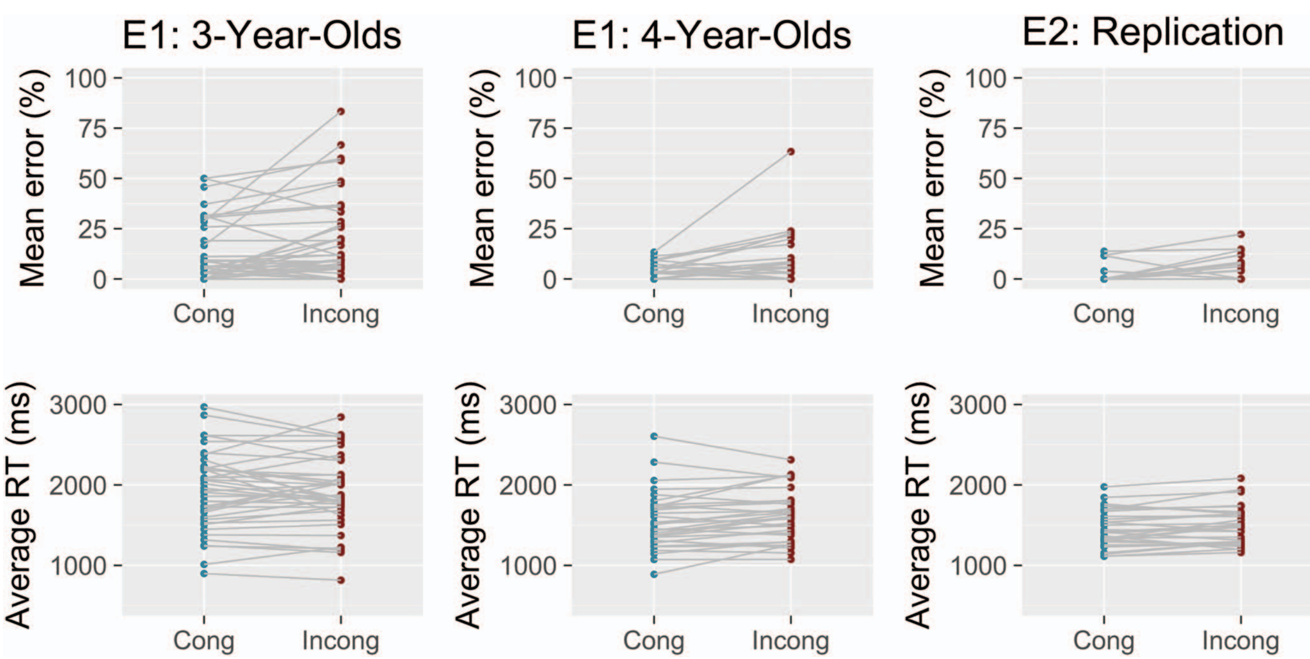

Figure A2. Average error rates (upper panels) and reaction times (RTs; lower panels) are shown for each individual child. Each dot reflects a child's average performance in one condition, with gray lines connecting their performance in both conditions. The first two columns reflect data from Experiment 1 (E1) and the last column shows data from the replication Experiment 2 (E2). See the online article for the color version of this figure.

(*Appendix continues*)

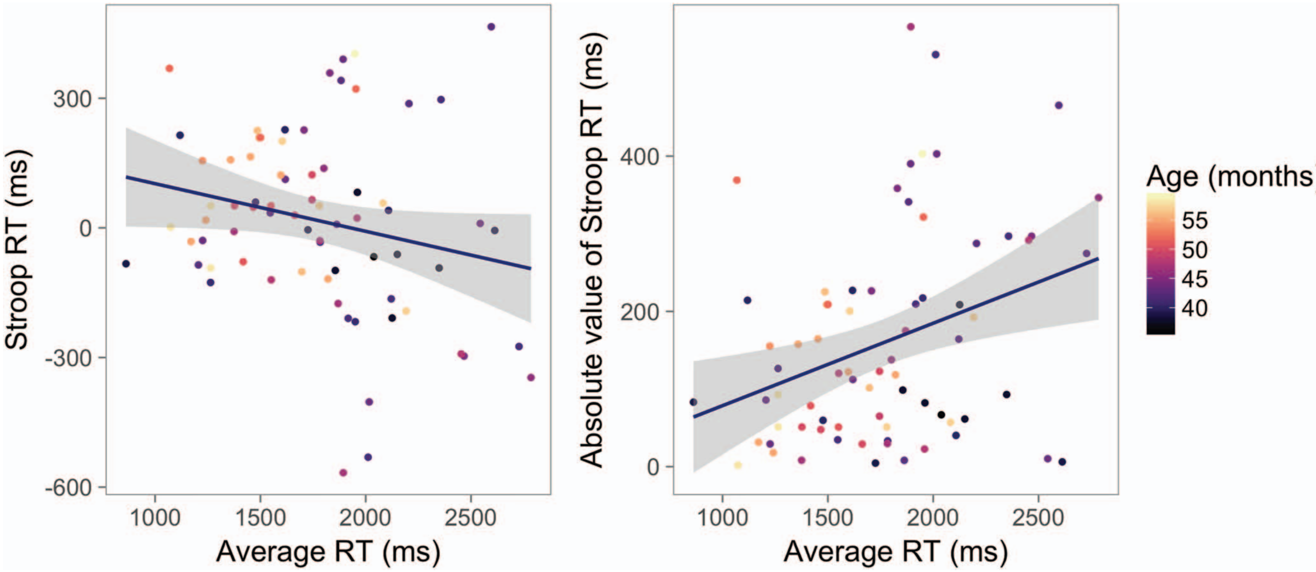

Figure A3. Size-Stroop effects for individual children are shown as a function of children's overall reaction time (RT) on the task across both conditions in the left panel. The absolute difference between incongruent and congruent conditions (i.e., the absolute value of the Size-Stroop effect) is plotted for all participants as a function of their average speed on the task in the right panel. See the online article for the color version of this figure.

