

**Modulation of visually-guided action by the image and familiar sizes of real-world  
objects**

Christine Gamble<sup>1</sup>, Joo-Hyun Song<sup>1,2</sup>

<sup>1</sup> Cognitive, Linguistic, & Psychological Sciences

Brown University

<sup>2</sup> Carney Institute for Brain Science

Brown University

Address for correspondence

Joo-Hyun Song

Department of Cognitive, Linguistic, and Psychological Sciences

Box 1821

Brown University

Providence, RI 02912

Email: Joo-Hyun\_Song@brown.edu

## Abstract

In daily life, two aspects of real-world object size perception—the *image size* of an object and its *familiar size* in the real world are highly correlated. Thus, whether these two aspects of object size differently affect goal-directed action (e.g., manual pointing), and how, has scarcely been examined. Here participants reached to touch one of two simultaneously presented objects based on either their image or familiar size, which could be *congruent* or *incongruent* (e.g., a rubber duck presented as smaller and larger than a boat, respectively). We observed that when pointing to target objects in the incongruent conditions, participants' movements were slower and were more curved towards the incorrect object compared to the movements in the congruent conditions. By comparing performance in the congruent and incongruent conditions, we concluded that both image size and familiar size influenced action even when task-irrelevant, indicating that both are processed automatically (Konkle & Oliva, 2012a). Image size, however, showed influence earlier in the course of movements and more robustly overall than familiar size. We additionally found that greater relative familiar size differences mitigated the impact of image size processing, and increased the impact of familiar size processing on pointing movements. Overall, our data suggest that image and familiar size perception interact both with each other and with visually-guided action, but that the relative contributions of each are unequal and vary based on task demands.

## 1 Introduction

2 In 2007 an art installation appeared floating in the harbor of Saint-Nazaire, a small town  
3 in Western France. “Rubber Duck” by Florentijn Hofman quickly gained international attention  
4 due to the unexpected scene created by a toy that typically measures 5.5 cm in height dwarfing  
5 nearby boats at 32 m. The *familiar size* of a rubber duck, the size that we know it typically would  
6 be based on past experience, and the *image size* of “Rubber Duck,” the size the piece appeared  
7 visually to viewers, were dramatically in conflict.

8 In normal daily life, objects’ *familiar size* and *image size* are highly correlated. When  
9 presented in the same context, real-world objects like rubber ducks which we know to be  
10 relatively small in the real world typically appear smaller than objects like boats which we know  
11 to be larger. Even when the sizes of objects on the retina vary, they are integrated with their  
12 environment via size constancy mechanisms. Thus, taking size constancy into account, in the  
13 real-world *image size* and *familiar size* are very rarely in conflict. Consequently, *image size* and  
14 *familiar size* processing are highly confounded—when we see an object and its image size and  
15 familiar size are congruent, how each of these two aspects of size affects our perception and  
16 goal-directed action is difficult to disambiguate.

17 Attempts have recently been made to disentangle image size and familiar size perception.  
18 Konkle & Oliva (2012a) implemented a Stroop-like paradigm, in which a pair of two objects  
19 were presented at different image sizes on the screen. Participants were asked to indicate which  
20 image size was bigger or smaller by key-press, while their familiar sizes were task-irrelevant.  
21 This experiment demonstrated that incongruence between familiar size and image size (e.g., a  
22 rubber duck presented with a larger image size than a boat) results in a “familiar size Stroop  
23 effect,” captured by slower reaction times for image size judgments. This result suggests that

1 these two highly different aspects of real-world object size, *image* size and *familiar* size, are both  
2 processed automatically, as in the classic Stroop Effect (Stroop, 1935).

3       However, despite disambiguating the two to some degree, Konkle & Oliva (2012a) only  
4 examined the unidirectional influence of familiar size on image size, as opposed to the  
5 bidirectional influence of each on the other. Furthermore, binary categories were used for large  
6 and small image sizes, and large and small familiar sizes, though both exist on a continuum in  
7 the real world. Therefore, the relative contributions of each type of object size on perception, and  
8 other visual processes such as guiding goal-directed actions, bear further investigation.

9       Object size is a key component of *vision for action* in addition to *vision for perception*,  
10 given that size constrains how we interact with the objects in our environment. For example, the  
11 width of a target modulates the speed of goal-directed pointing movements to it (Fitts, 1954), and  
12 grip aperture scales to the image sizes of objects (Jeannerod, 1984). Prior research has also  
13 extensively explored how dissociating the veridical physical and perceived image sizes of objects  
14 using contextual size illusions such as the Ebbinghaus, Ponzo, and Muller-Lyer illusions, affects  
15 various goal-directed action such as manual pointing and grasping as well as eye movements  
16 (e.g., Bernardis, Knox, Bruno, 2005; Binsted & Elliott, 1999; de Grave, Franz, & Gegenfurtner,  
17 2006; Franz, 2001; Glover & Dixon, 2002; Milner & Goodale, 2008; Knol et al., 2017; Gamble &  
18 Song, 2017).

19       However, less is known about how familiar size influences action. Image size is a low-  
20 level visual feature—it is represented retinally and in V1, one of the earliest visual processing  
21 regions (Murray, Boyaci, & Kersten, 2006). In contrast, familiar size is a high-level visual  
22 feature requiring object identification and the recruitment of memory, and is represented in a  
23 later visual processing region in the temporoparietal cortex (Konkle & Oliva, 2012b).

1           Though conflict between image and familiar size was demonstrated in the  
2   aforementioned paper (Konkle & Oliva, 2012a) using a discrete behavioral response paradigm, it  
3   is not known how action would respond to conflict between the two, again given that such  
4   scenarios are highly uncommon in the real world. It has been shown that discrete behavioral  
5   responses and action measures can produce different results for the same perceptual decision-  
6   making task. For example, it is typically thought that strongly salient distractors capture more  
7   attention and are more disruptive than weakly salient distractors (Theeuwes, 2010; Itti & Koch,  
8   2001). Counterintuitively, Moher et al. (2015) found dissociable effects of salience on discrete  
9   key-press and goal-directed action such as pointing. In a visually-guided pointing task, they  
10   required participants to reach to a shape-defined target while trying to ignore salient distractors.  
11   They observed that highly salient objects impacted hand movement trajectories *less* than less  
12   salient objects did. Thus, a strongly salient distractor triggers suppression during goal-directed  
13   action, resulting in enhanced efficiency and accuracy of target selection relative to when weakly  
14   salient distractors are present. In contrast, in a task requiring key-press to select a target, they  
15   found greater attentional interference from strongly salient distractors, reflected on slower  
16   reaction time. This counterintuitive result suggests that sufficiently strong distractors may trigger  
17   suppression, but only when a physical movement is required. These results also underscore the  
18   value and necessity of combining visually-guided actions with traditional perceptual approaches  
19   to fully understand how we resolve competing internal processes to achieve behavioral goals  
20   (Moher, Anderson, & Song, 2015).

21           Furthermore, the continuous nature of manual pointing movements has provided new  
22   insights into the temporal evolution of cognitive processes including language processing,  
23   numerical cognition, attention allocation, social perception, and cognitive control (for review,

Song & Nakayama, 2009; Song, 2017; Erb, 2018; Dotan et al., 2019). Previous studies have demonstrated that the evolution of reach trajectory curvature while selecting a target among alternative choices can reveal various aspects of decision-making such as timing of information process and the degree of competition (Erb, Moher, Sobel, & Song, 2016; McKinsty, Dale, & Spivey, 2008; Song, 2017; Song & Nakayama, 2009; Moher & Song, 2019). For instance, Song & Nakayama (2009) asked participants to determine whether a single Arabic numeral presented in a central square was less than, greater than, or the same as the number five by reaching to one of three corresponding squares on the screen. They observed that the greater the numeric deviation between the target and the number five, the greater the deviation of the trajectory from the standard trajectory. This provides direct evidence that the numeric magnitude of a target is spatially encoded, and that the proximity and order of numbers are spatially represented along a hypothesized mental number line. Such a methodology can therefore provide a tool to track how competition between the processing of image and familiar sizes evolves and in turn resolves over time.

Overall, gaps in the literature we seek to address in the present study are the potential influence of familiar size (a higher-level visual feature) on action, how it compares to the influence of image size (a lower-level visual feature) on action, and whether these interactions change over time. In order to address these questions, we employed a similar paradigm to Konkle & Oliva's aforementioned Stroop-like paradigm (Konkle & Oliva, 2012a), with the addition of A) a familiar size judgment task in which image size was the conflicting task-irrelevant feature, B) parametrically-varied familiar object sizes, and C) continuous action-based responses as opposed to discrete behavioral responses. In Experiment 1, we tested the hypothesis that the higher-level feature *familiar size* and the lower-level feature *image size* significantly

interfere with each other when the two are incongruent, while one type of size is task-irrelevant. Furthermore, we sought to investigate the relative strength of interference between familiar and image size processing by measuring temporal aspects (e.g., how fast movements are initiated) and spatial aspects (e.g., how much reach trajectories are curved towards a wrong choice) of goal-directed pointing movements. In Experiment 2, we evaluated the hypothesis that if familiar size is treated as a spectrum, varying the magnitudes of relative familiar size differences between pairs objects may lead to corresponding graded effects on action and on the interference between image and familiar size.

Overall, we hypothesized bidirectional interference between image and familiar size, but recognized that the influence of each might not be symmetrical. In other words, they might both exert influence on visually-guided action, but not to the same degree or on the same timescale. Specifically, in accord with classical models of object processing (Rosch et al., 1976; Collins & Quillian, 1969; Jolicoeur, Gluck, & Kosslyn, 1984), if image size is a more intrinsic feature with stronger, more fundamental neural correlates, it could cause more interference when it is task-irrelevant and more strongly *resist* interference when it is task-relevant. In this case we would expect to see image size exert a relatively strong influence in the familiar size task, and familiar size exert a relatively weak influence in the image size task. Furthermore, if there is a temporal component to the asymmetry, we would expect to see image size influence movements earlier in the familiar size task, and familiar size influence movements later in the image size task.

On the other hand, if familiar size is represented more strongly, we would see familiar size exert a relatively strong influence in the image size task, and image size exert a relatively weak influence in the familiar size task. Similarly, we could expect to see familiar size influence

movements earlier in the image size task, and image size influence movements later in the familiar size task.

## **Experiment 1: Congruency between target image and familiar size in visually-guided action**

### **Methods**

#### *Participants*

Fourteen right-handed participants (8 women; mean age 25.7 years) with normal color vision, and normal or corrected-to-normal visual acuity completed both tasks within Experiment 1. Participants provided their informed consent and were compensated monetarily (\$10 per hour) or with course credit for their participation. The experimental protocol was approved by the Brown University Institutional Review Board in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans.

#### *Apparatus*

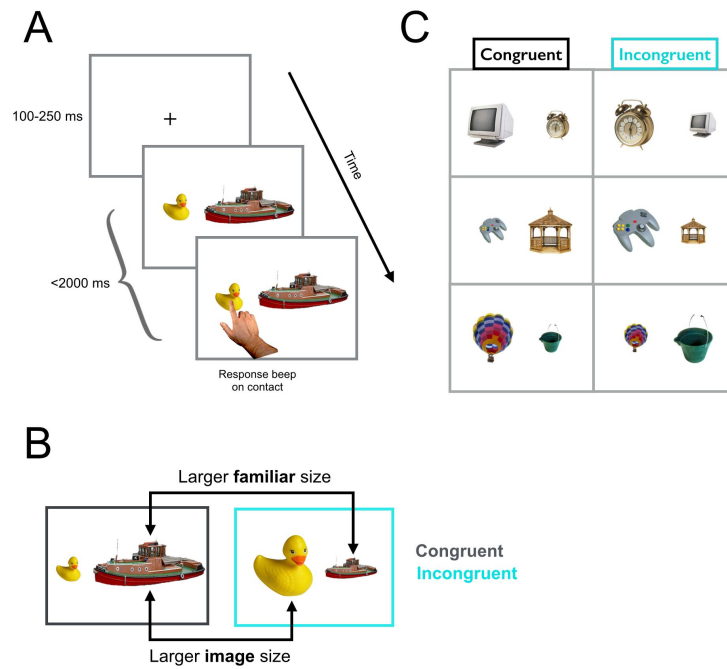
Stimuli were presented on an upright Plexiglas display facing the seated participant at a distance of approximately 55 cm. A projector behind the display projected a screen measuring  $44.1 \times 33.0$  cm ( $43.7^\circ \times 33.4^\circ$  visual angle), which participants viewed binocularly. Stimulus presentation was conducted using custom software designed with MATLAB (MathWorks, Natick, MA) and Psychtoolbox (Brainard, 1997). Three-dimensional hand position was recorded with an electromagnetic position and orientation recording system (Liberty; Polhemus,



Colchester, VT) at a rate of 160 Hz with a measuring error of 0.3 mm root mean square. A motion-tracking marker was fastened to the tip of each participant's right index finger using a Velcro strap. A foam starting block placed 27 cm in front of the participant, between the participant and the display (28 cm from the display), served as the starting position on which the index finger rested at the beginning of each trial.

