

This article is published in Nature Human Behavior:

<https://www.nature.com/articles/s41562-018-0485-2>

Title: **Attention Scales According to Inferred Real-World Object Size**

Authors: Andrew J. Collegio^{1*}, Joseph C. Nah¹, Paul Scotti^{1,2}, and Sarah Shomstein¹

Author affiliation: ¹Department of Psychology, The George Washington University;
²Department of Psychology, The Ohio State University

Corresponding author:

Andrew J. Collegio
Department of Psychology
George Washington University
2125 G St. NW
Washington, DC 20052
E-mail: ajcollegio@email.gwu.edu

Natural scenes consist of objects of varying shapes and sizes. The impact of object size on visual perception has been well-demonstrated, from classic mental imagery experiments¹, to recent studies of object representations reporting topographic organization of object size in occipito-temporal cortex². While the role of real-world physical size in perception is clear, the effect of inferred size on attentional selection is ill-defined. Here, we investigate whether inferred real-world object size influences attentional allocation. Across five experiments, attentional allocation within objects was measured in objects of equal retinal size, but varied in inferred real-world size (e.g., domino, bulldozer). Following each experiment, participants rated the real-world size of each object. We hypothesized that, if inferred real-world size influences attention, selection in retinal-size matched objects should be less efficient in larger objects. We also predicted that this effect should increase with greater attentional demand. These predictions were supported as human observers identified targets faster in objects inferred to be small than large, with costlier attentional shifting in large than small objects when attentional demand was high. Critically, there was a direct correlation between the rated size of individual objects and response times (and shifting costs) to targets within them. Finally, systematic degradation of the visual coherence, and thus size inference, of the objects proportionally reduced the strength of the size effect. These findings strongly suggest that, along with retinal size, inferred real-world object size parametrically modulates attention. These findings have important implications for models of attentional control and invite sensitivity to object size for future studies that use real world images in psychological research.

Visual attention is a spatially-deployed mechanism enabling efficient processing of a vast array of sensory information³⁻⁶. Whereas spatial selection is a fundamental aspect of attentional deployment, there is a rich history of attention research suggesting that objects can also be the

Object Size and Attentional Scaling

unit of attentional selection (i.e., above and beyond spatial contribution). In our environment, objects vary across multiple dimensions, including color, shape, boundaries, size, and semantic meaning. Such object properties enable an observer to distinguish between different objects, and also contribute to attentional deployment^{7–14}.

While the contributions of some object properties (e.g., object boundaries, strength of object representations) to attention have been well studied, the contribution of object size to attentional selection remains ill-defined. Principally, the size of an object on the retina and the inferred size of an object are independent^{15–19}. For example, the retinal size of a cup on your desk is larger than that of a car that can be seen from your window a block away. Changes in retinal object size, that can occur due to a variety of reasons (e.g., changes in depth)²⁰, do not correspondingly reflect changes in inferred object size, demonstrating the importance of object context^{17,21} and the principle of size constancy. Images of real-world objects, such as those employed in our study, are rich in high-level information, including information about each object's real-world size^{15,22–24}. Therefore, we hypothesize that spatial attentional selection might in fact be constrained by top-down inferences of this object property, or by inferred object size.

Whether inferred object size affects perception has been at the center of much object processing and mental imagery research since 1975. There is strong evidence suggesting the size of real-world object representations is inferred to be consistent with the true size of those objects in the real world, such that when participants are asked to imagine, draw, or adjust objects to a preferred viewing size, large real-world objects are consistently represented as larger than small real-world objects in each task^{1,15,25}. This consistent visual size information has been termed canonical size¹⁵, and is dependent on inferences of object size in the real world. Further demonstrating the impact of object size category on object perception, it was observed that

Object Size and Attentional Scaling

canonical object size representations are automatic, that mid-level object properties like texture and overall shape are sufficient to distinguish objects of different real-world sizes, and that the topography of occipito-temporal cortex is organized to reflect canonical object size^{2,22,23,26} in addition to an object's perceived size¹⁸. Interestingly, recent evidence also shows that observers are more likely to miss an object while performing a visual search of a natural scene if the object size is proportionally inconsistent with the scene in which it would typically appear (i.e., giant toothbrush placed on sink)^{27,28} further supporting evidence suggesting that not only is an object's canonical size computed automatically, but that it has consequence for perception.

Whereas cortical organization seems to reflect differentiation by an object's inferred size in addition to the object's retinal size, whether such cortical segmentation has direct impact on attentional selection, rather than general perception, remains unclear. Work by Castiello and Umiltà²⁹ delineated the effect of retinal object size on attentional allocation, showing that, in a spatial cueing task, attention was less efficiently allocated to locations cued by larger compared to smaller squares. Findings from this line of work suggest that attentional allocation within large objects is more diffuse while attention in smaller objects is more focused²⁹. That attention is influenced by the physical size of the object is then well established, but does the same influence extend to inferred, rather than physical, size of objects? The potential mechanism is as follows: following computation of the retinal size of the object, this representation is then influenced, augmented, by the inferred knowledge of an object's real-world size either from occipito-temporal canonical size topography², or potentially through the contribution from parietal cortex object representations that are action/size sensitive^{30,31}. These more complex object representations then serve as the basis for attentional selection.

Here, we directly investigate whether inferred real-world object size, rather than simply

Object Size and Attentional Scaling

retinal object size, influences the focus and shifting of attention within objects. If real-world object size has direct impact on attentional allocation, the efficiency of allocation should be modulated by whether an object is inferred to be canonically large (e.g., billiards table) or small (e.g., domino), as well as by the degree of attentional engagement. We hypothesized that if attention is influenced by, and thus constrained by, the functional localization of inferred object size in occipito-temporal cortex^{2,15}, then attention will be less efficiently deployed in canonically large objects, in a manner referred to henceforth as attentional scaling. Conversely, if attention operates solely according to retinal object size, then attention should be allocated at the same rate, without attentional scaling, across all objects of equal retinal size (independent of inferred size).

Inferred object size was directly manipulated by creating two experimental sets of objects. Objects were classified, a priori, into two groups (canonically small and large) according to a pre-determined metric: objects classified as “large” were body-sized or larger (e.g., billiards table, door), while “small” objects were those that could be held in hand (e.g., domino, cell phone). A modified Posner spatial cueing paradigm was employed⁵, with cues highlighting one end of a single line-drawing of a real-world object presented in the center of the screen (Fig. 1a). Line drawings were used to control for viewpoint and low-level information. According to several controls and metrics, there were no low-level differences between small and large objects (see Methods). Participants performed a T/L target discrimination task with targets appearing either at the cued (valid) or uncued (invalid) location within the object (see Methods). Importantly, while objects subtended the same physical size, they varied in inferred real-world

Object Size and Attentional Scaling

size (domino vs. billiards table).

