The Role of Real-World Size in Object Representation

by

Talia Konkle

B.A., Cognitive Science & Applied Mathematics
University of California, Berkeley (2004)

Submitted to the Department of Brain and Cognitive Sciences
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Cognitive Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2011

Author ....
Department of Brain and Cognitive Sciences
July 15, 2011

Certified by ....
Aude Oliva
Associate Professor of Cognitive Science
Thesis Supervisor

Accepted by ....
Earl K. Miller
Picower Professor of Neuroscience
Director, BCS Graduate Program

The Role of Real-World Size in Object Representation

by

Talia Konkle

Submitted to the Department of Brain and Cognitive Sciences
on July 15, 2011, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Cognitive Science

Abstract

Every object in the world has a physical size which is intrinsic to how we interact with it: we pick up small objects like coins with our fingers, we throw footballs and swing tennis rackets, we orient our body to bigger objects like chairs and tables and we navigate with respect to landmarks like fountains and buildings. Here I argue that the size of objects in the world is a basic property of object representation with both behavioral and neural consequences. Specifically, I suggest that objects have a canonical visual size based on their real-world size (Chapter 2), and that we automatically access real-world size information when we recognize an object (Chapter 3). Further, I present evidence that there are neural consequences of real-world size for the large-scale organization of object knowledge in ventral visual cortex (Chapter 4). Specifically, there are regions with differential selectivity for big and small objects, that span from along the dorsal and lateral surfaces of occipito-temporal cortex in a mirrored organization. Finally, I suggest that the empirical findings can be coherently explained by thinking about the experience of an observer situated in a three-dimensional world. This work provides testable predictions about retinal size biases in visual experience, and an approach in which to understand the neural representation of any object in the world.

Thesis Supervisor: Aude Oliva
Title: Associate Professor of Cognitive Science

Acknowledgments

Thanks to Aude Oliva, my advisor, for her tremendous ability to make courageous paradigms possible, and for reminding me, when I needed to hear it, that I have to find my own way of doing science and that I should dream big.

Thanks to Chris Moore, for including me as a part of his lab, for investing time and resources in me so generously, and for valuable conversation about possible mechanisms of neural circuits and systems.

Thanks to Nancy Kanwisher and Rebecca Saxe, both awe-inspiring role models, for their engagement and incisive feedback on my research over the years, and for continuing that commitment as part of my thesis committee. Thanks to Dan Kersten, outside member of my thesis committee, for his energy, smarts, and time.

Thanks to the people in the Oliva and Moore labs and in my graduate class, for being undeniably the most intelligent people I have ever met, and then choosing to hang out with me. Special thanks to Soojin Park, for launching me into the world of fMRI; to Barbara Hidalgo-Sotelo, for her compassionate ability to listen; to Mike Frank and Ed Vul, for defining my early graduate experience and enriching my life outside of science; and in particular to Tim Brady, my frequent coauthor and best friend, for being loud, selfless, and absolutely brilliant.

Thanks to George Alvarez. Your confidence and belief in my abilities as a scientist surpass my own, and that is an incredible thing to come home to. This thesis is dedicated to you. Thanks to my family, Ava, Maurice, and Nick. You are amazing individuals and I continually draw my inspiration from you.

Contents

1 Introduction **9**

Models of Object Representation 11

Organization of object representations in cortex 12

The size of objects in the natural world 13

Previous research on real-world size 16

Contributions of this thesis 17

2 Canonical visual size for real-world objects **19**

Introduction 19

Size Ranking 22

Experiment 1: Drawings from Memory 25

Experiment 2: Imagery 32

Experiment 3: Perception 36

Experiment 4: Miniatures 40

Experiment 5: Size Range 43

General Discussion 47

Appendix 55

3 A Familiar Size Stroop Effect: Real-world size is an automatic property of object representation **57**

Introduction 57

Experiment 1: Familiar Object Stroop Task 59

Experiment 2: One-Shot Learning Stroop Task 62

General Discussion 65

Supporting Information 68

4 The representation of objects in ventral temporal cortex depends on real-world size **71**

Introduction 71

Results 73

Discussion 81

Methods 88

Supporting Information 93

5 Conclusions **99**
Summary of Empirical Findings 99
A Situated-Observer Framework 103
Conclusion 107

Appendix:
Consequences of vision in a three-dimensional world **109**
Retinal Size and Eccentricity 110
Shape Biases, Spatial Frequency, and Orientation 111
Upper-Lower visual field biases 113
Correlated low-level features drive high-level regions 114
Summary 116

References **117**

Chapter 1

Introduction

Over a lifetime of visual experience, our visual system builds a storehouse of knowledge about the visual world. Incoming visual information rapidly makes contact with these existing object representations, enabling us to effortlessly recognize objects that are presented for only 100s of milliseconds (e.g. Grill-Spector & Kanwisher, 2006; Thorpe, Fize, & Marlot, 1996; for review see Logothetis & Shineburg, 1996). Two fundamental endeavors for visual cognition and cognitive neuroscience are to characterize the nature of these existing object representations and to understand how are they organized in the brain.