### *Stimuli*

Stimuli were adapted from Konkle & Oliva's "Object Size Stroop" database (Konkle & Oliva, 2012a). All stimuli were presented on a white background. A black eye fixation cross, measuring  $7 \times 7$  mm ( $0.73^\circ$  of visual angle) appeared at the center of the screen before each trial. As shown in Figure 1A two real world objects were displayed, one to each side of fixation ( $14.2$  cm or  $14.7^\circ$  measured from fixation to target center (35.5 cm diagonally from the foam starting block). Sixteen unique object pairs were used each for the *image size task*, which required participants to judge the sizes of the objects as they appeared on the screen, and for the *familiar size task*, which required participants to judge the sizes of the objects in the real world. An additional two unique object pairs were used for pre-experiment practice blocks. Thus, 34 total object pairs were used, with each individual object being assigned to only one other object. Example object pairs are shown in Figure 1B.



**Figure 1.** Task and Stimuli. **A.** While the index finger was positioned on the starting block, trials began with the eye-fixation cross for a variable period between 100 and 250 ms. Once stimuli were presented, participants had up to 2000 ms to complete the reach-to-touch hand movement indicating their choice. **B.** The objects' *image size* and *familiar size* could be presented as congruent (e.g. the left image where the rubber duck has a smaller image size than a boat) or incongruent (e.g. the right image where the rubber duck has a larger image size than a boat). Additionally, objects were counterbalanced such that the larger and smaller image size and familiar sizes were presented in both left and right positions. **C.** Examples of paired objects. The left column shows examples of *incongruent* conditions and the right-side *congruent* conditions.

In each trial one object had a larger relative familiar size and the other a smaller relative familiar size (see Figure 1C). Familiar sizes of objects ranged from 3 cm diagonally (a die) to

8776 cm diagonally (a cathedral), as reported by Konkle & Oliva (2012a). In parallel, each trial had one object presented with a larger image size and one presented with a smaller image size. Image sizes, as reported by Konkle and Oliva (2012a), were designed such that the objects with “small” image sizes were bounded by a rectangle with a diagonal equal to 17.5% of our screen height (5.8 cm or 6.0°) and objects with “large” image sizes were bounded by a rectangle with a diagonal equal to 30% of our screen height (9.9 cm or 10.3°). This method was selected to account for the variations in real-world objects’ aspect ratios (Konkle & Oliva, 2011, 2012a; Kosslyn, 1978).

In congruent trials, the object with the larger *familiar* size was presented with a larger *image* size and the object with the smaller *familiar* size was presented with a smaller *image* size. For example, in a congruent trial a rubber duck would be presented as 5.8 cm diagonally and a boat as 9.9 cm diagonally. In an incongruent trial, relative *familiar* and *image* sizes would be incongruent, e.g. a rubber duck would be presented as larger on the screen (9.9 cm diagonally) and a boat as smaller on the screen (5.8 cm diagonally).

### *Procedure*

Two blocks of each task (image or familiar size judgment) were performed in an ABBA order, with image and familiar size tasks assigned as A or B randomly. Task instructions appeared at the beginning of each block (“Make your selections based on the sizes you know the objects are in the real world” for the familiar size task, and “Make your selections based on the sizes of the objects on the screen” for the image size task).

Each block was broken into two sequential sub-blocks, each of which contained trials with the same task instructions (“Choose the object that is larger” or “Choose the object that is

smaller”). For example, in an *image size* task block, for the first sub-block participants would select the object with the larger image size, and in the second sub-block, the object with the smaller image size. The order of these sub-blocks was randomly determined (i.e. larger-smaller or smaller-larger), and reversed in the second block of each task. In total each of the 32 experimental object pairs was repeated eight times for a total of 128 trials per experimental task.

At the beginning of each participant’s session, nine-point calibration was conducted for the tracker. Participants also completed a practice block of 16 trials of the relevant size-judgment task before the first experimental block of that task. In both the image size and familiar size tasks, once participants rested their index finger on the starting position, each trial began with an initial fixation-cross presented for a variable amount of time (100–250 ms). Participants were asked to hold their eyes on the fixation cross though it was not monitored via eye tracker. This was followed by the presentation of the object pair. Participants were instructed to make their decisions as quickly and accurately as possible, by reaching out and touching the selected object on the screen. In both tasks, the display was presented for a maximum of 2000 ms. Auditory feedback was given when participants touched the display indicating either correct object selection (single high beep) or incorrect object selection (single low beep), or if the time limit had been exceeded (double low beep). An example trial of the task is diagrammed in Figure 1B.

### *Data analysis*

Data analysis procedures were largely adapted from methods reported in our previous work (Gamble & Song, 2017; Moher & Song, 2013, 2014).

Using custom MATLAB (MathWorks) software, we conducted off-line data analysis on the pointing movement data. Movement velocity was calculated from the 3D position traces after

1 filtering with a low-pass filter (cutoff frequency of 10 Hz). An algorithm using velocity criteria  
2 of 10 cm/s detected the beginning and end of pointing movements. The algorithm's identification  
3 of these movements was visually inspected to verify its accuracy for every trial (Gamble & Song,  
4 2017; Moher et al., 2015; Moher & Song, 2013, 2014); for trials in which the default threshold  
5 clearly missed capturing part of the movement or included substantial post-selection movement,  
6 thresholds were adjusted manually to more appropriate levels for that trial (~ 1% of all trials).

7 Pointing movements were classified as correct responses if they landed within a  
8 standardized target boundary used for all targets ( $6 \times 6$  cm or  $6.2^\circ \times 6.2^\circ$  of visual angle). Thus,  
9 *accuracy* was defined as the percentage of correct responses.

10 *Initiation latency* (IL) was defined as the time elapsed between stimulus onset and  
11 pointing movement onset. *Movement time* (MT) was defined as the time elapsed between  
12 movement onset and movement offset/target landing. *Maximum curvature* was calculated by  
13 tracing the path of the hand, and calculating an ideal direct path between the movement's starting  
14 and end points. The perpendicular deviation of the hand position from the ideal path was  
15 calculated at every time point over the course of the movement. The maximum of these  
16 perpendicular deviation lengths divided by the length of the ideal path results in a unitless ratio,  
17 referred to as *maximum curvature*. Larger ratios thus represent a greater maximum deviation  
18 from the ideal path, and greater overall curvature in the movement's trajectory. In order to  
19 compare congruent and incongruent trials within subjects for each measure (IL, MT and  
20 maximum curvature), and compare the *image* size and *familiar* size tasks directly, we performed  
21 2x2 repeated measures ANOVAs with factors of congruency (congruent vs. incongruent) and  
22 task (image size vs. familiar size judgments) for each measure. We also report partial eta  
23 squared, with values of .2 indicating a small effect size, .5 indicating a medium effect size, and .8

1 indicating a large effect size (Cohen, 1973). Within each task (image size and familiar size) we  
2 additionally performed paired t-tests comparing congruent and incongruent trials within subjects  
3 for each measure (accuracy, IL, MT and maximum curvature).

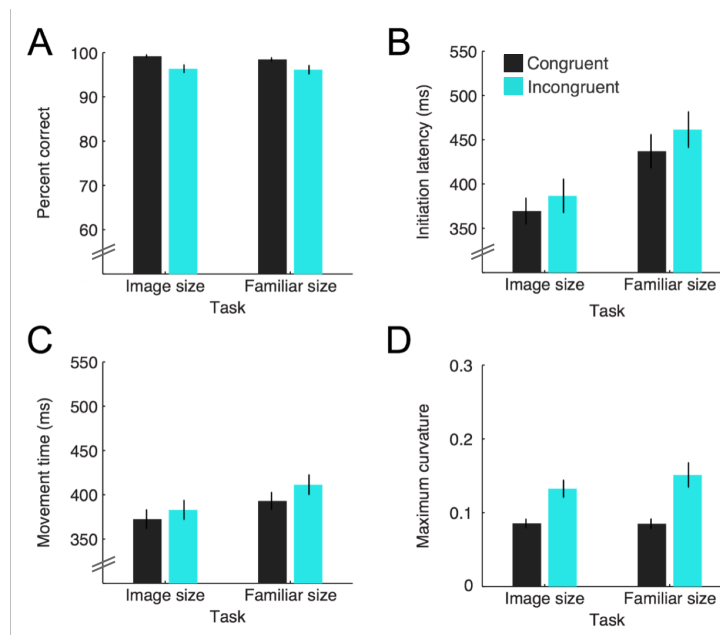
4 In addition to these analyses, we also examined the evolution of the reach trajectories  
5 over the course of the hand movements to the screen. In order to average and compare across  
6 participants and across trials, which naturally vary in length, we normalized all hand movements.  
7 First, movements to the left and right sides of the screen were collapsed. Additionally, we  
8 normalized all movements for each participant to 101 evenly spaced data points based on the  
9 linear distance of the hand from its starting point to the screen (the z-dimension, or “forward”  
10 from the participant). Then, to directly compare performance between the congruent and  
11 incongruent trials in each task, we focused on the lateralized horizontal movement on the x-  
12 dimension (left-right direction), the dimension along which the two competing potential targets  
13 differ. Specifically, at each of the 101 points, we calculated a difference between the averaged  
14 trajectories on the x-dimension between the congruent and incongruent conditions ( $x\text{-pos}_{\text{incongruent}}$   
15  $- x\text{-pos}_{\text{congruent}}$ ). The resultant difference score was calculated as positive if there was a  
16 measurable difference between the average position for the congruent and the incongruent  
17 conditions, indicating attraction of the hand movement toward the incorrect response alternative  
18 and significant interference from the task-irrelevant feature.

19 Comparing this difference score to zero reveals the points at which there was a  
20 significant difference in position between congruent and incongruent trials. Significant  
21 differences in position indicate interference from the task-irrelevant feature (e.g., in the *familiar*  
22 *size* task, difference scores significantly above zero indicate significant interference from *image*  
23 *size*).

These normalized positions were analyzed using a cluster-based permutation test. In order to correct for the multiple comparisons arising from conducting 101 t-tests, we used the Monte Carlo method to sample the datapoints in 500 iterations and find those points that were more significant (at a 95% confidence level) than the calculated test statistic. This analysis was performed using the Fieldtrip toolbox for Matlab (Oostenveld, Fries, Maris, & Schoffelen, 2011). The key datapoint in this subset is the first datapoint that was significantly greater than zero, indicating the point on average at which the trajectories of congruent and incongruent trials begin to deviate.

## Results

We excluded 5.34% ( $\pm 1.09\%$  s.e.) of trials from data analysis for the image size task and 5.19% ( $\pm 1.48\%$ ) for the familiar size task due to technical issues (e.g., expected occasional sampling drop). Accuracy was 97.47% ( $\pm 2.79\%$ ) for the image size task and 97.17% ( $\pm 2.83\%$ ) for the familiar size task. All subsequent analyses were restricted to correct trials.