Figure 1

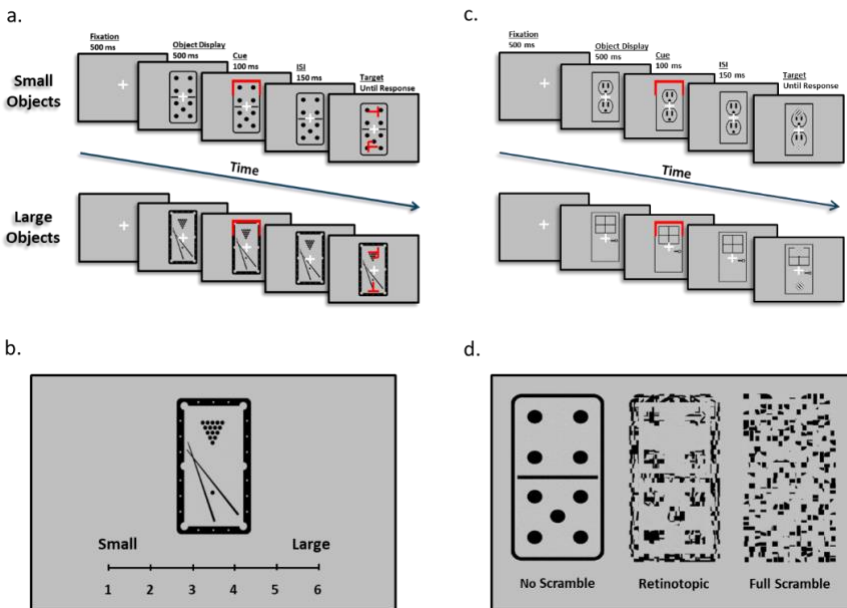


Figure 1: Experimental Design

a) A modified Posner cueing task with task-irrelevant real-world objects was employed in Experiments 1 and 2. Retinal size was controlled, while canonical size varied, and participants identified target letters. b) Post-experiment survey. Subjects ranked size of each object on a scale from 1-6. c) Modified Posner cueing task employed for Experiment 3, in which participants identified Gabor patch orientation. d) Examples of object stimuli (domino) used for scrambling, from left: intact, retinotopically scrambled, and fully scrambled.

We hypothesized that the standard attentional cueing facilitation should be observed independent of an object's inferred size, such that targets will be detected faster and more accurately in validly cued locations compared to invalidly cued locations, termed the spatial validity effect and reflecting spatial selections. Furthermore, and the novel contribution here, if attentional allocation scales with the inferred size of the object, target identification response time (RT) will be slower (and responses less accurate) in large real-world objects compared to smaller objects, in a diffuse/focused manner analogous to the effect of retinal object size delineated in prior research. Notably, the influence of object size may be more nuanced than a categorical difference between small and large objects, as observers have prior knowledge of real-world object size beyond category-level¹⁵. If allocation scales with increasing inferred real-

Object Size and Attentional Scaling

world size, a direct relationship should be observed between real-world object size and RT, such that as objects increase in size, attentional allocation should slow accordingly. To test this, upon completion of each experiment, participants rated the real-world size of each object, in random order, on a scale of one to six (one, very small; six, very large, see Fig. 1b). This served two purposes: (i) validating our a priori classification; and (ii) elucidating a direct relationship between individual's percept of object size and corresponding RTs.

In the first set of experiments (Experiments 1a and 1b), validity of the spatial cue (75% or 50% valid) was manipulated to test whether the top-down spatial predictability of the cue modulates the degree to which object size constrains attention. Given that inference of object's real-world size is a top-down process, it was important to dissociate top-down knowledge from top-down attention. If top-down inference is only triggered by top-down attentional allocation, object size should only constrain attention in the 75% valid condition. However, if object size has a general effect on attentional allocation, size-based attentional modulation should be observed independent of spatial predictability of the cue. To anticipate our findings, we observed direct evidence that attentional selection is modulated by inferred object size, independent of cue validity, in a manner consistent with the attentional scaling hypothesis proposed here (Figs. 2a and 2b). These results suggest more efficient deployment of attention in inferred smaller objects compared to larger objects, analogous to the impact of retinal object size on attention. Importantly, a strong correlation between participants' ratings of real-world object size and response times for individual objects strongly suggest that attentional allocation scales as a function of individual real-world object size, rather than just size category (Fig. 2).

In Experiment 1a (75% valid cue; Fig. 2a), a three-way analysis of variance (ANOVA) with object size (large, small), cue validity (valid, invalid), and object orientation (vertical,

Object Size and Attentional Scaling

horizontal) as within-subjects factors and RT as a dependent measure revealed main effects of inferred object size ($F(1, 19) = 14.50, p < .001, \eta_p^2 = .433$) and cue validity ($F(1, 19) = 22.94, p < .001, \eta_p^2 = .547$), such that attention was allocated less efficiently in canonically large than canonically small objects. No interaction between inferred size and cue validity was observed ($F(1, 19) = 0.52, p = .479$), indicating that while attention was allocated generally more slowly in large objects than in small objects, object size did not directly influence attentional shifts away from the valid location. A main effect of object orientation ($F(1, 19) = 51.65, p < .001, \eta_p^2 = .731$) was observed, suggesting that attention was allocated more efficiently across horizontally-oriented objects than vertically-oriented objects. Imperatively, neither the two-way interaction between inferred size and object orientation ($F(1, 19) = 2.40, p = .138$) nor the three-way interaction between inferred size, cue validity, and object orientation ($F(1, 19) = 0.06, p = .810$), were significant, suggesting that rotation away from an objects' canonical orientation did not modulate size-based attentional scaling. It should also be noted that should there have been any orientation bias influencing size effects (whether canonical orientation, or general orientation peculiarities) the orientation factor would have interacted with the size effect.

In a follow-up analysis, participants' size ratings were regressed against overall target RTs to measure the degree to which participants' perception of object size predicted attentional orienting. Crucially, individuals' ratings of object size were predictive of overall target response time (overall significant regression: $R_{adj}^2 = .314, p < .01$, Fig. 2a) following removal of two outliers (residuals greater than two standard deviation from mean). To assess whether this significant relationship was driven solely by a bimodal distribution (large and small objects), we performed a median split on RTs according to participants' size ratings for each object and conducted linear regression on each half of the data. This split yielded two results: (i) there was a

Object Size and Attentional Scaling

perfect correspondence between the pre-experiment size classification and participant ratings (median split on the data perfectly split objects into small and large) ($t(26) = 14.45, p < .001$); and (ii) size ratings were strongly predictive of RTs within the large object group, ($R_{adj}^2 = .321, p < .05$). Additionally, a tertile split across all objects, conducted to reduce the effect of moderately-sized objects on regression, revealed a significant relationship between size rating and RT within the small object group ($R_{adj}^2 = .385, p < .05$). Taken together, these results indicate that inferred knowledge of real-world object size directly predicts the efficiency of attentional allocation within objects of varying sizes. These findings are consistent with our prediction of attentional scaling according to object size, and imperatively, dovetail with the finding that attention is more diffuse and allocated less efficiently to locations within retinally large objects.