Much behavioral research has focused on understanding the nature of object representations, characterizing objects as either 3D-part-relationships (e.g. Biederman, 1987), or as image-based templates (e.g. Tarr et al., 1998). More recently, cognitive neuroscience approaches have been increasingly important for understanding object representation, but present an interestingly divided view. A few object categories drive responses in a spatially-contiguous and functionally-specific module of cortex along the ventral surface of cortex, indicating category-specific object representations (Kanwisher, 2010; Downing et al., 2006). However, most object categories drive responses in a large swath of this cortex to varying degrees, suggesting more category-general coding dimensions (e.g. Ishai et al., 1999; Haxby et al., 2001). The patterns of activity to different object categories are reliable even across subjects (Shinkareva et al., 2008), which suggests that there is some underlying organization

Figure 1: Real-world objects have a range of physical sizes, from objects you hold in two fingers to objects that you can walk around.

to object representation that we do not understand (Op de Beeck, 2008a, see also Freeman et al., 2011), and to date, there is no consensus on the features of visual object representations (Kourtzi & Conner, 2011; Cavanagh, 2011).

One reason why we have yet to understand the overarching organization of object representation is that it is not clear how to parameterize all of object knowledge. There seems to be no continuous parameter that applies across all objects—instead, objects have more typically been characterized using binary features (is it animate or inanimate? Is a face or not?). However, one often overlooked and intrinsic property of objects, that applies to every object in the world, is its physical *size* (Figure 1). The real-world size of objects dictates how we interact with them, and fundamentally shapes the distributions of our visual experience. My thesis is that real-world size is a basic and fundamental property of object representation, with consequences for both the nature of object representations and their organization in occipito-temporal cortex.

Here I briefly review research that focuses on the nature of object representations, and summarize our current state of understanding about how object representations are organized in the brain. Next I discuss why the real-world size of objects may be an important dimension for object representation, and review what is currently known about the role of real-world size in object representation. Finally I outline the contributions of this thesis, which presents both behavioral and neural evidence that real-world size is a basic property of object representation.

Models of Object Representation

There are two classic theories regarding the underlying nature of object representations. Structural accounts claim that early transformations of the visual system parse the world into simple geometric forms (e.g. geons), and objects are subsequently represented as a part-relations among simple 3D shapes (e.g. A suitcase is a thin rectangular solid with a curved cylinder on top; Biederman, 1987). View-based accounts claim that visual experience of the world is stored as a series of images or views, and objects are represented by prototypical image templates (e.g. Ashbridge & Perret, 1998). Distinguishing between these two accounts with behavioral measures has proven to be empirically challenging, as demonstrations of view-dependent processes do not necessarily imply view-based representation, nor do demonstrations of view-invariant processes necessarily imply structural representations (e.g. Bar, 2001).

More recent approaches have tried to gain insight into the nature of high-level object representations by focusing on the learning process. For example, Schyns and colleagues have argued that as we learn new kinds of objects, we have to learn which features will distinguish them from other objects (termed "functional features"); thus the underlying features of object representation will be those that are in service of categorization (e.g. Goldstone, Lippa, & Shiffrin, 2001; Schyns, Goldstone, & Thibaut, 1998; Schyns & Rodet, 1997). This resonates with recent computational approaches, which accomplish object recognition by using a set of hierarchal image-based fragments (e.g. Ullman, 2007). For example, in this model proposed by Ullman, small image fragments of car parts combine to make larger car fragments, which further combine to make a car. In this model, the features are learned for a particular category (e.g. fragments that help recognize a car), as opposed to category-general features (like geons).

In general, modern models of object recognition rely on a hierarchy of learned features, ranging from object-generic perceptual features like color and orientation to mid-level features that have some specificity to particular object classes (e.g., Ullman, Vidal-Naquet, & Sali, 2002) to very high-level conceptual features that are entirely object category-specific (e.g., Ullman, 2007). These models largely include a computational efficiency constraint, such that the learned features are efficiently represent the visual input. Such models derive a range of features, some of which are shared across multiple object categories while others are more category-specific (e.g., Epshtein & Ullman, 2005; Ommer & Buhmann, 2010; Riesenhuber & Poggio, 1999; Torralba, Murphy, & Freeman, 2004; Ullman, 2007). At the end of this thesis, I will propose that the size of objects gives rise to systematic biases in visual experience which may be extracted by efficient learning mechanisms in the visual hierarchy: specifically, there may be mid-level precursor object representations that capture features shared across objects of the same real-world size.

Organization of object representations in cortex

Object-responsive cortex is found along the ventral and lateral surfaces of the occipital and temporal lobes (Grill-Spector & Malach, 2004; Milner & Goodale, 1995; Ungerleider et al., 1982). Patients with lesions to these ventral temporal areas cannot recognize visually-presented objects though they can see that something is there (for review see Mahon & Caramazza, 2009), indicating that this large swath of cortex is the site of our stored visual object knowledge. What has the neural characterization of object representation in the brain revealed about object representation, especially with respect to the the different models described above?

Functional neuroimaging experiments have demonstrated that, within this cortex, there are a few spatially clustered regions that show responses only for a specific category. Specifically, this is true for faces, bodies, scenes, and letter strings (Cohen et al., 2000; Downing et al., 2001; Kanwisher et al., 1997; McCarthy et al., 1997). These kinds of representations are expected from the computational models that propose category-specific features at the top of the feature hierarchy. However, not all objects have a focal patch of category-selective cortex—e.g. there is no such region for chairs or for shoes. Instead most object categories drive responses in a large amount of ventral temporal cortex, but to varying degrees. Such distributed response profiles are more consistent with the models that learn shared features across object categories, or even with category-general features (e.g. Geons) that can be used to construct any shape (e.g. see Tanaka, 1996). These distributed patterns of activation for each object category are consistent and reliable within a person, across people, and even across species (Carlson et al., 2003; Cox & Savoy, 2003; Haxby et al., 2001; Norman et al., 2006; O'Toole et al., 2005; Shinkareva et al., 2008; Kriegeskorte et al., 2008). Thus, our current state of understanding about the organization of object representation across cortex is that there are islands of category selectivity amongst a sea of heterogenous unorganized (but systematic) responses.