**Figure 2.** Results for congruent versus incongruent trials in the image size and familiar size tasks in Experiment 1. All error bars represent between-subjects standard error. **A.** Accuracy. Accuracy was lower for the incongruent conditions than the congruent conditions in both tasks. **B.** Initiation latency. IL was faster for congruent compared to incongruent trials in both tasks, and faster overall for the image size task. **C.** Movement time. MT was faster for the congruent conditions in both tasks, and faster overall for the image size task. **D.** Maximum curvature. Movements were more curved in the incongruent trials for both tasks. Note, curvature values represent a unitless ratio.

#### *Image size task*

As seen in Figure 2A, in the image size task (leftmost bars) participants performed more accurately on congruent trials (black bars) compared to incongruent trials (cyan bars). This result indicates that interference from familiar size was strong enough to lead to incorrect choices when familiar size was in conflict with image size. This effect was supported by an overall main effect of congruency  $F(1,13) = 5.35$ ,  $p = .038$ ,  $\eta_p^2 = .292$ , with no interaction between congruence and task,  $F(1,13) = .71$ ,  $p = .794$ ,  $\eta_p^2 = .005$ . Further, a post-hoc t-test between congruent and incongruent trials for the image size task similarly showed a significant effect,  $t(13) = 2.83$ ,  $p = .014$ ,  $d = .697$ .

A congruency effect was also seen in initiation latency (Figure 2B) such that participants were faster overall to initiate movements when image size and familiar size were congruent compared when they were incongruent,  $F(1,13) = 28.02$ ,  $p < .001$ ,  $\eta_p^2 = .683$  with no interaction between congruence and task,  $F(1,13) = .88$ ,  $p = .365$ ,  $\eta_p^2 = .063$ ;  $t(13) = -2.84$ ,  $p = .014$ ,  $d = -.199$ . Similarly, congruency also affected online movement time (Figure 2C) such that



movements were executed more quickly overall when image size and familiar size were congruent compared to incongruent,  $F(1,13) = 41.14$ ,  $p < .001$ ,  $\eta_p^2 = .760$  with no interaction between congruence and task,  $F(1,13) = .519$ ,  $p = .484$ ,  $\eta_p^2 = .038$ ;  $t(13) = -3.62$ ,  $p < .001$ ,  $d = -.259$ .

Finally, we examined maximum curvature (Figure 2D) over the course of the movement as a measure of how much movements were pulled toward the incorrect object before ultimately landing on the correct object. Figure 2D shows that the average maximum curvature of pointing movements was greater when image size and familiar size were incongruent compared when they were congruent, showing that familiar size interfered with the image size task (e.g., when the task was to choose the larger object based on its image size, participants' hand movements were drawn toward the object with the larger *familiar* size),  $F(1,13) = 43.24$ ,  $p < .001$ ,  $\eta_p^2 = .760$  with no interaction between congruence and task,  $F(1,13) = .79$ ,  $p = .391$ ,  $\eta_p^2 = .057$ ;  $t(13) = -5.24$ ,  $p < .001$ ,  $d = -1.18$ .

Overall, the results from the image size task indicate that the higher-level feature *familiar size* significantly interferes with the lower-level *image size* when the two are incongruent, even when familiar size is task-irrelevant. This is consistent with prior findings on the “familiar size Stroop task” (Konkle & Oliva, 2012a).

#### *Familiar size task*

Figure 2A shows that, as in the image size task, participants were more accurate on average for congruent compared to incongruent trials for the familiar size task (rightmost bars). This demonstrates that image size interferes with familiar size even when it is task-irrelevant, suggesting a bidirectional influence of both aspects of size on reaching. This effect was

supported by the aforementioned main effect of congruency overall,  $F(1,13) = 5.35$ ,  $p = .038$ ,  $\eta_p^2 = .292$ , with no interaction between congruence and task, and a post-hoc t-test between congruent and incongruent trials for the familiar size task,  $t(13) = 2.43$ ,  $p = .030$ ,  $d = .802$ . Similarly, there was also an effect of congruency in IL (Figure 2B),  $F(1,13) = 28.02$ ,  $p < .001$ ,  $\eta_p^2 = .683$  with no interaction between congruence and task,  $t(13) = -3.58$ ,  $p = .003$ ,  $d = -.328$  and MT (Figure 2C),  $F(1,13) = 41.14$ ,  $p < .001$ ,  $\eta_p^2 = .760$  with no interaction between congruence and task,  $t(13) = -2.95$ ,  $p = .011$ ,  $d = -0.461$  such that participants were slower when image size and familiar size were incongruent. Finally maximum movement curvature (Figure 2D) was significantly greater when image size and familiar size were incongruent, as was also the case in the image size task,  $F(1,13) = 43.24$ ,  $p < .001$ ,  $\eta_p^2 = .760$  with no interaction between congruence and task,  $t(13) = -3.91$ ,  $p = .002$ ,  $d = -0.876$ .

Overall, these results suggest that there is significant interference from image size in the familiar size task such that performance is impaired when image and familiar size are incongruent, despite the fact that image size here is task-irrelevant.

#### *Comparing image and familiar size*

We additionally compared performance on the image and familiar size tasks in all the aforementioned measures. We observed no difference between the tasks for accuracy,  $F(1,13) = .068$ ,  $p = .799$ ,  $\eta_p^2 = .005$ , with no interaction between congruency and task  $F(1,13) = .071$ ,  $p = .794$ ,  $\eta_p^2 = .005$ , or for maximum curvature,  $F(1,13) = .636$ ,  $p = .439$ ,  $\eta_p^2 = .047$ , with no interaction between congruency and task  $F(1,13) = .789$ ,  $p = .391$ ,  $\eta_p^2 = .057$ . For initiation latency we observed that participants were faster to initiate movements in the image size judgment task compared to the familiar size judgment task,  $F(1,13) = 69.22$ ,  $p < .001$ ,  $\eta_p^2 = .842$ ,

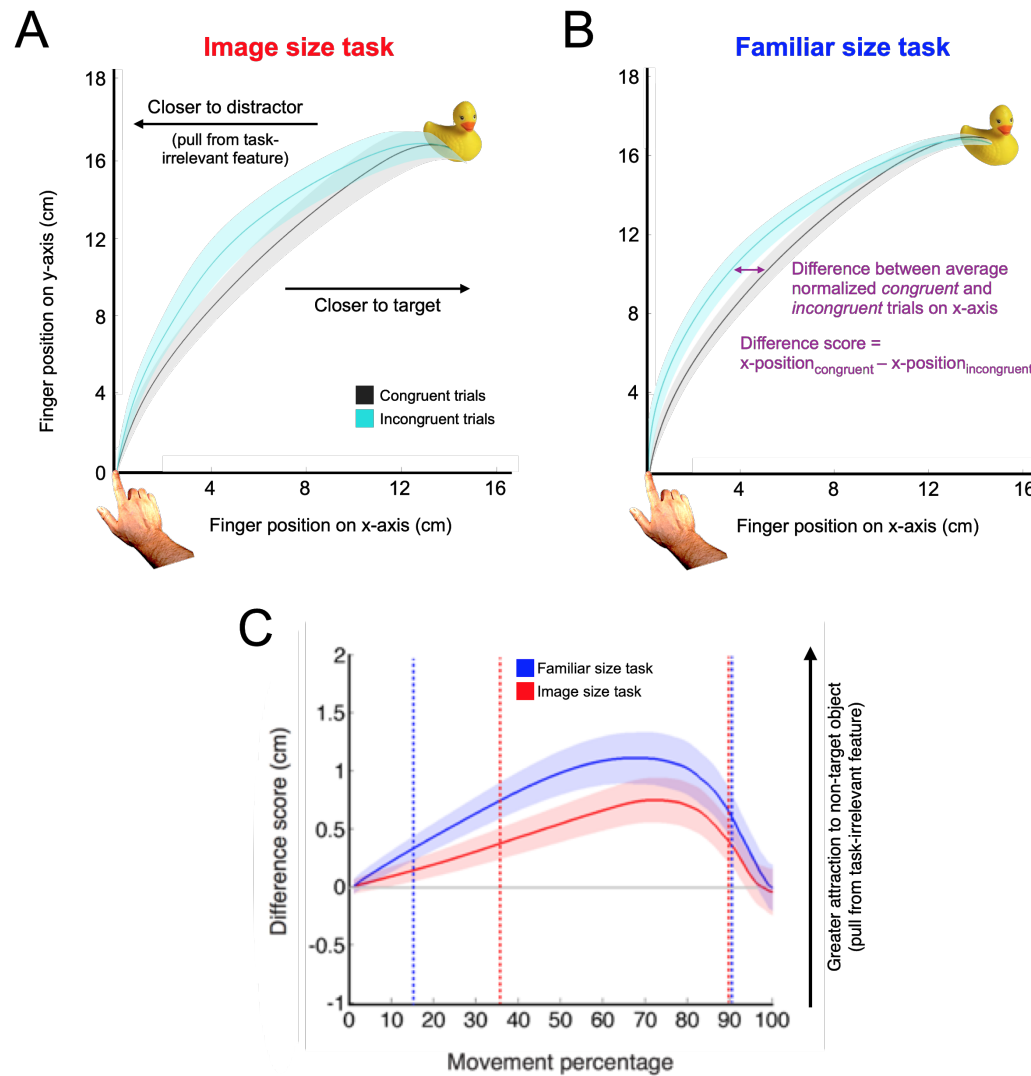
with no interaction between congruency and task  $F(1,13) = .880$ ,  $p = .365$ ,  $\eta_p^2 = .063$ . Similarly online movement execution was faster, with faster MTs observed for the image size task,  $F(1,13) = 11.96$ ,  $p = .004$ ,  $\eta_p^2 = .479$ , with no interaction between congruency and task  $F(1,13) = .519$ ,  $p = .484$ ,  $\eta_p^2 = .038$ .

Overall then, participants took longer to plan and execute their movements in the familiar size task compared to the image size task. This result, perhaps, is related to the fact that familiar size is characterized as *high level* and image size is characterized as *low level* meaning that evaluating familiar size requires more cognitive processing than evaluating image size. For instance, according to classical models of object processing, once the visual system extracts feature information such as image size, curvature, and depth, basic-level object recognition precedes accessing knowledge about that object such as its familiar size (Rosch et al., 1976; Collins & Quillian, 1969; Jolicoeur, Gluck, & Kosslyn, 1984). This effect though was not seen in accuracy or maximum curvature, suggesting that the key difference between the two tasks is temporal in nature.

#### *Time course of movement modulation in the image and familiar size tasks*

Our analysis of hand movement trajectory revealed key differences in maximum curvature between congruent and incongruent trials, in both the image and familiar size tasks (Figure 2D). In addition to the spatial measure of *maximum curvature*, which refers to a discrete point in a hand movement, we can additionally examine the impact of incongruency over the course of our continuous movement data. Comparing hand positions over the course of the movement for congruent and incongruent trials in the image and familiar size tasks provides insight into *when* image and familiar size interact in addition to previous measures of *how much*

they interact. As mentioned previously, movements were normalized to space with respect to reach distance from the starting point to the target (see, *Reach trajectory normalization in Methods* for details). Space-based normalization has the advantage of minimizing a potential confound from MT differences across trials. Thus, in evaluating the evolution of conflict between image and familiar size, we discuss the percentage into the course of the movement at which differences occur (Gallivan & Chapman, 2014).



**Figure 3.** Average normalized trajectory of hand movements in congruent (black line) and incongruent (cyan line) trials over the course of accurate reach-to-touch movements across all

1 participants. **A.** Image size task trials, where familiar size is the task-irrelevant interfering  
2 feature. **B.** Familiar size task trials, where image size is the task-irrelevant interfering feature.  
3 Movements to the left and right targets are collapsed across the midline, such that positions  
4 farther to the right are closer to the target, and farther to the left are closer to the midline and the  
5 incorrect object. Ribbons represent 95% confidence intervals. Hand and target are not to scale.  
6 **C.** Difference scores between congruent and incongruent trials representing the magnitude of  
7 interference by incongruency, in the image and familiar size tasks. Greater positive values  
8 represent greater attraction to the alternative (incorrect) object in incongruent trials, i.e., the  
9 strength of interference from the task-irrelevant feature. Scores in the familiar size task (where  
10 interference is from image size) rise significantly above zero earlier and peak earlier than the  
11 scores for the image size task (where interference is from familiar size). The ribbons represent  
12 95% confidence intervals.