29. In addition to RT, performance accuracy was also analyzed for all experiments, and is reported in the supplementary materials (see supplementary materials, Experiments 1-3). Accuracy results were consistent with those reported in RT. Namely, no effect of size was observed, suggesting that all significant effects were absorbed by the RT measure, and no speed-accuracy tradeoffs were observed, indicating consistency with RT results.

In Experiment 1a, more efficient attentional deployment was observed in small objects compared to larger objects, consistent with our original hypothesis that attention scales according to real-world object size. The observed top-down attentional scaling effects could be partly driven by top-down attentional signal from the spatial predictability of the cue. To test this possibility, cue validity was decreased to 50%, eliminating top-down spatial bias. While spatial validity effects should be reduced with the decrease in spatial predictability, the effect of size should remain. In Experiment 1b (50% valid cue; Fig. 2b), a three-way ANOVA for RT again revealed significant main effects of inferred size ($F(1, 19) = 49.01, p < .001, \eta_p^2 = .721$) and

Object Size and Attentional Scaling

validity ($F(1, 19) = 5.60, p < .05, \eta_p^2 = .227$), with no observed interaction between size and validity ($F(1, 19) = 1.21, p = .285$). Additionally, a main effect of object orientation was observed ($F(1, 19) = 69.78, p < .001, \eta_p^2 = .786$), but the two-way interaction between size and orientation ($F(1, 19) = 0.85, p = .368$), as well as the three-way interaction between size, validity, and orientation ($F(1, 19) = 0.33, p = .572$) did not reach significance, again suggesting that object orientation did not modulate the effect of object size on attention.

Figure 2

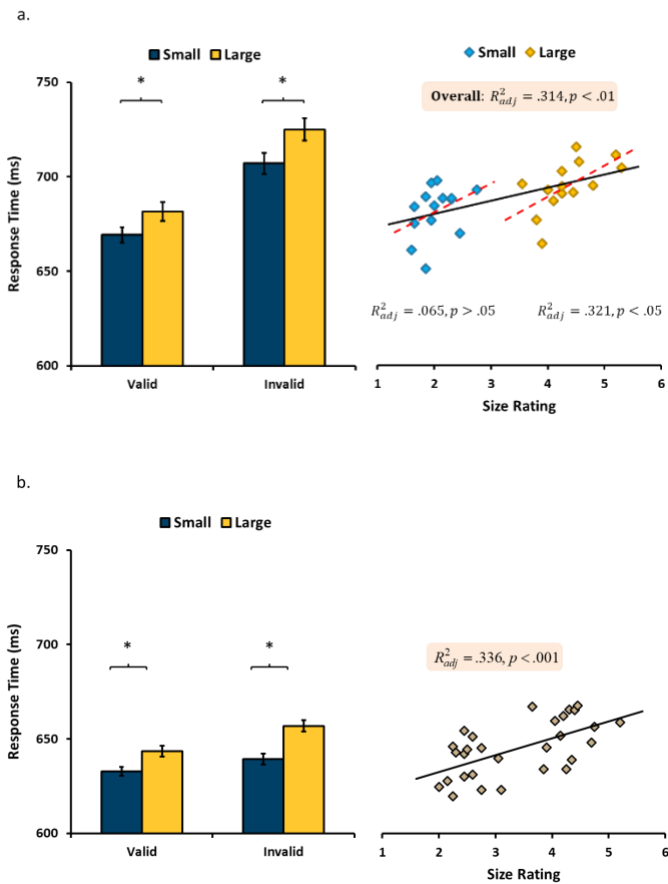


Figure 2: Data for Experiment 1

a) Results for Experiment 1a, 75% validity and short SOA. Significant main effects of both size and validity are observed. b) Results for Experiment 1b, with 50% validity and short SOA. Significant main effects of both size and validity are observed. All error bars represent standard error of the mean (SEM) corrected for within-subjects variance.

These results replicated our finding from Exp. 1a, indicating an effect of size-based attentional scaling: again, individuals' ratings of object size were predictive of response time

Object Size and Attentional Scaling

(overall significant regression: $R_{adj}^2 = .336, p < .001$). Additionally, a between-experiment four-way ANOVA with experiment (75% valid, 50% valid) as a between-subject variable, inferred object size, cue validity, and object orientation as within-subject factors, and RT as a dependent measure revealed main effects of size ($F(1, 38) = 42.82, p < .001, \eta_p^2 = .530$), validity ($F(1, 38) = 28.53, p < .001, \eta_p^2 = .429$), and orientation ($F(1, 38) = 115.67, p < .001, \eta_p^2 = .753$) such that, across both experiments, attention was less efficiently allocated in larger objects, in invalid locations, and in the vertical orientation. A two-way interaction between validity and experiment ($F(1, 38) = 10.43, p < .01, \eta_p^2 = .215$) was significant, consistent with the prediction that validity effect should be larger in the 75% valid condition (Exp. 1a). A two-way interaction between validity and object orientation was also significant ($F(1, 38) = 7.06, p < .05, \eta_p^2 = .157$), suggesting that validity effects are stronger in the horizontal orientation. Importantly, the influence of size on attention remained unaffected by experiment or orientation. In Experiments 1a and 1b, as well as two additional supplemental experiments, in which the time-course and robustness of these effects are tested (see supplementary materials, Experiments S1 and S2), we observe that attention is deployed less efficiently in canonically larger compared to canonically smaller objects. It could also be argued that the observed effect size is driven by the surprise or novelty associated with seeing small objects that share the same retinal size as the large objects. If this were the case, it would be anticipated that the effect of size would diminish with time spent on experiment (i.e., blocks). It was observed, however, that the effect of size remained stable over time (see Fig. S4)

Imperative to the original hypothesis of attentional scaling, inferred knowledge of object size has been shown to modulate deployment of attention within objects. If inferred object size is

Object Size and Attentional Scaling

driving this difference, then a systematic reduction in object identifiability should result in a parametric decrease in the contribution of inferred object size to attentional allocation, and thus lead to diminished differences in RT between large and small objects. In an additional set of two experiments (Experiments 2a and 2b), objects were scrambled to parametrically reduce knowledge of object size (Fig. 1d). Two forms of image scrambling were employed: (i) retinotopic scrambling, in Experiment 2a, preserved the gestalt of the object, such that the object might still be recognizable, despite a lack of low level-fidelity due to scrambled pixels; and (ii) full scrambling, in Experiment 2b, in which pixels were scrambled within an image irrespective of the gestalt of the original object. In this set of experiments, with 75% validity and a cue-to-target stimulus-onset asynchrony (SOA) of 250 ms (as in Experiment 1a), it was predicted that the magnitude of the size-driven effect should decrease parametrically with an increase in scrambling. Specifically, retinotopic scrambling will moderately reduce the effect of size on attentional deployment, while full scrambling will reduce the effect above and beyond retinotopic scrambling, due to the greater degree of object information loss in full scrambling compared to retinotopic scrambling.