(*Appendix continues*)

A. Images shown

B. 4-year-olds responses

BBQ	**Camping food cookingtable**
Desk	Machine
Grill	**Grill**
(no response)	**Grill**
(no response)	**Kitchen oven**
Kitchen Table	Something…
TV	**Grill**
Balcony/Cook	**Grill**
Table	**Grill**
(no response)	(no response)
Desk	**Grill**
Cook Thing	**A grill**

(no response)	Binoculars
Music Box	**Pencil Sharpener**
(no response)	(no response)
Spy Graph	Radio
(no response)	I don't know…never seen it before
Remote Control	Something with nails in it
Camera	Remote control
Tool Thingy	What you put on your eyes to see far away
Drawer	Binoculars
Binoculars	it looks like a plug
Tool	Binoculars
Instrument	Binoculars

Figure A4. Twenty-four 4-year-olds were asked "What does this look like?" about the depicted object shown in Panel A. Their responses are shown in Panel 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 (Panel A, left panel). See the online article for the color version of this figure.

(*Appendix continues*)

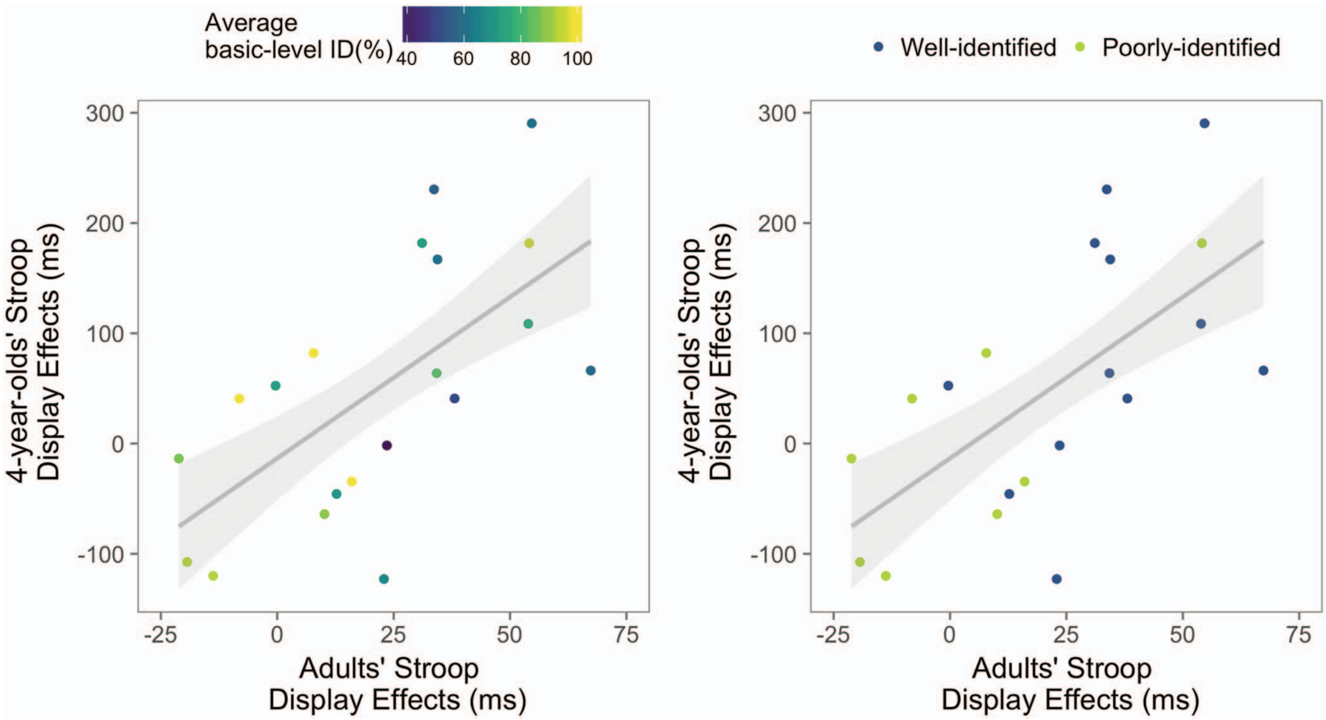

Figure A5. Size-Stroop item-pair effects (incongruent–congruent reaction time) are plotted for each pair of objects for all 4-year-olds in Experiments 1 and 2 as a function of adults' Stroop effects for the same pairs of displays. Shaded regions represent 95% confidence intervals around the mean. On the left panel, item pairs are colored the average recognizability of both items in a given pair. On the right panel, item pairs are colored according to whether preschoolers were able to recognize both images in a pair greater than 75% of the time, based on the experiment described in the Analysis 2: Object Identification section. See the online article for the color version of this figure.

Received August 14, 2018
Revision received November 14, 2018
Accepted November 21, 2018 ■