13       Figure 3 shows average normalized hand movements in real space for the image size task  
14 where familiar size is the interfering feature (Figure 3A), and the familiar size task where image  
15 size is the interfering feature (Figure 3B). For both tasks, trials where image and familiar size  
16 were congruent are shown in black, and trials where image and familiar size were incongruent  
17 are shown in cyan. Though trials were counterbalanced such that reaches were performed with  
18 equal frequency to the left and right sides of the screen, here we have collapsed all trials along  
19 the midline such that greater values on the x-axis represent hand positions closer to the correct  
20 target object and smaller values on the x-axis represent hand positions closer both to the starting  
21 position and to the incorrect non-target object. In both tasks, we observed that the movements  
22 were relatively direct to the target in the congruent trials (black lines). In contrast, movements  
23 were more pulled by the task-irrelevant feature and therefore curved toward the distractor object

1 in the incongruent trials (cyan lines), leading to significant deviations between movement paths  
2 for congruent and incongruent trials in both tasks.

3 In order to analyze the deviation between congruent and incongruent trials over time, and  
4 particularly to compare the deviations in the image size with the deviations in the familiar size  
5 tasks over time, we calculated a difference score between normalized congruent and incongruent  
6 trials along the x-dimension in each task, as shown in Figure 3C. Our goal in calculating a  
7 difference score was to isolate and quantify the interference from the task-irrelevant feature.  
8 Given that the correct target (defined by the task-relevant feature) and the incorrect distractor  
9 object (defined by the task-irrelevant feature) were separated in space on the screen only in the x-  
10 dimension, we analyzed this dimension exclusively to most precisely examine this interference.

11 Here the difference between incongruent and congruent trials is depicted in red for the  
12 image size task and in blue for the familiar size task. In the familiar size task (blue solid line),  
13 where image size is the competing feature, difference scores are significantly greater than zero  
14 between 15% - 91% (two vertical blue dotted lines) of the movement, indicating that the average  
15 paths of the congruent and incongruent trials were significantly different in this span. In contrast,  
16 in the image size task (red solid line), where familiar size is the competing feature, the congruent  
17 and incongruent trajectories deviated later than in the familiar size task, approximately between  
18 36% - 90% into the movement (two vertical red dotted lines). Thus, congruency becomes a  
19 factor in the familiar size task after a smaller portion of the movement; or said another way,  
20 image size influences movements at an earlier proportion than familiar size does (Bennett, 2007).  
21 There is no significant difference in the points at which the difference scores return to zero,  
22 indicating that congruent and incongruent trials converge at similar points of the movement for  
23 the image and familiar size tasks.

Taken together, these results suggest that the image sizes of objects impact movement to a greater degree than familiar size does, as demonstrated by a greater effect of incongruency in the familiar size compared to the image size task. Not only is the magnitude greater, but the impact of image size is observed earlier in the course of the decision-making process than familiar size is. This suggests that image size may be processed more robustly overall than familiar size, though once again both are automatic and robust enough to interfere with the other.

## Summary

The goal of Experiment 1 was to examine if, and how, the processing of *image size* and *familiar size* influences goal-directed pointing movements when one type of object size is task-irrelevant. Overall, we demonstrated that both the image and familiar sizes of real-world objects play a role in planning and generating goal-directed action. Specifically, when we manipulated the congruency of targets' image and familiar sizes, each of these two aspects of object perception interfered with the other. This suggests that both image size and familiar real-world size are aspects of object perception and identification that occur automatically, even when task-irrelevant, and furthermore are not independent from action. Despite the bidirectional interference, however, image size may be more robustly processed and represented, and interfere with familiar size judgments more than familiar size interferes with it (e.g., Figure 3C). This observation was investigated in more detail in Experiment 2, along with questions regarding absolute and relative familiar size differences and their relationship with action.

## **Experiment 2: Effects of familiar size difference magnitude on action**

In Experiment 1, like in much of the existing literature, we categorized the familiar size of the two objects presented concurrently by which of the two was “larger” and which was “smaller,” while largely ignoring the magnitude of this relative size difference. (Chao & Martin, 2000; Downing, Chan, Peelen, Dodds, & Kanwisher, 2006; Kanwisher, 2001; Macuga & Papailiou, 2012; Wang & MacKenzie, 1999). Thus, it is still unknown whether it only matters that one object’s familiar size is larger than the other, or whether *how much* larger it is has an impact as well. For example, does comparing two objects where one is only slightly larger than the other (e.g., a peanut and a paperclip) have the same effect as comparing two objects that are vastly different sizes (e.g., a bathtub and a paperclip)? In Experiment 2, to examine whether differing degrees of real-world size differences impact goal-directed pointing, we expanded our set of real-world objects to include a wider range of familiar sizes, and systematically manipulated the magnitude of the difference in familiar sizes between the paired targets. We expected that if familiar size is treated as a graded spectrum, varying the magnitudes of familiar size differences may lead to corresponding graded effects on action, and on the interference between image and familiar size.

## **Methods**

### *Participants*

Twelve new right-handed participants (10 female, mean age 21.6) who did not participate in Experiment 1, with normal color vision and normal or corrected-to-normal visual acuity completed both visually guided pointing tasks: familiar size judgment and image size judgment. Of these, eleven (10 female, mean age 21.7) additionally performed a



familiar size rating task to validate our relative familiar size category manipulation. Participants provided their informed consent and were compensated monetarily (\$10 per hour) or with course credit for their participation. The experimental protocol was approved by the Brown University Institutional Review Board in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans.

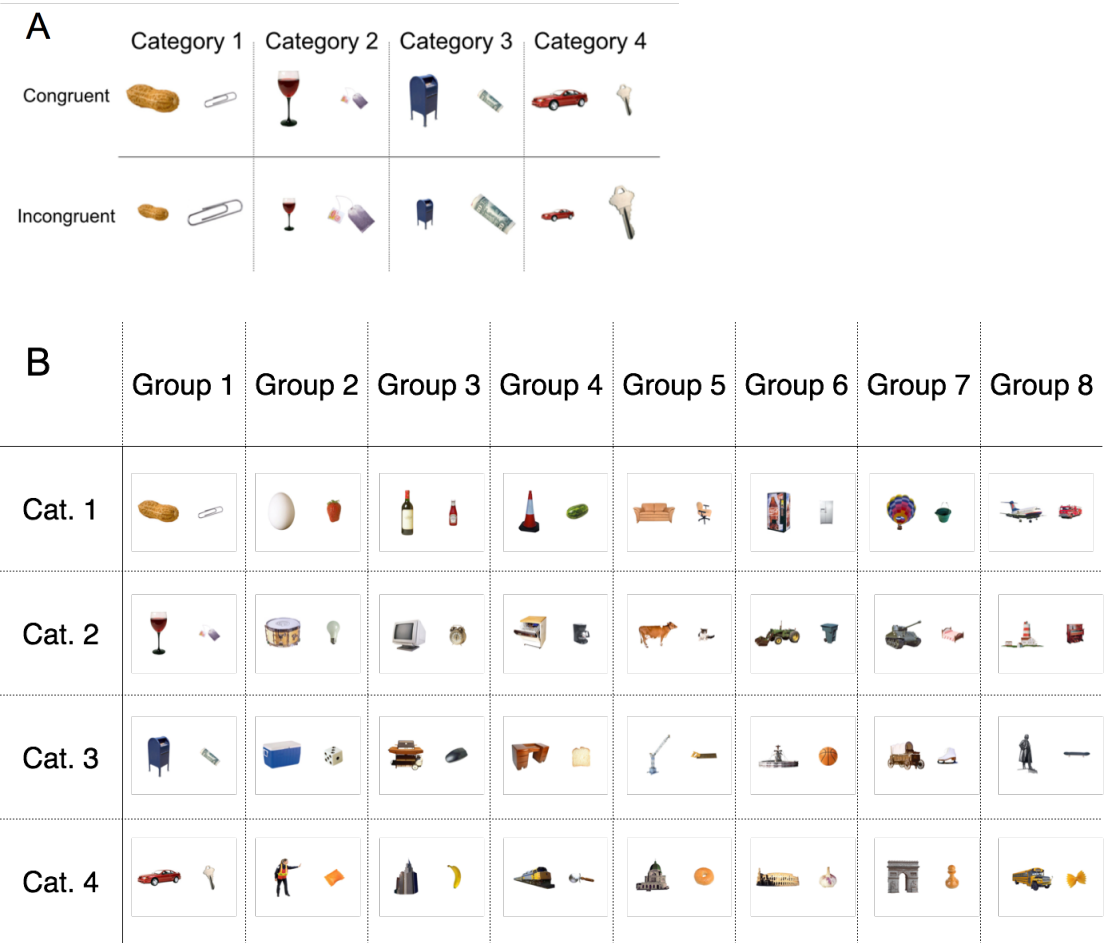
### *Apparatus*

The same apparatus was used as in Exp. 1.

### *Stimuli*

Experiment 2 stimuli were similar to those used in Experiment 1. We created image pairs from individual objects in the Konkle et al. “Object Size Range” database (Konkle & Oliva, 2011). Images in this database are divided into eight groups based on their familiar real-world sizes. These groups were independently defined and validated by participants (Konkle & Oliva, 2011). Using these eight familiar size groups, we defined four categories of paired images ranging from small familiar size differences to large familiar size differences. Objects in Category 1 were selected from within the same familiar size group (e.g. a peanut and a paperclip [both group 1], or a space shuttle and an airplane [both group 8]), Category 2 were two groups apart (e.g. a wineglass [group 3] and a teabag [group 1]), Category 3 were four groups apart (e.g. a cooler [group 5] and a die [group 1]), and Category 4 were six groups apart (e.g. a car [group 7] and a key [group 1]). All eight groups defined by familiar size by Konkle et al. (2011) were represented in all of our four categories defined by relative familiar size difference between pairs. Thus, in this four category by eight group manipulation there was a total of 32 image pairs.

Examples of pairs from each category, in their congruent and incongruent configurations, are shown in Figure 4A. All image pairs can be seen in Figure 4B.



**Figure 4.** Representative examples of familiar size categories. **A.** Representative examples of each of the four relative familiar size categories for object pairs, in both congruent and incongruent conformations. Categories 1-4 have progressively greater differences in familiar size magnitude. **B.** Every image pair used in its congruent conformation, arranged by category and group.

## Procedure

The Exp. 2 procedure was the same as in Exp. 1. Our four relative familiar size category manipulation based on the established image dataset was validated by participants' ratings of how different the familiar sizes of the objects in each pair were on a 0-10 scale representing no difference to extreme size differences. All image pairs from the pointing tasks were presented with a rating bar at the bottom of the screen representing a continuum from 0 to 10. Participants were instructed to use the mouse to click anywhere within the bar rating how large the relative familiar size difference was for the object pair shown (e.g., "How different are the sizes of a rubber duck and a boat in the real world?"). As in Experiment 1, object pairs were presented with relative image size and familiar size either congruent or incongruent. This task took place after both the image and familiar size judgment tasks in order not to bias performance.

## Data analysis

All data analysis was identical to Experiment 1, with the following exceptions. The measure *rating* was defined as the average value given to each object pair when participants were instructed to rate the pairs based on their relative *familiar size* differences. To analyze the effects of task and congruency in movement accuracy, initiation latency, movement time, and maximum curvature, and to compare between the two tasks as in Exp. 1, we performed a series of 2x4x2 ANOVAs with factors of congruency (congruent or incongruent), category (1-4), and task (image size vs. familiar size judgment).