A between-experiment three-way ANOVA was conducted with scrambling level (no scrambling, Exp. 1a; retinotopic scrambling, Exp. 2a; and full scrambling, Exp. 2b) as a between-subject variable, and object size (large, small) as a within subject factors (Fig. 3). ANOVA revealed a significant effect of object size ($F(1, 57) = 17.18, p < .001, \eta_p^2 = .232$) and importantly, a significant two-way interaction between object size and scrambling type ($F(2, 57) = 6.35, p < .01, \eta_p^2 = .182$) was observed, such that size effect was only significant in the no-scramble condition ($F(1, 19) = 14.40, p < .001, \eta_p^2 = .431$) and the retinotopic scrambling experiment ($F(1, 19) = 8.58, p < .01, \eta_p^2 = .311$; Figs. 3 and S3a), but not in the full scramble

Object Size and Attentional Scaling

experiment ($F < 1$; Figs. 3 and S3b). Interestingly, dovetailing with the findings, the retinotopically-scrambled objects were rated consistently with our pre-defined large/small groups with 93% accuracy, suggesting a somewhat preserved size inference. On the other hand, size rating consistency was at chance (53.33%) for the fully scrambled objects. Taken together, these results indicate that the size effect, as predicted, was reduced (Fig. 3) in the full scramble condition (no canonical size inference, Exp. 2b, with difference in RTs between large and small objects of $\Delta = -0.67$ ms) relative to the retinotopic condition (somewhat preserved size inference, Exp. 2a, $\Delta = 7.73$) and the no-scramble condition (intact size inference, Exp. 1a, $\Delta = 13.79$ ms).

Figure 3

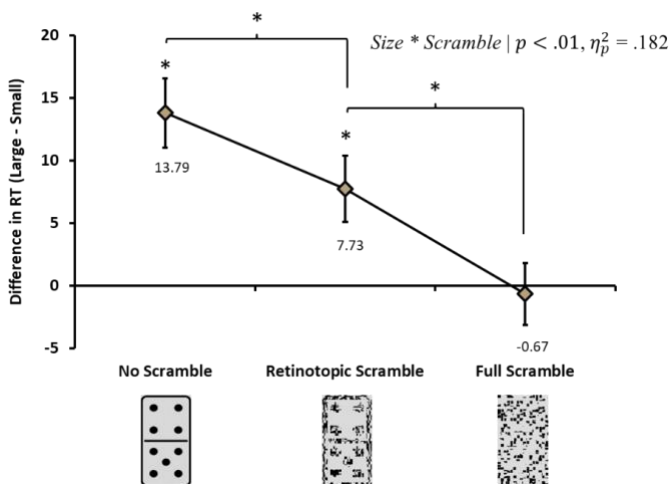


Figure 3: Data for Experiment 2 – Comparison across all scrambling conditions

Scrambling of objects reduced contribution of object size to attention. As objects become more difficult to identify, the contribution of size is reduced. Error bars represent standard error of the mean.

Experiment 3 was designed with three aims: (i) internal replication with a different target; (ii) testing the effect of increased attentional demand; and (3) addressing a possible confound of target spatial frequency. One of the important aspects of object size is a consideration that objects of varying size also vary in terms of the spatial frequency of their features (i.e., large objects, low spatial frequency; small objects, high spatial frequency)³², and that the results demonstrated in Experiment 1 and 2 could partly be confounded by a mismatch between target

Object Size and Attentional Scaling

spatial frequency and object spatial frequency. High spatial frequency targets may be processed less efficiently in low spatial frequency, large objects, leading to the attentional scaling effect. To address this potential confound, and to provide additional internal replication of our original results, in Experiment 3 participants performed a target identification task, as in Experiments 1 and 2, but the nature of the target was different. In this experiment, instead of performing a T/L discrimination task, participants identified the orientation of a target Gabor patch (45° or -45°) (Fig. 1c). Additionally, the spatial frequency of the target Gabor patches varied between 4 cycles per degree (low frequency) and 8 cycles per degree of visual angle (high spatial frequency) in a within-subjects design. If attention is affected by the spatial frequency of targets as they relate to object spatial frequency, a congruency effect should be observed, such that attention will be allocated more efficiently when target spatial frequency matches object spatial frequency (e.g., high spatial frequency, small object; low spatial frequency, large object). Alternatively, if attention is only constrained by inferred object size, as we have argued up to this point, we expect to observe two findings: (i) a main effect of object size, same as in prior experiments; and (ii) an interaction with frequency such that effect of size should be large in a more difficult, high frequency (less discriminability), task.

For Experiment 3, an RT ANOVA was conducted with object size (large, small), cue validity (valid, invalid), object orientation (vertical, horizontal), and target spatial frequency (low, high) as within-subject factors (Fig. 4). ANOVA revealed main effects of inferred object size ($F(1, 19) = 25.11, p < .001, \eta_p^2 = .569$) with faster RTs to small objects; cue validity ($F(1, 19) = 29.35, p < .001, \eta_p^2 = .607$) with faster RTs to validly cued targets; object orientation ($F(1, 19) = 6.20, p < .05, \eta_p^2 = .246$) with faster RTs in horizontally-oriented objects; and target spatial frequency ($F(1, 19) = 43.99, p < .001, \eta_p^2 = .698$), with faster RTs for the low frequency targets.

Object Size and Attentional Scaling

In addition to replicating main effects of object size and validity, the main effect of target spatial frequency indicates that participants are slower for high spatial frequency targets compared to low frequency targets, regardless of object size, indicating this is a result of task difficulty and not spatial frequency mismatch.

A two-way interaction between inferred object size and cue validity was marginally significant ($F(1, 19) = 4.09, p = .057, \eta_p^2 = .177$), suggesting that, in addition to general attentional scaling resulting from a more diffuse focus of attention in large than small objects (observed in Experiments 1 and 2), as predicted, attentional shifts between cue valid and invalid locations are also modulated by inferred object size, such that cue validity effects are larger in large objects ($\Delta = 56.6$ ms) compared to small objects ($\Delta = 45.8$ ms). Crucially, a significant three-way interaction between inferred object size, cue validity, and target spatial frequency was also significant ($F(1, 19) = 7.99, p < .05, \eta_p^2 = .296$), indicating that target spatial frequency (low or high) modulates the degree to which attentional shifts are constrained by object size. When target spatial frequency is high, the task is more difficult and attentionally demanding, and modulation of attentional shifts by object size is observed (large objects, $\Delta = 70.8$ ms; small objects, $\Delta = 40.3$ ms). This was further supported by a significant relationship observed between object size rating and cost of attention shift for high spatial frequency targets ($R_{adj}^2 = .210, p < .01$, Fig. 4c). When target spatial frequency is low, modulation of attentional shifts by object size is reduced (large objects, $\Delta = 43.0$ ms; small objects, $\Delta = 52.5$ ms), and attention is constrained by inferred object size in a more general, scaling fashion (i.e., less efficient in large objects). No relationship was observed between object size rating and attentional shifting costs with low spatial frequency targets. No other effects, including any congruency effect of inferred object

Object Size and Attentional Scaling

size and target spatial frequency, were observed (all F s < 1).