Interestingly, approaches to understanding these patterns of activity have largely ignored how they are arranged spatially across the cortex. However, we know that the spatial organization of information across cortex is far from random (Kass, 1997). This is evidenced in primary sensory cortices most clearly, with body maps, motor maps, tonotopic maps, and retinotopic maps. In visual cortex these maps are mirrored—where visual areas are aligned by eccentricity and smoothly flip along polar angle between vertical and horizontal meridian (e.g. Wandell, Dumoulin, & Brewer, 2007). Beyond the sensory areas, there is further evidence of potentially meaningful proximity, with face-selective regions adjacent to body-selective regions (e.g. Schwarzlose et al., 2008), and partial overlap between neural regions responsive to tools, hands, and motion (e.g. Beauchamp et al., 2002). Assuming that cortex has meaningful topographic representations, and given that people show similar distributed activation patterns for objects, this suggests there are organizing dimensions of object representation that we have not discovered (Op de Beeck, 2008a).

The size of objects in the natural world

The nature and organization of object representation has been approached through a number of different frameworks. However, a fundamental observation about objects is that they are physical entities in a three-dimensional world. Our experience with objects, both in our life time and over evolutionary time, arises as we move through the world. The physical size of objects in the world thus has a dramatic impact on our experience of them, both (i) in how we coordinate our motor actions with objects, and (ii) how object information projects onto the retina. Both of these consequences of object size may have an impact on how object knowledge is organized across the cortex.

First, the real-world size of objects is a fundamental dimension for active visual experience. When entering a movie theatre or attending a lecture, how do we choose the best seat? When making this decision, we actively maneuver to place the visual information of interest into a particular part of our visual field, adjusting our angle and distance to an object like the screen or podium, based on its size in the world. More generally, object information arrives in retinotopic coordinate frames, and if we want to act on this object we have to transform the information for the relevant effector, be it hand-, head-, or body-centered coordinates (e.g. see Cohen & Andersen, 2002). The real-world size of the object reduces the complexity of this mapping problem: not all object are equally relevant for all reference frames. Small objects like keys or paperclips require finger-based action plans, while chairs require body-coordinate frames, and the exact position of the fingers with respect to the chair is largely irrelevant (Figure 1). Thus, demands of action systems on visual processes may place top-down organizing constrains on object knowledge in the ventral pathway such that it is computationally efficient to be grouped by real-world size.

Second, there are systematic biases in visual experience driven by the size of objects in the world–in retinal size, eccentricity, height in visual field, head-angle, shape, and spatial frequency. Due to the geometric structure of the world, objects of different sizes are interacted with at different distances, and thus gives rise to systematic distributions of retinal size (Figure 2). A peanut at arms length subtends 3 degrees visual angle, while a car at a typical distance subtends 30 degrees visual angle. Given that our eyes view the world from a height off the horizon, big objects will also tend to be higher up in our visual field than small objects. Further, small objects tend to be shaped for the hand and are rounder whereas larger objects withstand gravity and provide structural support and are boxier: these differences in shape statistics can

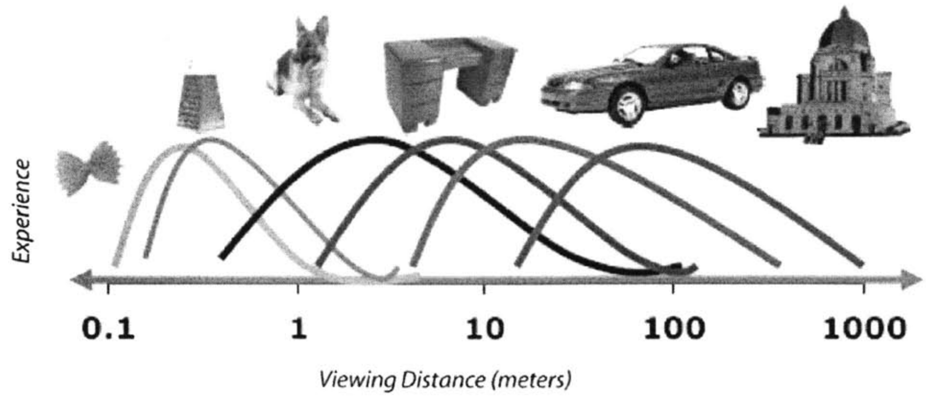


Figure 2: Hypothetical distributions of visual experience based on typical interaction distances. Objects of different physical sizes in the world have a distribution of visual experience over viewing distances. Estimates of typical interaction distances for objects indicate that objects of increasing physical sizes are experienced at increasing retinal sizes (see also Hubbard, Kall, & Baird, 1989).

be measured by combinations of image measurements that capture power at different spatial frequencies and scales and positions across the image (see Oliva & Torralba, 2001; see also Haldane, 1928). Head angle is another dimension that is affected by object size, where we tend to look down to small objects in our hand, progressively up to big objects at the horizon. In the appendix at end of the thesis, I explain these visual biases in greater detail and suggest that they can explain a number of low-level response properties that have been observed in high-level visual areas.