## Results

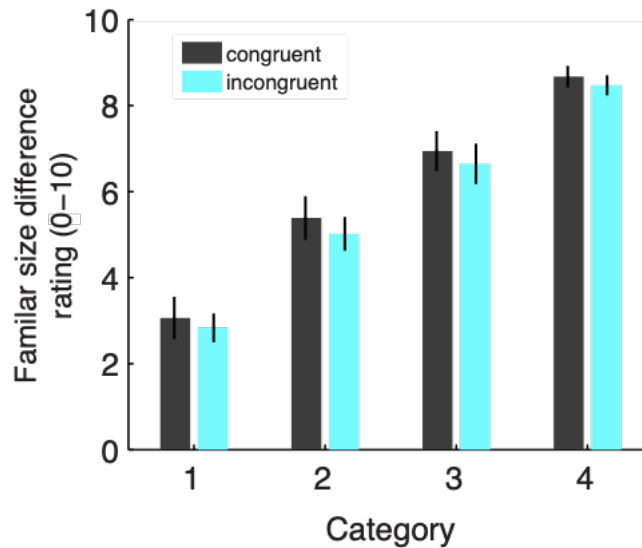
In total, 10.75% ( $\pm 1.02\%$ , s.e) of trials were excluded from data analysis from the image

size task and 10.93% ( $\pm 1.27\%$ ) from the familiar size task due to technical issues (e.g. sampling drop). Accuracy was 89.25 % ( $\pm 1.02\%$ ) for the image size task and 88.09 ( $\pm 1.27\%$ ) for the familiar size task. All data analyses were restricted to correct trials.

#### *Ratings*

To validate the four categories that we designed based on relative familiar size difference magnitude, we asked participants to rate this difference for each pair of objects. Additionally, we presented each pair in both its congruent and incongruent conditions, in order to evaluate the effect of congruency on subjective ratings of familiar size difference.

Figure 5 shows the average ratings assigned by participants to the image pairs in each of the four relative familiar size difference categories (1-4), when the relative familiar and image sizes were congruent (black) and incongruent (cyan). Participant ratings validate our manipulation of creating image pairs that fit these four categories—Figure 6 shows a clear trend of average rating increasing in a stepwise manner across categories 1-4 (small to large familiar size difference). This was supported by a 2x4 repeated measures ANOVA with factors of congruency (congruent vs. incongruent) and category (1-4), which revealed a main effect of category,  $F(3,30) = 55.80, p < .001, \eta_p^2 = .848$ .

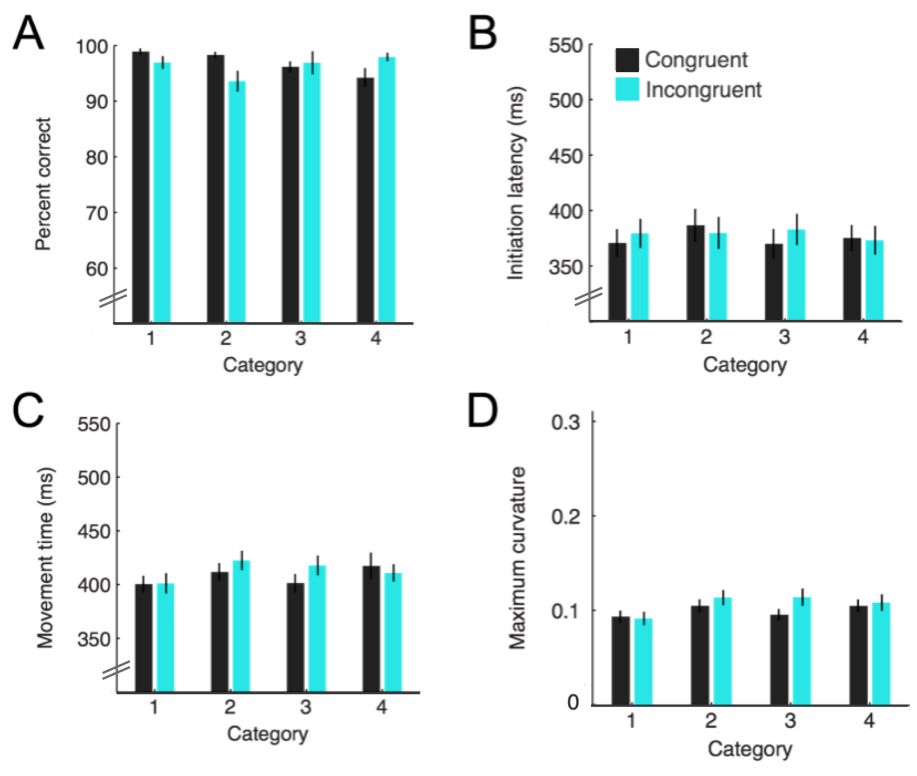


**Figure 5.** Average ratings of relative familiar size differences for image pairs in each relative image size category, when paired objects' image and familiar sizes are congruent and incongruent. Ratings supported our category definitions, with higher ratings given to categories designed to show larger familiar size differences. There was no significant difference in rating based on congruency. All error bars represent between-subjects standard error.

Figure 5 appears to show a trend such that ratings of the familiar size differences were higher on average in the congruent compared to incongruent condition. Such an effect would suggest that participants attended to the objects' image sizes despite task-irrelevance, and that congruent *image size* differences magnified the perceived difference in familiar size. However, there was no significant effect of congruency,  $F(1,10) = .874$ ,  $p = .372$ ,  $\eta_p^2 = .080$ , and no interaction between congruency and category,  $F(3,30) = .705$ ,  $p = .557$ ,  $\eta_p^2 = .066$ . This could be due to the absence of a meaningful effect, or insufficient power. This latter possibility may reflect the relative weakness of a perceptual button-press task compared to the more robust action measures employed in the main task (which show consistent effects of congruency) in accord with past research (Finkeiner et al., 2008).

## Image size task

**Accuracy.** As shown in Figures 6A, there were no clear trends in accuracy across conditions in the image size task. There was no main effect of congruency (*congruent* vs. *incongruent*),  $F(1,11) = .223$ ,  $p = .648$ ,  $\eta_p^2 = .024$ , or category (1-4),  $F(3,33) = 1.42$ ,  $p = .259$ ,  $\eta_p^2 = .136$ . There was however a significant interaction between the two such that the strength of the congruency effect was different in different categories,  $F(3,33) = 6.62$ ,  $p = .002$ ,  $\eta_p^2 = .424$ . This difference does not appear to follow a systematic trend across categories as familiar size difference magnitude varies, and thus it is impossible to make any claims about a meaningful effect of category on congruency or vice versa.



**Figure 6.** Results for congruent versus incongruent trials across relative familiar size conditions in the image size task of Experiment 2. All error bars represent between-subjects standard error.

**A. Accuracy.** Accuracy was consistent across all conditions. **B. Initiation latency.** IL was consistent across all conditions. **C. Movement time.** MT was slower overall for categories with larger familiar size differences. **D. Curvature.** Movements were more curved in the incongruent trials overall, and for categories with lower relative familiar size differences.

Initiation latency. In order to examine how congruency and the relative differences between real-world objects' familiar sizes affected movement preparation we again analyzed initiation latency (IL). Figure 6B shows that for the image size task, unlike in Exp. 1, IL was consistent across conditions. There was no significant difference based on congruency,  $F(1,11) = 1.27, p = .282, \eta_p^2 = .096$ , based on category,  $F(3,33) = 2.37, p = .087, \eta_p^2 = .165$ , and no interaction between the two,  $F(3,33) = 2.34, p < .085, \eta_p^2 = .166$ .

Movement time. Again, to examine how congruency and familiar size difference affected the online control of pointing we analyzed movement time (MT). Figure 6C shows a marginal effect of congruency on MT in the image size task,  $F(1,12) = 4.02, p = .068, \eta_p^2 = .251$ . Additionally we observed that with larger familiar size differences, MTs were slower overall,  $F(3,33) = 10.10, p < .001, \eta_p^2 = .457$ . As in the accuracy data, there was an interaction between congruency and category,  $F(3,33) = 5.36, p = .004, \eta_p^2 = .309$ . Again though, this effect does not appear to vary systematically with category; Figure 6C shows that categories 2 and 3 display greater differences between congruent and incongruent trials than 1 and 4, so we can draw no conclusions about the interaction in terms of MT.

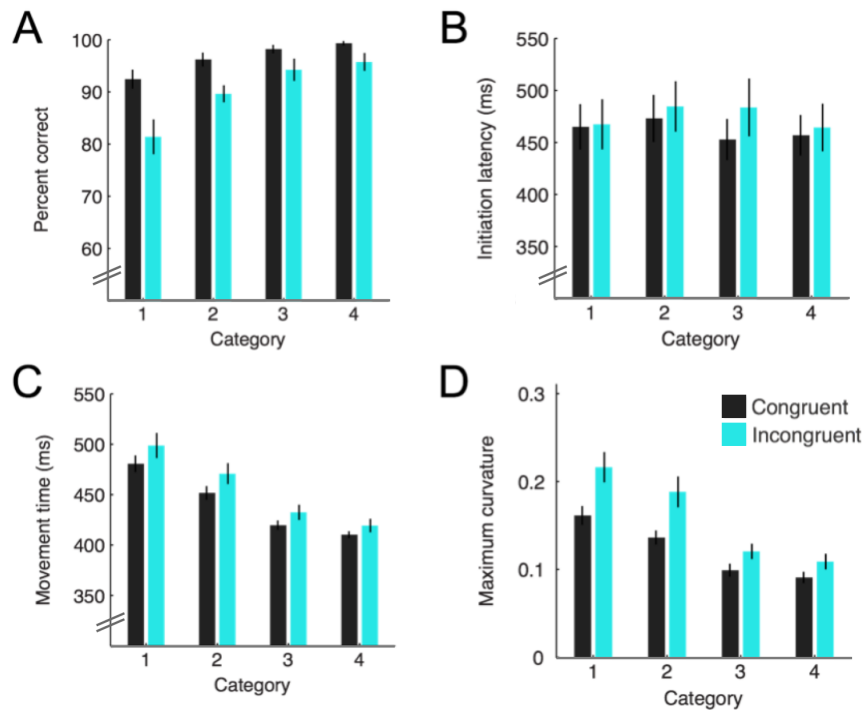
Maximum Curvature. In the image size task, maximum curvature results are consistent with those seen in movement time, and replicate the congruency effect seen in maximum curvature in Exp. 1 (Figure 2D, left). Maximum curvature was greater overall in the incongruent

conditions,  $F(1,11) = 6.72, p = .024, \eta_p^2 = .359$ , suggesting interference from familiar size when the two were incongruent (Figure 6D). Maximum curvature was also larger when familiar size differences were larger,  $F(3,33) = 7.68, p < .001, \eta_p^2 = .390$ . There was additionally an interaction between congruency and category,  $F(3,33) = 2.99, p = .044, \eta_p^2 = .199$ , indicating that congruency doesn't have the same impact across categories where familiar size differences vary.

### *Familiar size task*

Accuracy. In the familiar size task, we see several effects in accuracy not seen in the image size task. As seen in Figure 7A, there was a significant effect of congruency such that accuracy was lower when image and familiar size were incongruent,  $F(1,11) = 11.37, p = .008, \eta_p^2 = .558$ . There was additionally an effect of category such that accuracy was lower for the conditions with smaller familiar size differences,  $F(3,33) = 32.48, p < .001, \eta_p^2 = .783$ . This suggests that judgments of relative familiar size were more difficult when paired objects were closer in their familiar size. Here familiar size was the task-relevant feature while it was task-irrelevant in the image size task. Thus, it is not surprising to see an effect of category here but not in the image size task (Figure 6A). In Figure 7A there appears to be a trend such that the congruency had a smaller effect in larger categories, suggesting that larger familiar size differences were processed more robustly and more successfully resisted interference from image size, however this interaction was only marginally significant,  $F(3,33) = 2.61, p = .072, \eta_p^2 = .225$ .