Importantly, the 4-way interaction between size, validity, frequency, and orientation was non-significant ($F(1, 19) = .44, p = .513$), suggesting that object orientation did not affect the crucial three-way interaction between size, validity, and frequency (see supplementary orientation analyses section for additional lower-level orientation interactions).

Figure 4

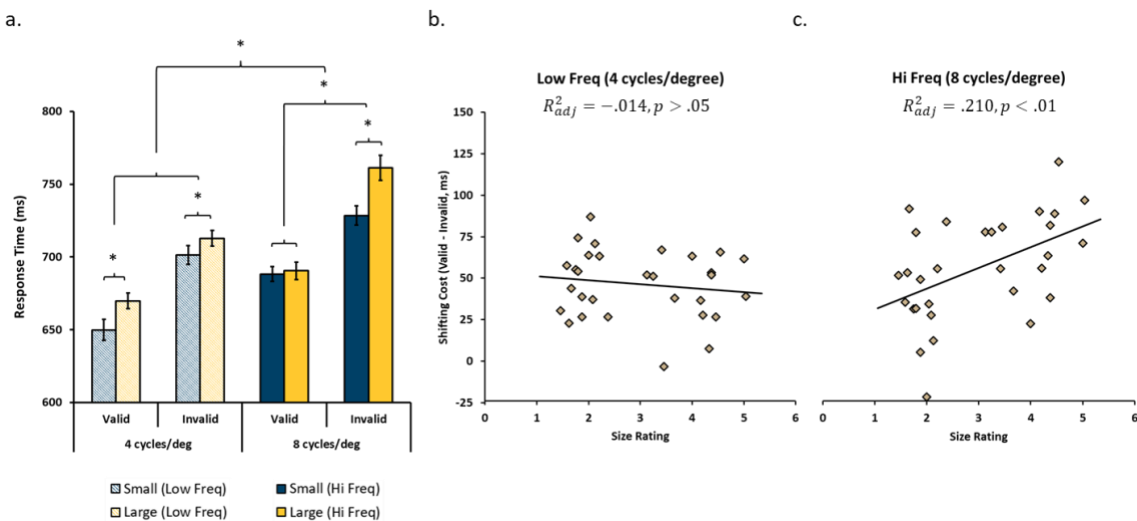


Figure 4: Data for Experiment 3

a) Results for Experiment 3, in which target Gabor spatial frequency varied between 4 cycles/degree and 8 cycles/degree. Low frequency data shown in striped bars; high frequency data shown in solid bars. Error bars represent standard error of the mean corrected for within-subjects variance. b) No correlation was observed between shift costs and object size ratings with low spatial frequency targets. c) Strong positive correlation between object size ratings and shift costs with high spatial frequency targets.

More sluggish processing of spatial locations within canonically larger objects compared to smaller objects suggests that inferred real-world size modulates attentional allocation, in a scaling fashion. Critically, this effect is modulated by the degree of top-down knowledge available about the object, as evidenced by the reduction of size-based attentional modulation according to the degree of information loss about object size due to image scrambling (Fig. 3). Finally, under attentionally demanding conditions, shifts of attention are also constrained by

inferred object size, with slower attentional shifts within large objects.

As much progress has been made in the understanding of how objects guide attention, it is becoming imperative to move toward real-world objects to better understand what properties of objects facilitate attentional selection. Real-world objects contain a rich array of high-level information that may also constrain attentional selection, alongside low-level features of more basic, two-dimensional objects, such as the rectangles in a traditional two-rectangle display¹⁰. These novel findings provide the first such evidence of high-level size-based attentional constraint, in the form of attentional scaling within line-drawings of real-world objects.

While novel with respect to equal retinal size, the findings dovetail with prior research, suggesting that attention scales according to both the physical and the perceived size of an object^{16–18,29,33,34}. One elegant study aimed to disentangle the complex relationship between retinal and perceived object size through the use of the Ponzo illusion and rectangles, concluding that perceived object size constrains attentional selection¹⁶, and that, in concordance with work conducted by Castiello and Umiltà, attention is more tightly focused in objects perceived to be smaller. The study presented here concludes that attention is not influenced by perceived size alone, but by an inferred knowledge of object size in the real-world. This inferred canonical size, an inherent high-level property of real-world objects, constrains topographic organization within object-selective cortical regions, but also influences how attention is deployed. Both the focus and shifting of attention are affected by inferred real-world object size, rather than retinal object size alone.

Mechanistically, our findings point to a possible functional relationship between object-selective cortical regions of occipito-temporal cortex, mapped according to inferred object size, and parietal cortex and the attentional priority map, in that real-world object size perception

affects attentional deployment^{2,16,17}. This has yet to be explored, and the time is ripe for pursuits of the neural basis of size-based attention scaling. It should be noted that the real-world size-specific functional response in the inferior temporal cortex (IT) lies in good register with the known maps of eccentricity, in many cases extending them. Such retinotopic constraints on the high-level visual representations are widespread in both human and monkey³⁵ for both animate and inanimate objects. Importantly, this organization is present even in studies where the actual retinal size and position is held constant, as in our study. Thus, our results may be related to the presence of this inferred real-world object size organization in IT. The precise mechanism relating these two factors will need to be the subject of further study.

This research also holds strong implications for scene processing as well, as scene processing is a complex behavior by which numerous objects are parsed. According to an observer's prior experience with an object, the context of an object may provide information about an object's size, which will impact the manner in which the object engages attentional selection. High-level object representations and object context modulate attentional allocation, and novel work must be conducted to investigate the role of real-world objects and inferred size on attention, both space- and object-based. Additional high-level properties of objects, such as how large an object is imagined to be via mental imagery, or whether an object is known to be animate, may also constrain attentional deployment in a scaling manner. Previous work suggests that if participants are told an object is a toy, the canonical size of the toy object is smaller than the object it represents²², suggesting a change in the mental representation of the object, while other work suggests that imagined objects are also topographically organized according to real-world size in occipito-temporal cortex²⁶. However, animate objects do not show a size-based topographic organization³⁶. If this topographic organization is necessary for size-based

Object Size and Attentional Scaling

attentional scaling, animate objects should not elicit a scaled attentional response. Taken together, our findings and these implications suggest that caution should be taken when using object images in psychological research, as inferred object size constrains processing of and attention to objects.