Taking these visual biases all together, a consequence of natural visual inputs is that objects of different sizes will have systematically different low-level statistics arriving in early visual cortex. If visual systems are tuned to efficiently encode visual input by extracting covariances (Attneave, 1954; Carlson et al., 2011; Field, 1987), then these low-level correlated statistics are viable candidates. This would lead to mid-level visual representations that apply to all objects of a particular real-world size. Due to the way eccentricity is laid out across cortex, these mid-level visual representations, in essence pre-cursor high-level object representations, may naturally be arrayed by real-world size in more anterior visual cortex.

Previous research on real-world size

Up to this point I have been referring to "visual object representations", which are the representations along the visual processing stream that enable recognition. In the broader study of memory systems, this is referred to as the perceptual representation system, as distinct from the semantic system (e.g. Schacter & Tulving, 1994). The former concerns visual form (e.g. what an object looks like, allows us to name an object from a picture), and the latter concerns associated semantic facts (e.g. can move on it's own, is edible, is big, is a cow). Interestingly, information about the real-world size of objects straddles this divide, and research has been done in both areas.

Semantic information about real-world size is exemplified by a fact-based knowledge: just as you can know the capital of Colorado is Denver, so you can know that a building is 152 ft tall, or that an elephant is bigger than a mouse. However, given the task to say which of two named objects is bigger in the world, the time it takes is related to the log of their real-world size ratios; this is a classic signature of perceptual comparisons, e.g. indicating which of two lines is longer (Moyer, 1973; Pavio, 1975; Rubinsten & Henik, 2002; Srinivas, 1996). This has led some to argue that real-world size knowledge may actually be stored in some analog or perceptual format that preserves real-world size, and suggests that real-world size may not have a purely semantic (non-perceptual) representation. However, understanding "size" as a semantic fact is outside the scope of this thesis. Rather, here I focus on if, and how, real-world size influences the perceptual representation system.

Surprisingly, very few studies on the perceptual representation system, i.e visual object representation, have focused on the dimension of real-world size. In Biederman's discussion of scene schemas, objects presented at the wrong real-world size in the scene are one of the 5 scene violations: objects were detected more slowly when presented at an atypical sized were detected more slowly than objects at a typical size in the scene context (Biederman, et al., 1982). Knowledge about the real-world size of objects ("familiar size") can serve as a cue to depth, but the nature of this familiar size knowledge has not been directly examined. In contrast, there have been many studies examining the dimension of *visual size* in object representation (Biederman & Cooper, 1992; Fiser & Biederman, 1995; Cave & Squire, 1992; Srinivas, 1996; Furmanski & Engel, 2000; Jolicoeur, 1987; Millikan & Jolicoeur, 1992). Here, "visual size" refers to the visual angle of the object projected to the retina. One potential reason for this focus on visual size in perceptual representation is that visual size is the input to the visual system, whereas physical size (and viewing distance) have to be inferred from the retinal projection. Further, in theory an object of any physical size can subtend any visual angle, by moving closer or farther away. However, as previously discussed, typical viewing distances are constrained by geometry and interaction. This over-simplification has made it easy to miss the impact of real-world size on visual experience with consequences for object representations.

Contributions of this thesis

My thesis is that real-world size is a basic and fundamental property of object representation, with implications for the object representations and their organization in occipito-temporal cortex.

Chapter 2 examines how visual size information is represented in our existing object representations. In a series of experiments which required observers to access existing object knowledge, we observed that familiar objects have a consistent visual size at which they are drawn, imagined, and preferentially viewed. This visual size was not the same for all objects, but was instead proportional to the logarithm of the known size of the object in the world. Akin to the previous literature on canonical perspective (Palmer, Rosch, & Chase, 1981), we term this consistent visual size information the *canonical visual size.*

Chapter 3 presents evidence that real-world size is an automatically-accessed property of object representations. In this study, two real-world objects were presented at different visual sizes observers had to indicate which was bigger (or smaller) *on the screen.* Even though the known size of the objects was irrelevant for this task, we observed a familiar size stroop effect. A second experiment demonstrated that this effect was not cognitively penetrable, implying that real-world size knowledge is a part of visual representations acquired with repeated experience.

Chapter 4 presents neuroimaging data, and examines the impact of real-world size on the spatial distribution of neural representations for objects without selective regions of cortex. In a series of studies, I demonstrate that the representation of objects in ventral temporal cortex depends on their real-world size, and suggest there is a large-scale mirrored organization across ventral and lateral occipito-temporal cortex based on real-world size.

Finally, Chapter 5 summarizes and interprets these empirical findings within a situated-observer framework. The object representations studied here were learned over a life time; visual experience was not manipulated in any of these studies. However, I suggest that these behavioral and neural data can be coherently interpreted in framework which takes into account the visual experience of a situated observer in the three-dimensional world and assumes that the visual system is tuned to statistical regularities of experience. I also describe areas of future research where this situated-observer framework can provide insight into the underlying representations of objects and can make testable predictions about the neural representation of any object.