**Figure 7.** Results for congruent versus incongruent trials across relative familiar size conditions in the familiar size task of Experiment 2. All error bars represent between-subjects standard error. **A.** Accuracy. Accuracy was higher for congruent compared to incongruent conditions, and for larger relative familiar size category. **B.** Initiation latency. IL was faster for congruent trials. **C.** Movement time. MT was faster for congruent trials, and for larger relative familiar size difference categories. **D.** Curvature. Movements were more curved in the incongruent trials overall, and for categories with lower relative familiar size differences.

Initiation latency. Figure 7B shows that, unlike in the image size task, ILs were marginally slower for incongruent compared to congruent trials,  $F(1,11) = 4.53, p = .055, \eta_p^2 = .274$ . Additionally, there was a significant effect of category,  $F(3,33) = 5.35, p = .004, \eta_p^2 = .308$ , and interaction between congruency and category,  $F(3,33) = 3.48, p = .026, \eta_p^2 = .225$ .

1 In the image size task, familiar size was the task-irrelevant interfering factor, whereas  
2 here in the familiar size task, image size is the interfering factor. Overall, familiar size failed to  
3 exert any influence on IL in the image size task (Figure 7B), but image size significantly  
4 interfered with the speed of movement initiation for the familiar size task (Figure 7B).  
5 Additionally, unlike the image size task, differences in familiar size impacted IL, as might be  
6 expected given that image size here was the task-relevant feature. However, these differences  
7 between category do not appear to follow any systematic trend, but rather be driven primarily by  
8 Category 3. Thus, no strong claims can be made based on this effect.

9  
10 Movement time. Figure 7C shows that in the familiar size task, movements were faster  
11 overall when image and familiar size were congruent,  $F(1,11) = 12.74, p = .004, \eta_p^2 = .515$ .  
12 Additionally, MTs were significantly faster when familiar size differences were larger (category  
13 1-4),  $F(3,33) = 78.06, p < .001, \eta_p^2 = .867$ . There was no interaction between these factors,  
14  $F(3,33) = .216, p = .885, \eta_p^2 = .018$ . As with IL, it is not surprising that category, which is based  
15 on relative familiar size difference, would modulate the familiar size task more than the image  
16 size task. That said, a clear asymmetry exists between image size's influence on familiar size,  
17 and familiar size's influence on image size, as seen in the difference between congruency effects  
18 in Figure 6C and 7C.

19  
20 Maximum Curvature. As in the image size task, maximum curvature was again greater  
21 overall in the incongruent compared to congruent conditions,  $F(1,11) = 18.47, p < .001, \eta_p^2$   
22  $= .606$ , (Figure 7D). Here though, there was less maximum curvature overall when familiar size  
23 differences were smaller compared to larger (categories 1-4),  $F(3,33) = 68.95, p < .001, \eta_p^2$

1 = .852. Again, familiar size here is the task-relevant factor while it is the interfering factor in the  
2 image size task, leading to opposite effects of increasing familiar size difference in the two tasks.  
3 What is consistent across both is that with larger familiar size differences, there is a larger effect  
4 of familiar size.

5 Similarly, there was an interaction between these two factors such that as relative familiar  
6 size difference (category) increased, the incongruency effect decreased,  $F(3,33) = 40.96, p$   
7  $< .001, \eta_p^2 = .773$ . Again, this reflects a decreasing degree of interference from image size as  
8 relative familiar size differences increase, and is expectedly the reverse of the effect seen in  
9 image size.

#### 11 *Comparisons between image and familiar size tasks*

12 In addition to the above within-task measures, we again compared the image and familiar  
13 size tasks to each other. We found greater accuracy,  $F(1,11) = 6.48, p = .031, \eta_p^2 = .419$ , faster  
14 initiation latency,  $F(1,11) = 40.06, p < .001, \eta_p^2 = .770$ , faster movement time  $F(1,11) = 46.41, p$   
15  $< .001, \eta_p^2 = .795$ , and smaller maximum curvature,  $F(1,11) = 107.25, p < .001, \eta_p^2 = .899$  in  
16 the image than the familiar size task. Taken together, this suggests that the familiar size task may  
17 have been more difficult than the image size task, consistent with participants' subjective  
18 reports.

19 There were additionally task by congruency interactions in accuracy,  $F(1,11) = 7.20, p$   
20  $= .025, \eta_p^2 = .444$ , movement time,  $F(1,11) = 84.62, p < .001, \eta_p^2 = .876$ , maximum curvature,  
21  $F(1,11) = 10.81, p = .006, \eta_p^2 = .474$ , and a marginally significant effect in initiation latency,  
22  $F(1,11) = 4.37, p = .059, \eta_p^2 = .267$ . Together, these all point to image size interfering in the  
23 familiar size task more than familiar size did in the image size task.

We also observed interactions between category (i.e. the degree of familiar size difference between paired objects) and task for both movement time and maximum curvature, such that categories with larger familiar size differences led to shorter movement times,  $F(3,33) = 84.32, p < .001, \eta_p^2 = .89$ , and less curvature,  $F(3,33) = 85.12, p < .001, \eta_p^2 = .89$ , in the familiar size compared to the image size task. This effect is not surprising given that familiar size was task-relevant in the familiar size task, and task-irrelevant in the image size task.

Finally, for maximum curvature, there was a three-way interaction between task, congruency, and category,  $F(3,33) = 11.21, p = .007, \eta_p^2 = .51$ . Such an interaction is expected given both the individual main effects seen in the previous analysis, and the way our categories are defined based on magnitude of familiar size difference. In the image size task (Figure 6D), larger familiar size differences (category) led to more curved trajectories  $F(3,33) = 7.68, p < .001, \eta_p^2 = .390$ . However this effect seems to be driven entirely by the incongruent conditions—maximum curvature increases across the incongruent conditions based on category, while the congruent trials are stable across category, consistent with a category by congruency interaction,  $F(3,33) = 2.99, p = .044, \eta_p^2 = .199$ . This suggests that increasing the familiar size difference increases the interference of familiar size in performing the image size task. However, when familiar size and image size are congruent familiar size difference plays no role, or a negligible role compared to the perception of image size.

In the familiar size task, there was an overall decrease in maximum curvature as the familiar size difference (category) increased (Figure 7D),  $F(3,33) = 68.95, p < .001, \eta_p^2 = .852$ . This is the reverse of the effect seen in the image size task because here participants' decisions were made based on *familiar size* as opposed to *image size*, while category was defined by relative familiar size difference in both cases. Thus, the lower maximum curvature for the greater

familiar size differences may reflect lower task difficulty. Similarly, the interaction effect suggests that the effect of congruency is different across categories. Specifically, the congruency effect is greater in the smaller familiar size difference categories, showing that image size interferes with participants' ability to judge familiar size more when the familiar size difference is smaller (i.e., the familiar sizes of the paired objects are closer). This decrease in the effect of congruency across category was supported by a post-hoc linear trend analysis in which we found that difference scores ( $\text{curvature}_{\text{incongruent}} - \text{curvature}_{\text{congruent}}$ ) significantly decreased across category, as familiar size difference increased,  $r = -.32$ ,  $p = .028$ .

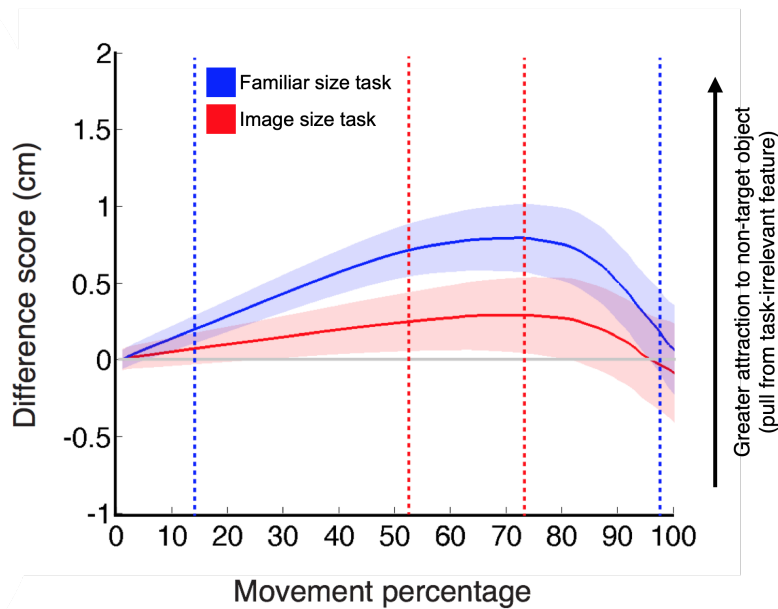
Taken together with converging evidence that image size is processed more robustly than familiar size, this suggests that image size is the greater influence on trajectory overall, but that familiar size exerts a larger influence the larger the magnitude of familiar size differences. This increased modulation of movement by familiar size leads both to greater interference from familiar size in the image size task, and greater resistance against interference from image size in the familiar size task.

#### *Time course of movement modulation in the image and familiar size tasks*

As in Experiment 1, we again calculated difference scores from the normalized congruent and incongruent trials in each task to examine the influence of image and familiar size over time (Figure 8). Specifically, we were interested in the effect of varying the magnitude of the familiar size difference between paired images on the time course of both image size's and familiar size's impact on movement.

We first performed the same time course analysis as in Exp. 1 by comparing the difference scores between congruent and incongruent trials for the image size task (solid red line)

and familiar size task (solid blue line). Figure 8 shows that the difference scores were significantly above zero 53% - 71% into the movement for the image size task (vertical red dotted lines) and 14% - 97% into the movement for the familiar size task (vertical blue dotted lines). This replicates our Exp. 1 findings that image size “came online” and influenced movements earlier in the course of movements than familiar size did.

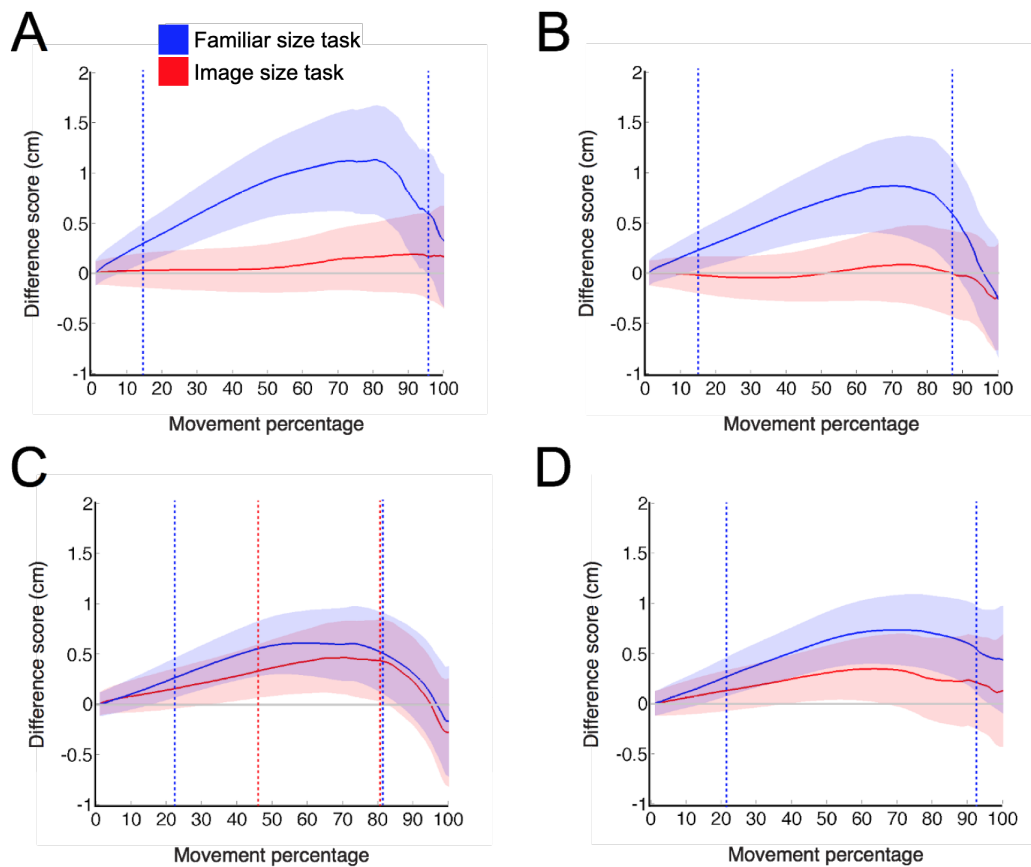


**Figure 8.** Difference scores between congruent and incongruent trials representing the magnitude of interference by incongruency, in the image and familiar size tasks. Scores in the familiar size task (blue), where image size is the interferer rise significantly above zero earlier than the scores for the image size task (red), where familiar size is the interferer. The ribbons represent 95% confidence intervals.