Methods

Participants. For Experiment 1a, 25 participants were recruited. Five participants were excluded for failing to meet the 90% accuracy criteria. For Experiment 1b, 25 participants were recruited. Five participants were excluded for failing to meet the 90% accuracy criteria. For Experiment 2a, 26 participants were recruited. Six participants were excluded for failing to meet the 90% accuracy criteria. For Experiment 2b, 26 participants were recruited. Five participants were excluded for failing to meet the 90% accuracy criteria. For Experiment 3, 25 participants were recruited. Five participants were excluded for failing to meet the 90% accuracy criteria. All participants were recruited from The George Washington University participant pool, reported normal or corrected-to-normal visual acuity, and were naïve to the purpose of the experiment. All experiments consisted of completely independent sets of participants.

Apparatus and Stimuli. For all experiments, stimuli were presented on a 19" Dell 1908FP Color LCD monitor with a refresh rate of 60 Hz, positioned at a distance of 62 cm from the viewer. All experiments were generated and presented using PsychoPy v1.82^{37,38}. Stimuli consisted of line drawings of real-world, everyday objects drawn both from The Noun Project, an online repository of object icons and clip art (<https://thenounproject.com>), as well as from a set of images rendered by the authors. The object stimuli consisted of 15 line-drawings of small objects, and 15 line-drawings of large objects. The small objects were as follows: soda can, cell

Object Size and Attentional Scaling

phone, cigarette lighter, credit card, domino, envelope, iPod, mason jar, computer mouse, electric outlet, playing card, salt shaker, aerosol canister, toothpaste tube, and watch. The large objects were: basketball hoop, billiards table, bulldozer, car, sofa, door, grandfather clock, jukebox, kayak, mirror, phone booth, porta potty, throne, bathtub, and vending machine. All objects are displayed in supplementary materials (Figure S5).

For each experiment, the display consisted of a single central fixation point, which subtended a visual angle of $1^\circ \times 1^\circ$, and a single central line-drawing of an object, subtending a visual angle of $4^\circ \times 8^\circ$ oriented either vertically (as shown in Fig. S5a) or horizontally (rotated 90° clockwise, as shown in Fig. S5b). Object orientation was full counterbalanced within subject for each experiment. Following the presentation of an object, a cue was presented at either end of the object, with equal probability. For Experiments 1 and 2, following the offset of the cue, a single target letter (T or L, equiprobable) always appeared at either of the two ends of the object, while a distractor letter (T/L hybrid) always appeared at the opposite end of the object. For Experiment 3, a single Gabor patch oriented 45° or -45° equiprobably appeared at either of the two ends of the object, while a distractor checkerboard appeared at the opposite end. Gabor spatial frequency equiprobably varied between 4 cycles/degree of visual angle and 8 cycles/degree of visual angle. Note that we used a factor of 2 rather than factor of ~ 10 (more reflective of real-world differences) for spatial frequencies with the goal of increasing spatial attention demands and not necessarily matching real world frequency differences. Targets and distractors each subtended a visual angle of $1^\circ \times 1^\circ$ and were presented with a center-to-center distance of 5° . Line drawings were rendered in black (RGB = 0, 0, 0) against a light gray background (RGB = 192, 192, 192). Cues, targets, and distractors were rendered in red (RGB = 255, 0, 0) in Experiments 1 and 2, while targets and distractors were rendered in black (RGB = 0,

Object Size and Attentional Scaling

0, 0) and white (RGB = 255, 255, 255) in Experiment 3. Additionally, all stimuli were controlled for overall luminance, and tested to ensure that small and large objects did not differ.

To demonstrate that size effects in the two main and two supplementary experiments are truly driven by inferred object size knowledge, rather than any low-level group or individual differences between canonically large and small objects, low-level features (specifically, luminance and retinal size) of each object were directly compared¹⁴. A 100-bin histogram based on luminance values was created for each object. Intensities greater than 128 were discounted from analysis, as these pixels belonged to the gray background of each image. Histograms were normalized, and all objects were compared to each other and to the group mean. The pixel counts for each object were compared across conditions to account for potential differences in retinal object size. Two-sample t-tests for both relative pixel intensity ($t(28) < 1$) and retinal object size ($t(28) = 1.68, p = .10$) revealed no significant differences between groups, and one-sample t-tests comparing relative pixel intensity ($t(14) < 1$) and retinal object size (small objects: $t(14) = 1.66, p = .12$; large objects: $t(14) < 1$) to the overall group average revealed no individual object differed from the group mean, suggesting that less efficient allocation of attention within larger objects was not due to systematic differences in low-level features between large and small objects.

To control for any further possible conflation between inferred object size and low-level features, individual ratings of inferred real-world object size were also regressed against average luminance value for each object, and luminance values were regressed with RT. This was done to test whether average object luminance was predictive of performance (i.e., worse performance in darker objects), despite a lack of group differences in low-level features. For Experiment 1a, neither a size by luminance ($R_{adj}^2 = -.014, p = .447$) nor a luminance by response time ($R_{adj}^2 = -$

Object Size and Attentional Scaling

.005, $p = .361$) interaction were observed, indicating that inferred object size, and not the average luminance of each object, was predictive of response time. For Experiment 1b, neither size by luminance ($R_{adj}^2 = -.026$, $p = .619$) nor luminance by response time ($R_{adj}^2 = -.020$, $p = .515$) interactions were observed.

While low-level features (luminance, retinal object size) were controlled across objects, it remains possible that clutter, another low-level feature, is responsible for differences between canonically large and small objects. Due to remaining concerns about possible clutter differences between object groups, a clutter analysis assessing contrast was conducted across all objects, and then analyzed for differences between groups³⁹. Per a two-samples t-test of clutter values for large and small real-world objects, large objects were on average more cluttered than small objects ($t(28) = 2.08$, $p < .01$). However, when eliminating the top 5 most-cluttered large items (i.e., grandfather clock, jukebox, phonebooth, porta-potty, vending machine) and top 5 least-cluttered small items (i.e., computer mouse, lighter, playing card, soda can, toothpaste tube) from analysis, leaving only the middle 20 objects, the groups no longer differed in average clutter ($t(18) = 2.14$, $p = .46$ (two-tailed)), yet the size-based modulation of attentional allocation remained when accounting for only the middle 20 objects in Experiment 1a ($F(1, 19) = 16.66$, $p < .001$, $\eta_p^2 = .467$) and Experiment 1b ($F(1, 19) = 25.98$, $p < .001$, $\eta_p^2 = .578$). Furthermore, the degree of contrast clutter in the middle 20 objects did not significantly predict RT in a regression analysis in either experiment (all F s < 1), despite the significant prediction of RT by object size rating. Finally, in Experiment 3, size-based modulation of attentional shifting also remained when accounting for only the middle 20 objects, as the analysis revealed a significant main effect of size ($F(1, 19) = 31.61$, $p < .001$, $\eta_p^2 = .625$), a marginally significant interaction between size and validity ($F(1, 19) = 3.52$, $p = .076$, $\eta_p^2 = .156$), and a significant three-way interaction

Object Size and Attentional Scaling

between size, validity, and target spatial frequency ($F(1, 19) = 5.19, p = .034, \eta_p^2 = .215$), indicating that both size and spatial frequency modulated attentional shifts. Crucially, the four-way interaction between size, validity, object orientation, and frequency, was not significant ($F(1, 19) = 0.10, p = .757$). This result suggests that object orientation does not affect the critical three-way interaction between object size, cue validity, and target frequency. Additional lower-level orientation interactions are included in the Clutter Control section of supplementary orientation analyses. Taken together, these results further suggest that the effect of size is not driven by low-level features or object orientation.