Chapter 2

Canonical visual size for real-world objects[1]

Real-world objects can be viewed at a range of distances and thus can be experienced at a range of visual angles within the visual field. Given the large amount of visual size variation possible when observing objects, we examined how internal object representations represent visual size information. In a series of experiments which required observers to access existing object knowledge, we observed that real-world objects have a consistent visual size at which they are drawn, imagined, and preferentially viewed. Importantly, this visual size is proportional to the logarithm of the assumed size of the object in the world, and is best characterized not as a fixed visual angle, but by the ratio of the object and the frame of space around it. Akin to the previous literature on canonical perspective, we term this consistent visual size information the *canonical visual size.*

Introduction

In the real world, the particular view of an object (i.e., its projected retinal image) depends on where the observer is standing with respect to that object. This fact is implicitly understood by observers choosing where to sit in a movie theatre, where to stand in an art gallery, or where to move to get a better view of an item of interest. When observers walk around an object, changing the viewing angle of an object without changing its distance, this image transformation is called a perspective

[1]This chapter was published as Konkle, T. & Oliva, A. (2011). Canonical visual size for real-world objects. *Journal of Experimental Psychology: Human Perception and Performance, 37*(1):23-37.

change. Similarly, when observers approach or back away from an object to change its retinal size within their visual field without changing the viewing angle, the image transformation is called a visual size change. Given the many possible object views that can be experienced by an observer, what information about perspective and size is present in object representations?

Seminal research by Palmer, Rosch, and Chase (1981) examined how object viewpoint information was accessed in a number of different tasks, and found evidence for consistently preferred viewpoints. For example, during goodness judgments of photographs of objects over different viewpoints, three-quarter perspectives—in which the front, side, and top surfaces were visually present—were usually ranked highest[2]. The best" view was also the perspective imagined when given the name of the object, the view most photographed, and enabled fastest naming of objects. The consistencies across observers and across tasks led Palmer, Rosch, and Chase (1981) to term this view the "canonical perspective."

Two main explanations have been suggested for why objects have a preferred, canonical perspective. One account is motivated by object properties, where the canonical perspective maximizes surface information visible with the least degree of self-occlusion. The other account argues that canonical perspective arises based on the distribution of visual experience. Evidence for the latter involves studies that control exposure with novel objects, and find speeded recognition arises at more-often experienced viewpoints (e.g. Bulthoff & Edelman, 1992; Tarr, 1995; Tarr & Pinker, 1989). However, canonical viewpoints can be found for novel objects that have been experienced equally from all angles in the viewing sphere (Edelman & Bulthoff, 1992), suggesting that a purely experiential account cannot fully predict the occurrence of canonical viewpoints. These explanations for canonical viewpoints reflect a trade-off between constraints of object-centered properties, where shape and orientation determines the best viewing angle, and viewer-centered properties, where accumulated episodes with that object influence the preferred viewing angle. Likely

[2]There were a few objects for which this was not true, such as a clock, for which a pure front view was ranked highest, probably due to the frequency with which it is viewed in this perspective.

both of these factors contribute to canonical perspective (Blanz, Tarr, & Bulthoff, 1999).

Perspective is determined by the physical orientation of the object relative to the direction of gaze of the observer. Similarly, visual size is determined by the physical size of the object relative to the distance of the observer to the object. Given that there is evidence for canonical perspective, is there similar evidence for canonical visual size? Here, we employed memory, imagery, and perceptual preference tasks and asked whether these different mental processes yield consistent visual sizes across observers. In addition, we examined the contributions of two factors that might influence an object's canonical visual size—real-world size and framing.

First, we might expect knowledge about the real-world size of the object to matter for an object's canonical size. Intuitively, smaller objects in the world subtend smaller visual angles on average than larger objects in the world. For example, a typically sized car would subtend about 30 degrees visual angle at a typical viewing distance of ~9 m. For a penny to subtend that same visual angle it would have to be held only ~3 cm away from one eye; at a more typical arms-length viewing distance, it subtends 3.5 degrees. Thus, natural experience with objects might predict a systematic relationship between real-world size and canonical visual size. Alternatively, maximizing the available object information could determine canonical size, e.g. if the object is centered in the high-acuity foveal or parafoveal region of the visual field. Such an account might predict that all objects would have the same canonical visual size that is related to acuity falloff with eccentricity, possibly modulated by the internal complexity of the surfaces features of the object.

Second, size judgments are strongly influenced by the relative size of an object within a fixed frame of space. In typical real-world viewing situations, a chair looks the same physical size as we approach it, despite the increasing visual size it projects on the retina—a phenomenon known as size constancy. However, failures of size constancy can be found when the frame of space around an object is manipulated. For example, Rock and Ebenholtz (1959) had observers adjust the length of one line to match the length of a standard line. The standard line was framed in a small rectangle, while the adjustable line was framed in a larger rectangle. Observers were strongly biased to preserve the ratio of the line within the frame, adjusting the line to be much larger than the standard, even though the task was to match the physical length of the two lines (see also Kunnapas, 1955).

This framing effect occurs not only for simple stimuli but also for objects in the real world, and is known as the vista paradox (Walker, Rupick, & Powell, 1989; see also Brigell, 1977; Senders, 1966). Approaching an object makes it physically closer, but approaching that object through the view of a window creates an illusion that the object is both shrinking in physical size and getting farther away. On the retina, both the visual size of the frame and the visual size of the object increase as one approaches; however, the ratio of the object in the frame decreases because the frame grows much more quickly than the more distant object. This illusion demonstrates that our perception of an object's physical size and distance away are subject to relative framing ratios, and are not derived from visual angle alone.