It is worth noting that the precise timing and the magnitude of interference in Exp. 2 did not completely replicate the results in Exp. 1. In the image size task specifically, familiar size interfered later and less robustly in Exp. 2 than in Exp. 1. This could reflect the fact that the

range of familiar sizes was much larger in Exp. 2 compared to Exp. 1, and thus there was a wider range of interference effects and more noise overall.

We additionally compared the influences of image and familiar size on the movements between categories. Figure 9 shows the time course of difference scores for each category, with the image size task presented in red and the familiar size task in blue: Figures 9A-D represent data from Category 1 to Category 4, which contained the smallest to the largest familiar size difference magnitudes, respectively.



**Figure 9.** Time course of different scores across categories in Experiment 2. Red and blue lines represent mean difference scores in the image size task and familiar size task, respectively. The

1 ribbons represent 95% confidence intervals. **A.** Category 1. **B.** Category 2. **C.** Category 3. **D.**  
2 Category 4.

3 For the image size task (red), where familiar size was the interfering factor, there was an  
4 overall main effect of congruency. However, only Category 3 independently displays a  
5 difference scores significantly above zero, 46% - 82% into the movements. Thus, familiar size  
6 interference in the known size task cannot be compared between all familiar size categories.

7 In the familiar size task (blue) where image size was the interfering factor, difference  
8 scores were significantly greater than zero for all categories, 15% - 96% into the movement for  
9 category 1, 16% - 88% for category 2, 23% - 81% for category 3, and 22% - 93% for category 4.  
10 Thus, there was no overall effect of category in terms of where in the movement interference  
11 from image size came online, indicating that the magnitude of the familiar size difference did not  
12 significantly impact how congruence effects unfolded.

13 Overall then, we replicate the time course effect seen in Exp. 1 (Figure 3C), with image  
14 size influencing trajectory earlier in the course of the movement than familiar size. However,  
15 when we separate out the categories defined by familiar size difference, we see no differences in  
16 the timing of a congruency effect in the familiar size task, and no reliable congruency effect in  
17 the image size task. Thus, the magnitude of relative familiar size difference (i.e. category)  
18 modulated movements in both the image and familiar size tasks in terms of *trajectory* (i.e.  
19 maximum curvature), but not in terms of *timing*. This suggests that differences in the familiar  
20 sizes of objects lead to corresponding differences in the degree of interference between image  
21 and familiar size, but not *when* the interference occurs. This result might also suggest that the  
22 conflict resolution process in our paradigm is sensitive during movements, and may even be  
23 dynamically tailored to this period.



## Summary

The goal of Experiment 2 was to examine whether differing degrees of real-world size difference impact goal-directed pointing to differing degrees, and to do this we expanded our set of real-world objects to include a wider range of familiar sizes and systematically manipulated the magnitude of the difference in familiar sizes between paired targets. Here, we again demonstrated bidirectional interference between the perception of *image size* and *familiar size* when the two are presented in an incongruent manner, though here we also found evidence of image size exerting greater influence on familiar size than vice versa. Furthermore, we observed several trends associated with greater magnitudes of familiar size difference between paired objects. Namely, in the familiar size judgment task, the larger the familiar size difference, the lower the maximum curvature of goal-directed pointing movements overall. This is consistent with greater differences in familiar size making familiar size judgments easier.

Additionally, the incongruence effect in the familiar size task diminished with increasing familiar size difference magnitude. This indicates that image size did interfere with familiar size judgment across categories, but did so less when the relative familiar size difference was greater. Similarly, in the image size judgment task, the interference from familiar size increased with increasing familiar size difference across the four categories. Overall our result is consistent with image size perception being the more automatic of the two, and exerting greater influence on the decision making process and on hand movements. However, the greater the difference between the familiar real-world sizes of the objects, the greater the effect that familiar size has, and the more it mitigates/interferes with image size's influence.

## General Discussion

Our perception of real-world objects involves conceptualization of how large or small they are in the real world—we expect a rubber duck to be physically smaller than a boat (though artists have taught us that this doesn't have to be the case) (Konkle & Oliva, 2011, 2012a, 2012b). In the current study we presented participants with pairs of real-world objects, and manipulated their relative *image* and *familiar* sizes. By using a visually-guided pointing paradigm and decision-making task based on either image size or familiar size, we were able to assess the relationship between these two aspects of object perception and their impacts on visually-guided action. Overall, we observed that the conflict between image and familiar sizes is resolved over the course of the movement and results in curved trajectories, suggesting early processing of each feature, and a later conflict resolution between the two.

### *Mechanisms of image and familiar size perception*

Object perception relies on a hierarchy of perceptual processes representing increasingly complex object features (Riesenhuber & Poggio, 1999). Image size is a low-level feature, and is represented essentially by the size of the stimulus on the retina—far earlier in the visual processing pathway than anything related to object identity. However, recent studies have demonstrated that familiar size can be represented and processed earlier than classical models of object processing assume (Rosch et al., 1976; Collins & Quillian, 1969; Jolicoeur, Gluck, & Kosslyn, 1984).

For instance, as early in the image processing pathway as V1, representations have been shown to reflect viewers' subjective perception of objects' sizes as opposed to their veridical image size. Sperandio, Chouinard, and Goodale (2012) found that, despite constant retinal image

size, V1 activity reflected perceived object size in a size constancy task that manipulated target viewing distance, suggesting that image size perception is influenced by other aspects of perception even at this early stage in processing.

Furthermore, the familiar sizes of real-world objects have been shown to have neural representations independent of image size. Konkle and Oliva (2012b) showed that familiar object size is represented in the occipitotemporal cortex (OT) much like object categories, with larger objects represented in medial OT and smaller objects in lateral OT. Thus, our observation that both image size and familiar size are processed automatically is consistent with known mechanisms underlying object perception. In addition, recent studies have shown that processing mid-level perceptual features can be sufficient to distinguish objects of different familiar sizes without real-world context (Long et al., 2016; Long et al., 2018).

Here, we report evidence of bidirectional interference that suggests interactions between image and familiar size mechanisms. Image size was consistently the more dominant of the two—it was processed earlier and more robustly, and participants anecdotally reported finding it more salient than familiar size. That said, we demonstrated in Experiment 2 that the robustness of familiar size’s influence on movements, and we infer therefore its representation strength, increased with increasing familiar size magnitude, and that this robustness mitigated image sizes’ effects when the two were in conflict. Essentially the more robust process—image size judgment—became *less* robust and more susceptible to influence from the task-irrelevant feature familiar size when the familiar size difference was greater in magnitude.

Similarly we observed that, in the familiar size task, task-irrelevant image size processing interfered *less* the greater the magnitude of the familiar size difference. Thus, in both tasks greater familiar size differences more strongly counteracted the influence of image size. This is

possibly due to increased salience of familiar size as a feature when the familiar sizes of the two paired real-world objects was more discrepant. Furthermore, the parametric nature of our manipulation in Experiment 2 and the resulting graded effects of familiar size suggest that familiar size may be represented in a continuous manner as opposed to the binary “large object” and “small object” areas found by Konkle & Oliva (2012a).

Overall, we present converging evidence that image size is processed earlier and more robustly than familiar size, but that increasing familiar size differences somewhat mitigate this effect, indicating dynamic interactions between image and familiar size perception, and action.

#### *Processing conflict in the familiar size Stroop task*

The “familiar size Stroop” task used here was inspired by the classic Stroop task in which identifying the color a word is printed in is impaired when the word is the name of an incongruent color (Konkle & Oliva, 2012a; Stroop, 1935). In order to accurately perform the Stroop task participants must suppress the automatic response produced by a *direct processing pathway* in order to respond to the relevant feature for their current goal, processed by an *indirect processing pathway* (Botvinick et al., 2001; Ridderinkhof, van der Molen, & Bashore, 1995; Shenhav, Botvinick, & Cohen, 2013; Van der Stigchel, van Koningsbruggen, Nijboer, List, & Rafal, 2012).

The prioritization of task-relevant over task-irrelevant features is a key aspect of cognitive control, and the Stroop task has been investigated extensively within the cognitive control literature (Memelink & Hommel, 2013). Resolving conflict in the Stroop task is thought to comprise three processes: monitoring conflict between the direct pathway and indirect pathway, adjusting the response threshold by inhibiting motor output, and recruiting top-down

1 processes in order to resolve conflict between the two pathways (Botvinick et al., 2001; Erb et  
2 al., 2016; Shenhav et al., 2013). Erb et al. (2016) investigated these individual processes in a  
3 classic Stroop task using a visually-guided pointing paradigm, and showed that they manifested  
4 in different aspects of visually-guided action. Specifically, movement initiation latency was  
5 shown to reflect the response threshold adjustment process while maximum curvature reflected  
6 the final conflict resolution between the direct and indirect pathways.

7 By using a similar visually-guided pointing paradigm and the familiar size Stroop task we  
8 are able to speculate about the processing of image size and familiar size, and the conflict  
9 between the two. When objects' image and familiar sizes were in conflict participants were  
10 slower to initiate movements, and their movements were more curved toward the incorrect  
11 response. We propose that the former effect results from motor output being suppressed to allow  
12 for more processing time, an early conflict resolution process. The latter effect suggests that  
13 representations of both alternatives persist even after movements are initiated (Erb et al., 2016).  
14 Thus, conflict resolution processes begin early, but conflict is not resolved until later. This also  
15 demonstrates that both image size and familiar size, as well as their conflict, influence both  
16 movement preparation and movement execution, with a larger effect from image size.

17 Again though, despite the fact that image size influenced familiar size more than familiar  
18 size did image size, we did observe bidirectional interference. While the classical Stroop effect is  
19 largely reported in terms of the unidirectional influence of reading a word on reporting a color, a  
20 "reverse Stroop effect" has also been observed in which participants' ability to read a color name  
21 is impaired by incongruent text color (MacLeod, 1991). However, the effect of the written word  
22 on printed color identification is more pronounced than the impact of ink color on reading the

1 written word. Thus, interference in the classical Stroop effect is bidirectional but asymmetrical,  
2 like our reported effects in the familiar size Stroop.

3       Based on these results, processing image size in the familiar size Stroop task parallels  
4 reading the word in the classical Stroop task, and familiar size processing parallels reporting the  
5 text color. From this we speculate that the perception of an object's *image size* is a direct  
6 pathway process and the perception of *familiar size*, an indirect pathway process, again  
7 consistent with participants' subjective reports of their relative ease.

8       These observations have important implications for target selection, action control, and  
9 perception/action integration more broadly. There are a number of reasons image size processing  
10 could be a direct pathway process in this action-based Stroop paradigm. First, image size is  
11 arguably more relevant for the guidance of action than familiar size is—we can process the size  
12 of an object in front of us and interact with it even if it is not an object we are familiar with, and  
13 when familiar objects' image and familiar sizes are incongruent their current image size is far  
14 more important for action than their typical familiar size.

15       In the classic Stroop task the direct pathway process is reading a word while the indirect  
16 process is naming the color of the ink, again somewhat counterintuitive given that reading is a  
17 higher order process than color perception. However, reading proceeds more automatically due  
18 to the relative frequency of performing this learned behavior. Similarly, attending to image size  
19 in isolation may be performed more frequently in daily life than processing familiar size in  
20 isolation. These two explanations are not mutually exclusive. It is likely that the automaticity of  
21 image size processing and its influence on action is the result of its frequency and utility in daily  
22 life.