Possible mid- and high-level differences between small and large objects were investigated using a two-pronged approach (see Supplementary Experiment 3). First, in a separate Mechanical Turk survey, participants were asked to rate the overall curvature and familiarity of each object in the object set, as previous work has suggested that large and small objects may differ in both curvature (mid-level) and familiarity (high-level) due to the ergonomic features of small objects. Second, a steerable pyramid image decomposition analysis was performed to assess the angular features and relative curvilinearity of small and large objects. To preface the findings of Experiment S3, no differences were observed across object groups in either the Mechanical Turk survey or the steerable pyramid analysis. These results suggest that size-based attentional scaling is not due to differences in shape features or familiarity between large and small objects.

Design & Procedure. In Experiment 1, cue validity (75%, 50%) was manipulated between subjects (Experiments 1a and 1b). For each experiment, a 2 x 2 x 2 within-subjects factorial design was employed, with inferred object size (large, small), cue-target relation (valid, invalid), and object orientation (vertical, horizontal) as within-subject factors. Response times

Object Size and Attentional Scaling

(RT) were the primary measure of interest. The letters T and L were used as target stimuli, while a T/L hybrid, which contains components of both a T and an L, was used as a distractor (see Fig. 1a). The “T” target was mapped onto the “c” keyboard response, while the “L” target was mapped on the “m” keyboard response. Targets were considered valid if they appeared in the same location as the cue, and invalid if they appeared in the opposite location.

In Experiment 2, the procedure was identical to that used in Experiment 1a. However, line-drawing images were scrambled using MATLAB (Mathworks, Natick, MA) in order to reduce the contribution of inferred object knowledge to attention. Two forms of scrambling were used. In Experiment 2a, images were scrambled using a retinotopic technique, in which each object was divided into 2500 6x6 pixel grains (36 px² region). The 6x6 grains were then scrambled across all objects in a manner that preserved the average luminance of each 6x6 pixel grain location. This conservative method roughly preserved the boundaries of each object, but reduced the quality of each image. In Experiment 2b, a more general scrambling technique was used, in which no pixel information was preserved, conserving no identifiable information about each object.

In Experiment 3, the procedure was also identical to that used in Experiment 1a, with the exception that targets were Gabor patches rather than T/L renderings. The left-leaning Gabor target (-45°) was assigned to the “c” keyboard response, while the right-leaning Gabor target (45°) was assigned to the “m” keyboard response.

Each trial began with a 500 ms presentation of one object line-drawing. Following the presentation of the line-drawing, a red cue was presented on one of the two ends of the object, followed by a brief interval prior to the presentation of target stimuli. In Experiments 1, 2, and 3, the cue lasted 100 ms, while the interstimulus interval (ISI) was 150 ms. Following the ISI,

Object Size and Attentional Scaling

targets and distractors remained on the screen for two seconds or until response. Participants were encouraged to respond as quickly and as accurately as possible. Each participant in each experiment completed a practice block consisting of 30 trials, followed by eight experimental blocks, each consisting of 120 trials.

In the post-experiment survey, a single line drawing was presented centrally, with a 6-point rating scale presented underneath (Fig. 1a). Participants were instructed to rate the size of each of the objects on a scale of one to six, one being very small and six being very large. The size ratings were used as an independent measure of each object's real-world size, and participants were agnostic to the size classifications of the objects made prior to the experiments.

Data Analyses. Data for each experiment were analyzed using repeated-measures analysis of variance. Two-sample and paired t-tests were also employed where indicated in the text and were two-tailed unless otherwise specified. Anticipatory response times faster than 100 ms as well as any response time longer than 1500 ms were removed from analysis (1.9%, 2.9, 1.1%, 5.0%, and 6.6% data removal for Exps. 1a, 1b, 2a, 2b, and 3, respectively. Mean response times for each condition (small, large object; valid, invalid location) were calculated for each subject.

Data Availability. The data that support the findings of this study are available from the corresponding author upon reasonable request.

Code Availability. The custom PsychoPy and MATLAB programs generated for this study are available from the corresponding authors upon reasonable request.

References

1. Kosslyn, S. M. Information representation in visual images. *Cogn. Psychol.* **7**, 341–370 (1975).
2. Konkle, T. & Oliva, A. A Real-World Size Organization of Object Responses in Occipitotemporal Cortex. *Neuron* **74**, 1114–1124 (2012).
3. Carrasco, M. Visual attention: The past 25 years. *Vision Res.* **51**, 1484–1525 (2011).
4. Desimone, R. & Duncan, J. Neural Mechanisms of Selective Visual Attention. *Annu. Rev. Neurosci.* **18**, 193–222 (1995).
5. Posner, M. I., Snyder, C. R. R. & Davidson, B. J. Attention and the detection of signals. *J. Exp. Psychol. Gen.* **109**, 160–174 (1980).
6. Eriksen, C. W. & St. James, J. D. Visual attention within and around the field of focal attention: A zoom lens model. *Percept. Psychophys.* **40**, 225–240 (1986).
7. Duncan, J. Selective Attention and the Organization of Visual Information. *J. Exp. Psychol. Gen.* **113**, 501–517 (1984).
8. Valdes-Sosa, M., Bobes, M. a, Rodriguez, V. & Pinilla, T. Switching attention without shifting the spotlight object-based attentional modulation of brain potentials. *J. Cogn. Neurosci.* **10**, 137–151 (1998).
9. O’Craven, K. M., Downing, P. E. & Kanwisher, N. fMRI evidence for objects as the units of attentional selection. *Nature* **401**, 584–587 (1999).
10. Egly, R., Driver, J. & Rafal, R. D. Shifting Visual Attention Between Objects and Locations : Evidence From Normal and Parietal Lesion Subjects. *J. Exp. Psychol. Gen.* **123**, 161–177 (1994).
11. Drummond, L. & Shomstein, S. The timecourse of space- and object-based attentional prioritization with varying degrees of certainty. *Front. Integr. Neurosci.* **7**, 1–10 (2013).
12. Muller, N. G. & Kleinschmidt, A. Dynamic Interaction of Object- and Space-Based Attention in Retinotopic Visual Areas. *J. Neurosci.* **23**, 9812–9816 (2003).
13. Shomstein, S. & Behrmann, M. Cortical systems mediating visual attention to both objects and spatial locations. *Proc. Natl. Acad. Sci.* **103**, 11387–11392 (2006).
14. Malcolm, G. L., Rattinger, M. & Shomstein, S. Intrusive effects of semantic information on visual selective attention. *Attention, Perception, Psychophys.* **78**, 2066–2078 (2016).
15. Konkle, T. & Oliva, A. Canonical visual size for real-world objects. *J. Exp. Psychol. Hum. Percept. Perform.* **37**, 23–37 (2011).
16. Jefferies, L. N., Gmeindl, L. & Yantis, S. Attending to illusory differences in object size. *Attention, Perception, Psychophys.* **76**, 1393–1402 (2014).
17. Robertson, L. C. & Kim, M.-S. Effects of Perceived Space on Spatial Attention. *Psychol. Sci.* **10**, 76–79 (1999).
18. Murray, S. O., Boyaci, H. & Kersten, D. The representation of perceived angular size in