In the current experiments, we examined whether or not existing object representations show evidence for a canonical visual size. Using a drawing task (Experiment 1), an imagery task (Experiment 2), and a perception task (Experiments 3, 4, and 5), we found that all these tasks gave rise to consistent visual sizes across observers and mental processes. We also observed a systematic and reliable correlation between canonical visual size of objects and the logarithm of their assumed size in the world. Further, we demonstrate that this canonical visual size is best characterized not as a fixed visual angle, but as a ratio reflecting the object size relative to the frame of space within which it is viewed.

Size Ranking

Observers have prior knowledge about the size of objects in the world, often referred to as "assumed size" (e.g. Ittleson, 1951; Baird, 1963; Epstein 1963). In the following experiments, we aimed to assess whether the assumed size of objects influences the visual size at which objects are accessed across different tasks. Thus, first we gathered

100 images of real-world objects and had observers sort these objects into 8 groups of increasing real-world size. These data will give us size ranks that reflect the assumed size of objects in the real world, and will be used in the rest of the experiments. The object images spanned the range of real-world sizes from small objects (e.g. a paper clip) to large objects (e.g. the Eiffel Tower; see Figure 1). Additionally, we examined how the size ranks compared with the actual real-world size of such objects.

Methods

Six observers (age range 18-35) gave informed consent and received $5 for their participation. One hundred color pictures of real-world objects were selected from a commercial database (Hemera Photo-Objects, Vol. I & II), and all objects appeared on a white background (see Figure 1). The sorting procedure was adopted from Oliva and Torralba, 2001. Thumbnails of 100 objects were arrayed on a 30 in (64.5 x 40.5 cm) screen, with a line separating the left and right half of the screen. Participants were instructed to drag and drop the objects so that the large objects (large in their real-world size) were on one half of the screen and the small objects (small in real-world size) were on the other half of the screen. Next, the screen divided into fourths, and participants refined the two sets of objects into four groups. This processes repeated one more time so that the objects were divided into 8 groups, ranked by their size in the real world. Here, a rank of 1 represents the smallest object size and a rank of 8 represents the largest object size. Participants were told that they did not have to have an equal number of objects in each group and that instead they should make sure each category of objects had roughly the same physical size in the world. Participants could double click on a thumbnail to view a larger image of that object (15cm x 15cm). Stimuli were presented using software written in MATLAB.

Observers were instructed to sort objects based on their real-world size", and we did not explicitly instruct observers how to think of real-world size (e.g. volume, area, extent). To obtain a measure of the "actual size" of each depicted object, we used the following procedure. For each image a corresponding real-world object was measured or approximated. In the case of the larger objects, the dimensions were found using

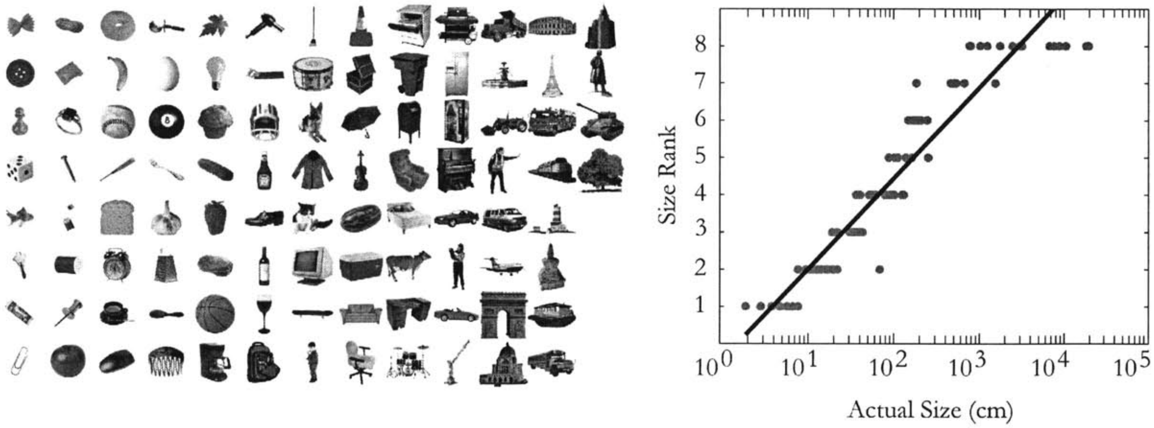


Figure 1: Left: Database of 100 objects. Right: All objects were sorted into 8 groups based on their assumed size in the world. These ranks are plotted as a function of the actual real-world size of the object (cm), on a logarithmic scale. The graph shows is a systematic logarithmic relationship between the actual physical size of the object and the size ranks.

internet searches. The actual size of the object was quantified in cm (rather than cm^3), measured as the diagonal of its bounding box (i.e., the smallest rectangle that completely enclosed the object), ignoring the depth of the object[3].

Results

The left panel of Figure 1 shows thumbnails of the object set. We defined the size rank of each object as the mode of its rank distribution over the six observers. There were 9 to 23 objects for each size rank (mean 13 objects/size rank). Next we examined the relationship between the size ranks and the actual size of such objects in the world. The right panel shows the actual size of each object, plotted as a function of its size rank, with the actual size plotted on a logarithmic axis. The graph shows that size ranks and actual size are related by a logarthmic function. The correlation between size rank and log10(actual size) is $r^2 = .91, p < .001$.