1           That said, our results show that familiar size is processed automatically as well. This  
2 learned behavior is more relevant to visual perception than it is to visually guided action in a  
3 classical perception/action dissociation framework. However, as we have shown, perception and  
4 action are not truly distinct and perception does influence action. Thus, if image size influences  
5 action more readily and familiar size influences perception more readily, this doesn't mean that  
6 familiar size doesn't influence action, just that the influence emerges later as in other  
7 perception/action integration processes.

8           It is additionally worth noting that just as the Stroop effect can be replicated with features  
9 other than color names, and can inform our understanding of *direct* and *indirect* processing  
10 pathways, we believe that the action-based Stroop effect seen here is not limited to incongruency  
11 between image and familiar size. While object size has a significant impact on reach-to-point as  
12 well as other types of hand movements, the conflict revealed in these experiments is as much  
13 about cognitive control and conflict-resolution as about visual perception and action control.  
14 Thus, we hypothesize that these results would be replicated with other forms of perceptual  
15 incongruency or conflict tasks, and the same asymmetry observed as long as one feature  
16 represents a *direct* and the other an *indirect* pathway.

## 18 **Conclusion**

19           By examining the modulation of goal-directed hand movements by the image sizes and  
20 familiar sizes of real-world objects, the present study can contribute to a more complete picture  
21 of how objects are perceived and identified, how hand movements are guided, and how these  
22 processes interact. Comparing the impact of image size and familiar size, image size appears to  
23 exert more impact, and be processed faster than familiar size is. That said, the more complex and

difficult perceptual process of judging the familiar sizes of objects, requiring higher level perceptual identification and the recruitment of prior experiences and memory, also occurs automatically and robustly enough to interfere with judgments of image size. Critically, both image and familiar size are processed even when task-irrelevant.

The strength of this bidirectional influence of image size on familiar size judgment, and familiar size on image size judgment is not absolute, however. Experiment 2 demonstrated that when the familiar size difference between paired objects was larger, the effects of familiar size on action became more robust. This had two effects. First, choosing the larger or smaller object based on its familiar size in the real world became easier the larger the difference between the two, despite the fact that a difference was always readily apparent. Secondly, and more importantly, familiar size interfered with the effects of image size—image size judgments in the image size task, and image size *interference* in the incongruent trials of the familiar size task—more strongly with greater relative familiar size differences, suggesting greater salience and representation robustness.

Overall, the present study provides evidence that even high-level aspects of visual perception—the identification of real-world objects and the integration of prior knowledge regarding their sizes in the real world—interact with visually-guided action automatically and systematically. This points to a far more integrated view of perception and action than classically hypothesized. However, the current study alone is insufficient to explain these interactions on a mechanistic level. Further investigations are needed to identify the mechanisms responsible for the perception of object size (image size, familiar size, and other aspects), decision-making, visually-guided action, and their relationships.



1    **Acknowledgments**

2    This project is supported by NSF BCS 1849169 to JHS. CMG is supported by NSF Graduate

3    Research Fellowship.

## References

- Binsted, G., & Elliott, D. (1999). The Müller–Lyer illusion as a perturbation to the saccadic system. *Human Movement Science*, 18(1), 103-117.
- Bennett, D. J. (2007). Does an Estimate of Environmental Size Precede Size Scaling on a Form-Comparison Task? *Perception*, 36(3), 375–390.
- Bernardis, P., Knox, P., & Bruno, N. (2005). How does action resist visual illusion? Uncorrected oculomotor information does not account for accurate pointing in peripersonal space. *Experimental Brain Research*, 162(2), 133-144.
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., Cohen, J. D., & Botvinick, M. (2001). Evaluating the demand for control: Anterior cingulate cortex and conflict monitoring. *Psychological Review*, 108(3), 624–52.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10, 433–436.
- Chao, L. L., & Martin, A. (2000). Representation of manipulable man-made objects in the dorsal stream. *NeuroImage*, 12(4), 478–484.
- Cohen, J. (1973). Eta-Squared and partial eta-squared in fixed factor ANOVA designs. *Educational and Psychological Measurement*, 33, 107–112.
- Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of verbal learning and verbal behavior*, 8(2), 240-247.
- de Grave, D. D., Franz, V. H., & Gegenfurtner, K. R. (2006). The influence of the Brentano illusion on eye and hand movements. *Journal of Vision*, 6(7), 5-5.
- Dotan, D., Pinheiro-Chagas, P., Al Roumi, F., & Dehaene, S. (2019). Track it to crack it: Dissecting processing stages with finger tracking. *Trends in Cognitive Sciences*, 23(12), 1058-1070.

- Downing, P. E., Chan, A. W.-Y., Peelen, M. V., Dodds, C. M., & Kanwisher, N. (2006). Domain Specificity in Visual Cortex. *Cerebral Cortex*, 16(10), 1453–1461.
- Erb, C. D. (2018). The developing mind in action: measuring manual dynamics in childhood. *Journal of Cognition and Development*, 19(3), 233-247.
- Erb, C. D., Moher, J., Sobel, D. M., & Song, J.-H. (2016). Reach tracking reveals dissociable processes underlying cognitive control. *Cognition*, 152, 114–126.
- Finkbeiner, M., Song, J.-H., Nakayama, K., & Caramazza, A. (2008). Engaging the motor system with masked orthographic primes: a Kinematic analysis. *Visual Cognition*, 16, 11–22.
- Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47(6), 381–391.
- Franz, V. H. (2001). Action does not resist visual illusions. *Trends in cognitive sciences*, 5(11), 457-459.
- Gamble, C. M., & Song, J.-H. (2017). Dynamic modulation of illusory and physical target size on separate and coordinated eye and hand movements. *Journal of Vision*, 17(3), 23.
- Gallivan, J. P., & Chapman, C. S. (2014). Three-dimensional reach trajectories as a probe of real-time decision-making between multiple competing targets. *Frontiers in neuroscience*, 8, 215.
- Glover, S., & Dixon, P. (2002). Dynamic effects of the Ebbinghaus illusion in grasping: Support for a planning/control model of action. *Perception & Psychophysics*, 64(2), 266-278.
- Itti, L., & Koch, C. (2001). Feature combination strategies for saliency-based visual attention systems. *Journal of Electronic imaging*, 10(1), 161-169.

- 1 Jeannerod, M. (1984). The Timing of Natural Prehension Movements. *Journal of Motor*  
2 *Behavior*, 16(3), 235–254.
- 3 Jolicoeur, P., Gluck, M. A., & Kosslyn, S. M. (1984). Pictures and names: Making the  
4 connection. *Cognitive psychology*, 16(2), 243-275.
- 5 Kanwisher, N. (2001). Neural events and perceptual awareness. *Cognition*, 79(1–2), 89–113.
- 6 Knol, H., Huys, R., Sarrazin, J. C., Spiegel, A., & Jirsa, V. K. (2017). Ebbinghaus figures that  
7 deceive the eye do not necessarily deceive the hand. *Scientific reports*, 7(1), 1-17.
- 8 Konkle, T., & Oliva, A. (2011). Canonical visual size for real-world objects. *Journal of*  
9 *Experimental Psychology: Human Perception and Performance*, 37(1), 23–37.
- 10 Konkle, T., & Oliva, A. (2012a). A familiar-size Stroop effect: Real-world size is an automatic  
11 property of object representation. *Journal of Experimental Psychology: Human*  
12 *Perception and Performance*, 38(3), 561–569.
- 13 Konkle, T., & Oliva, A. (2012b). A Real-World Size Organization of Object Responses in  
14 Occipitotemporal Cortex. *Neuron*, 74(6), 1114–1124.
- 15 Kosslyn, S. M. (1978). Measuring the visual angle of the mind’s eye. *Cognitive Psychology*,  
16 10(3), 356–389.
- 17 Long, B., Konkle, T., Cohen, M. A., & Alvarez, G. A. (2016). Mid-level perceptual features  
18 distinguish objects of different real-world sizes. *Journal of Experimental Psychology:*  
19 *General*, 145(1), 95.
- 20 Long, B., Yu, C. P., & Konkle, T. (2018). Mid-level visual features underlie the high-level  
21 categorical organization of the ventral stream. *Proceedings of the National Academy of*  
22 *Sciences*, 115(38), E9015-E9024.

1 MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review.  
2 *Psychological Bulletin*, 109(2), 163–203.

3 Macuga, K. L., & Papailiou, A. P. (2012). Motor imagery of tool use: relationship to actual use  
4 and adherence to Fitts' law across tasks. *Experimental Brain Research*, 218(2), 169–79.

5 McKinstry, C., Dale, R., & Spivey, M. J. (2008). Action Dynamics Reveal Parallel Competition  
6 in Decision Making. *Psychological Science*, 19(1), 22–24.

7 Memelink, J., & Hommel, B. (2013). Intentional weighting: a basic principle in cognitive  
8 control. *Psychological Research*, 77(3), 249–259.

9 Milner, A. D., & Goodale, M. A. (2008). Two visual systems re-  
10 viewed. *Neuropsychologia*, 46(3), 774–785.

11 Moher, J., Anderson, B. A., & Song, J.-H. (2015). Dissociable Effects of Salience on Attention  
12 and Goal-Directed Action. *Current Biology*, 25(15), 2040–2046.

13 Moher, J., & Song, J.-H. (2013). Context-dependent sequential effects of target selection for  
14 action. *Journal of Vision*, 13(8), 10–10.

15 Moher, J., & Song, J.-H. (2014). Target selection bias transfers across different response actions.  
16 *Journal of Experimental Psychology: Human Perception and Performance*, 40(3), 1117–  
17 1130.

18 Moher, J., & Song, J. H. (2019). A comparison of simple movement behaviors across three  
19 different devices. *Attention, Perception, & Psychophysics*, 81(7), 2558–2569.

20 Murray, S. O., Boyaci, H., & Kersten, D. (2006). The representation of perceived angular size in  
21 human primary visual cortex. *Nature Neuroscience*, 9(3), 429–434.

- Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, 2011, 1–9.
- Ridderinkhof, K. R., van der Molen, M. W., & Bashore, T. R. (1995). Limits on the application of additive factors logic: Violations of stage robustness suggest a dual-process architecture to explain flanker effects on target processing. *Acta Psychologica*, 90(1–3), 29–48.
- Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2(11), 1019–1025.
- 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.
- Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The Expected Value of Control: An Integrative Theory of Anterior Cingulate Cortex Function. *Neuron*, 79(2), 217–240.
- Song, J.-H. (2017). Abandoning and modifying one action plan for alternatives. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 372(1718), 20160195.
- Song, J.-H., & Nakayama, K. (2009). Hidden cognitive states revealed in choice reaching tasks. *Trends in Cognitive Sciences*, 13(8), 360–366.
- Sperandio, I., Chouinard, P. A., & Goodale, M. A. (2012). Retinotopic activity in V1 reflects the perceived and not the retinal size of an afterimage. *Nature Neuroscience*, 15(4), 540–542.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18(6), 643–662.

1 Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta psychologica*,  
2 135(2), 77-99.

3 Van der Stigchel, S., van Koningsbruggen, M., Nijboer, T. C. W., List, A., & Rafal, R. D.  
4 (2012). The role of the frontal eye fields in the oculomotor inhibition of reflexive  
5 saccades: evidence from lesion patients. *Neuropsychologia*, 50(1), 198–203.

6 Van Donkelaar, P. (1999). Pointing movements are affected by size-contrast illusions.  
7 *Experimental Brain Research*, 125(4), 517–520.

8 Wang, Y., & MacKenzie, C. L. (1999). Object manipulation in virtual environments. In  
9 *Proceedings of the SIGCHI conference on Human factors in computing systems the CHI*  
10 *is the limit - CHI '99* (pp. 48–55). New York, New York, USA: ACM Press.