- human primary visual cortex. *Nat. Neurosci.* **9**, 429–434 (2006).
19. Konkle, T. & Oliva, A. Normative Representation of Objects: Evidence for an Ecological Bias in Object Perception and Memory. *Proc. Twenty-Nine Annu. Conf. Cogn. Sci. Soc.* 407–412 (2007).
 20. Epstein, W. & Baratz, S. S. Relative Size in Isolation As a Stimulus for Relative Perceived Distance. *J. Exp. Psychol.* **67**, 507–513 (1964).
 21. Bar, M. Visual objects in context. *Nat. Rev. Neurosci.* **5**, 617–629 (2004).
 22. Long, B., Konkle, T., Cohen, M. A. & Alvarez, G. A. Mid-Level Perceptual Features Distinguish Objects of Different Real-World Sizes. *J. Exp. Psychol. Gen.* **145**, 95–109 (2016).
 23. Konkle, T. & Oliva, A. A familiar-size Stroop effect: Real-world size is an automatic property of object representation. *J. Exp. Psychol. Hum. Percept. Perform.* **38**, 561–569 (2012).
 24. Malcolm, G. L. & Shomstein, S. Object-based attention in real-world scenes. *J. Exp. Psychol. Gen.* **144**, 257–263 (2015).
 25. Biederman, I. & Cooper, E. E. Size invariance in visual object priming. *J. Exp. Psychol. Hum. Percept. Perform.* **18**, 121–133 (1992).
 26. Kosslyn, S. M., Thompson, W. L., Kim, I. J. & Alpert, N. M. Topographical representations of mental images in primary visual cortex. *Nature* **378**, 783–785 (1995).
 27. Eckstein, M. P., Koehler, K., Welbourne, L. E. & Akbas, E. Humans, but Not Deep Neural Networks, Often Miss Giant Targets in Scenes. *Curr. Biol.* 1–6 (2017). doi:10.1016/j.cub.2017.07.068
 28. Wolfe, J. M. Visual Attention: Size Matters. *Curr. Biol.* **27**, R1002–R1003 (2017).
 29. Castiello, U. & Umiltá, C. Size of the Attentional Focus and Efficiency of Processing. *Acta Psychol. (Amst)*. **73**, 195–209 (1990).
 30. Martin, A. The Representation of Object Concepts in the Brain. *Annu. Rev. Psychol.* **58**, 25–45 (2007).
 31. Chao, L. L. & Martin, A. Representation of manipulable man-made objects in the dorsal stream. *Neuroimage* **12**, 478–484 (2000).
 32. Sekuler, R. & Blake, R. *Perception*. (Alfred A. Knopf, 2005).
 33. Galera, C., Von Grünau, M. & Panagopoulos, A. Automatic focusing of attention on object size and shape. *Psicologica* **26**, 147–160 (2005).
 34. Herrmann, K., Montaser-Kouhsari, L., Carrasco, M. & Heeger, D. J. When size matters: attention affects performance by contrast or response gain. *Nat. Neurosci.* **13**, 1554–1559 (2010).
 35. Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G. & Mishkin, M. The ventral visual pathway: An expanded neural framework for the processing of object quality. *Trends Cogn. Sci.* **17**, 26–49 (2013).
 36. Konkle, T. & Caramazza, A. Tripartite Organization of the Ventral Stream by Animacy

- and Object Size. *J. Neurosci.* **33**, 10235–10242 (2013).
37. Peirce, J. W. PsychoPy-Psychophysics software in Python. *J. Neurosci. Methods* **162**, 8–13 (2007).
 38. Peirce, J. W. Generating stimuli for neuroscience using PsychoPy. *Front. Neuroinform.* **2**, 1–8 (2009).
 39. Rosenholtz, R., Li, Y. & Nakano, L. Measuring visual clutter. *J. Vis.* **7**, 17.1-22 (2007).

Acknowledgements

This work was supported by a National Science Foundation grant BCS-1534823 to S.S. We thank Dr. George Malcolm, Dr. Dwight Kravitz, and Dr. Marlene Behrmann for insightful comments on an earlier version of this manuscript. Special thank you to Dr. Mary Peterson for suggesting a spatial frequency control experiment reported in Experiment 3.

Author Contributions

A.J.C. and S.S. contributed to theoretical motivation, developed study design, and wrote the paper. A.J.C. programmed and conducted the experiments, performed data collection and analysis. J.C.N. and P.S.S. assisted with theoretical motivation, programming experiments, and data collection, and provided revisions on manuscript drafts.

Competing Interests

The authors declare no competing interests.

Materials & Correspondence

Correspondence and requests for materials should be addressed to A.J.C. or S.S.

Figure Legends

Figure 1: Experimental Design

a) A modified Posner cueing task with task-irrelevant real-world objects was employed in Experiments 1 and 2. Retinal size was controlled, while canonical size varied, and participants identified target letters. b) Post-experiment survey. Subjects ranked size of each object on a scale from 1-6. c) Modified Posner cueing task employed for Experiment 3, in which participants identified Gabor patch orientation. d) Examples of object stimuli (domino) used for scrambling, from left: intact, retinotopically scrambled, and fully scrambled.

Figure 2: Data for Experiment 1

a) Results for Experiment 1a, 75% validity and short SOA. Significant main effects of both size and validity are observed. b) Results for Experiment 1b, with 50% validity and short SOA. Significant main effects of both size and validity are observed. All error bars represent standard error of the mean (SEM) corrected for within-subjects variance.

Figure 3: Data for Experiment 2 – Comparison across all scrambling conditions

Scrambling of objects reduced contribution of object size to attention. As objects become more difficult to identify, the contribution of size is reduced. Error bars represent standard error of the mean.

Figure 4: Data for Experiment 3

a) Results for Experiment 3, in which target Gabor spatial frequency varied between 4 cycles/degree and 8 cycles/degree. Low frequency data shown in striped bars; high frequency data shown in solid bars. Error bars represent standard error of the mean corrected for within-subjects variance. b) No correlation was observed between shift costs and object size ratings with low spatial frequency targets. c) Strong positive correlation between object size ratings and shift costs with high spatial frequency targets.