These results suggest that when sorting objects by assumed size, judgments about which sizes are similar follow Weber-Fechner-like scaling (as do judgments about most

[3]The actual size of the object could also be quantified as the diagonal of the three-dimensional bounding box (height x width x depth). Because of the correlation between height, width, and depth of these objects, the 3d diagonal and the frontal diagonal are negligibly different on a log scale.

other psychophysical variables, e.g. weight, sound intensity, frequency, etc; Stevens, 1957). For example, two objects at 1m and 10m in size are more different that two objects at 1001m and 1010m. Similar ranking procedures and results were found by Paivio (1975) and Moyer (1975).

These size ranks formed 8 groups of objects which were used in subsequent experiments. While we could use the actual size measured from real-world objects, the size ranks are used because (i) they reflect empirically gathered data about assumed size, and (ii) provide natural bins of the assumed size dimension. However, it should be noted that the size rank reflects a logarithmic scaling of real-world size, thus any systematic relationship found with size rank also shows a similar systematic relationship with the logarithm of the real-world size of the object.

Experiment 1: Drawings from Memory

In Experiment 1, we used a drawing task to probe existing object representations, which is a task that requires reconstruction from long-term memory. Similar tasks have been used for studies of visual memory, but have typically been used as a measure of visual free recall of a previously studied image (e.g. Carmichael, Hogan, & Walters, 1932; Intraub & Richardson, 1989). Here, we instead probed pre-existing long-term memory representations. The observers' task was simply to draw a picture of the named object on the page. Unbeknownst to the observers, we were interested in the size at which they drew these objects.

One possibility is that all objects would be drawn at the same size on the page (or at the same visual angle). This might be predicted by classic alignment models of object recognition, which assume that all objects are stored at a specified visual size in memory, and recognition proceeds by first mentally scaling the input or the fixed template (e.g., Ullman, 1989). Another possibility is that there will simply be no consistent relationship between the drawn size of objects and the assumed size of those objects. Alternatively, there may be a systematic relationship between drawn size and assumed size, where a number of quantitative relationships are possible.

Importantly, the task of drawing objects does not require explicit reasoning about the assumed size of the object nor does it require making judgments about the drawn size.

We also examined the role of the frame of space in which the object was drawn by manipulating the paper size across observers. If the frame serves as a ceiling for drawn object sizes, then we might predict that the physically small objects would be drawn the same size across paper sizes, but the physically larger objects would be drawn increasingly larger with bigger paper sizes. However, another possibility is that objects might be drawn with a consistent ratio of the object to the frame across paper sizes. This might be predicted if object representations are reactivated from long-term memory representations relative to a space around them.

Methods

Sixty-four naïve observers (age range 18-35) participated in Experiment 1. All gave informed consent and received a candy bar and a beverage for their participation. Twenty observers drew on the small paper size, 22 observers draw on the medium paper size, and 22 observers drew on the large paper size. Participants sat at a table and were given 18 sheets of paper (all of the same size) and a list of items to draw. They were instructed to draw one object per page and were explicitly told that we were not interested in artistic skills. We told participants to draw each object relatively quickly (within 1 minute). When delivering the instructions, the word "size" was never used.

The list of items contained 16 different objects that spanned the range of real-world sizes, with two objects at each size rank. The objects were: paperclip, key, pet goldfish, apple, hairdryer, running shoe, backpack, computer monitor, German shepherd, chair, floor lamp, soda machine, car, dump truck, 1-story house, light house. The order of objects was randomized for each observer. After all 16 objects had been drawn, observers next drew two scenes, a beach and a park, in random order.

Across observers, we manipulated the size of the drawing paper. Observers were not aware of this manipulation. The small paper size was 7.6 x 11.4 cm (3 x 4.5

inches), the medium size 18.5 x 27.9 cm (was 7.3 x 11 inches), and the large size was 30.5 x 45.7 cm (12 x 18 inches), thus all three sizes had approximately the same aspect ratio. All observers used a fine black sharpee marker to draw (i.e., the pen width was fixed, and did not scale with the paper size).

To measure the drawn size of the objects, all drawings were scanned at a fixed resolution (150 dots per inch). Custom software was written in MATLAB to automatically find the bounding box around the object in the image, and these dimensions were converted from pixels into centimeters using the known resolution. Drawn size was calculated as the length of the diagonal of the bounding box around the object. Using the diagonal, rather than as the height or width alone, better takes into account variation in aspect ratio and has been shown to account for more explained variance in relative size measures than height, width, principle axis, and area (Kosslyn, 1987). The software proceeded one drawing at a time, and each object's identity and the corresponding bounding box was verified by eye.

Results

The first author and one additional observer used a strict criterion to filter any drawings with extraneous objects (e.g. trash bins behind the dump truck, a worm sticking out of the apple, cords connecting the floor lamps, headlight beams on cars, air coming out of the hairdryer), which constituted 21% of the images. The analysis reported below was conducted on the filtered data set (887 drawings)[4].

Figure 2 (left panel) shows the drawn size of the objects (in cm) plotted as a function of the size rank of the object. The three lines represent the three different paper sizes. A two-way ANOVA was conducted on drawn size with paper size as a between-subject factor and object size rank as a within-subject factor. There was a significant main effect of the size rank of the object on the drawn size of the object ($F(7, 391) = 30.1, p < 0.001, \eta_p^2 = .35$). That is, objects that are small in the world

[4]The patterns in the data are unchanged when the analysis is conducted on drawn images using a more moderate exclusion criteria (connected objects such as worms and wires included) or with full inclusion (including the trash cans behind the dumptruck).

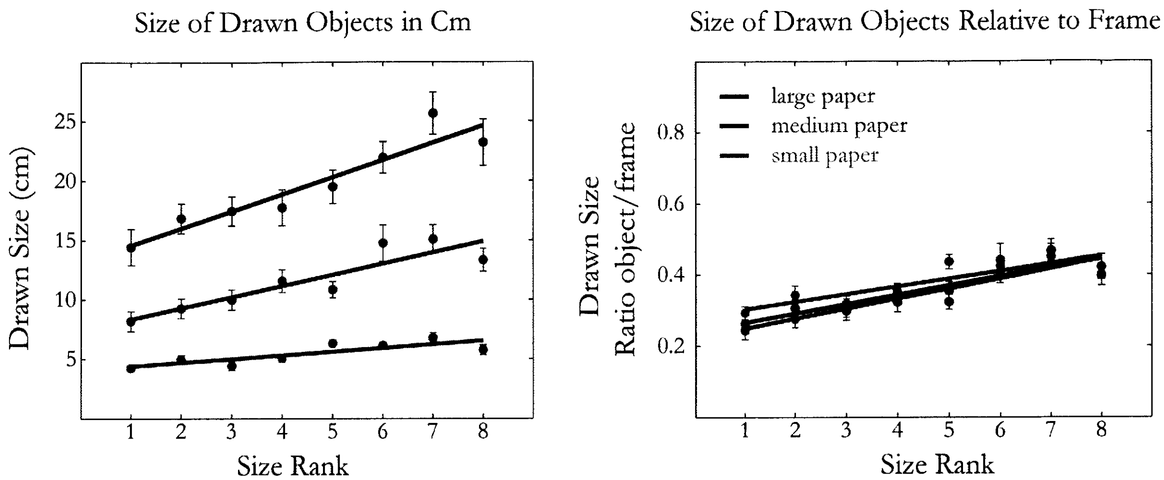


Figure 2: Left: Drawn size of objects (measured in centimeters) as a function of their size rank, for small, medium, and large paper sizes. Right: Drawn size of objects (measured as the ratio of the drawn object and paper diagonal length), as a function of size rank for small, medium, and large paper size. There was a separate group of observers for each paper size. Error bars represent ±1 S.E.M.

were drawn smaller on the page than objects that are large in the world. There was also a significant effect of paper size on drawn size ($F(2, 41) = 70.9, p < 0.001, \eta_p^2 = .78$), where the average drawn size of objects increased as the paper size increased. Additionally, there was a significant interaction between the paper size and the effect of the object size ($F(14, 391) = 4.3, p < 0.001, \eta_p^2 = .13$). In other words, there was a smaller range of drawn object sizes on the small paper, with progressively greater ranges of drawn sizes on the medium and large paper.

These data show a clear linear relationship between the drawn size and the size rank ($r^2 = 0.88, p < 0.001$, collapsing across paper size). Thus, this also demonstrates that the drawn size of an object is proportional to the logarithm of its real-world size. For each participant, a regression analysis was used to estimate a slope and intercept for their drawn sizes as a function of the size rank. ANOVAs were conducted on these slopes and intercepts, with paper-size as a between-subject factor. There was a significant effect of paper size on slope ($F(2, 61) = 28.7, p < 0.001, \eta^2 = .48$), and a significant effect of paper size on intercept ($F(2, 61) = 15.1, p < 0.001, \eta^2 = .33$).

Across the 16 objects, the systematic variation in the drawn object sizes was highly consistent. The effective reliability R, which is the aggregate reliability from a set of judges (see Rosenthal & Rosnow, 1991) was R=.97.

Figure 2 (right panel) contains the same data as in the left panel, replotted to show the ratio of the drawn size of the object to the paper size. This was calculated as the diagonal length of the drawing divided by the diagonal length of the paper size. When considering the drawn size ratio, there was no longer an effect of paper size ($F(2, 41) = 1.01, n.s.$), whereas size rank still significantly influenced the drawn size of the object in the frame ($F(7, 391) = 38.19, p < 0.001, \eta_p^2 = .41$). However, there was a small, but significant, interaction between paper size and size rank ($F(14, 391) = 1.82, p < 0.05, \eta_p^2 = .06$), which indicates that some of the items had a slightly different ratio from small to medium to large paper sizes. For example, the smallest objects drawn on the smallest paper size show slightly larger ratios than for the medium or large paper. One possible explanation is that because all observers used the same sharpee marker for drawing across paper size, they may have drawn the smallest objects on the small paper size somewhat larger than on the larger paper sizes. Separate ANOVAs conducted on the single subject regression fits revealed no difference between the slopes across paper sizes ($F(2, 61) = 2.1, n.s.$), nor any difference between the intercepts ($F(2, 61) < 1, n.s.$). Figure 3 shows example drawings, both to scale and with normalized paper sizes.

Discussion

When observers are instructed to draw an object from an existing representation in visual long-term memory, the drawn size of the object depends on at least two factors. First, the drawn size of the object depends on the assumed size of the object in the world. Small objects in the world are drawn small on the page; large objects in the world are drawn larger on the page. Further, this relationship is systematic: the drawn size of an object is proportional to the size rank (and thus to the logarithm of its actual real-world size). Second, the drawn size of the object depends on the scale of the space it can occupy. Small objects such as a keys occupied 27% of the image (as measured by the diagonal of their bounding boxes relative to the diagonal of the paper), whereas large objects like houses occupied 41%. Critically, the raw

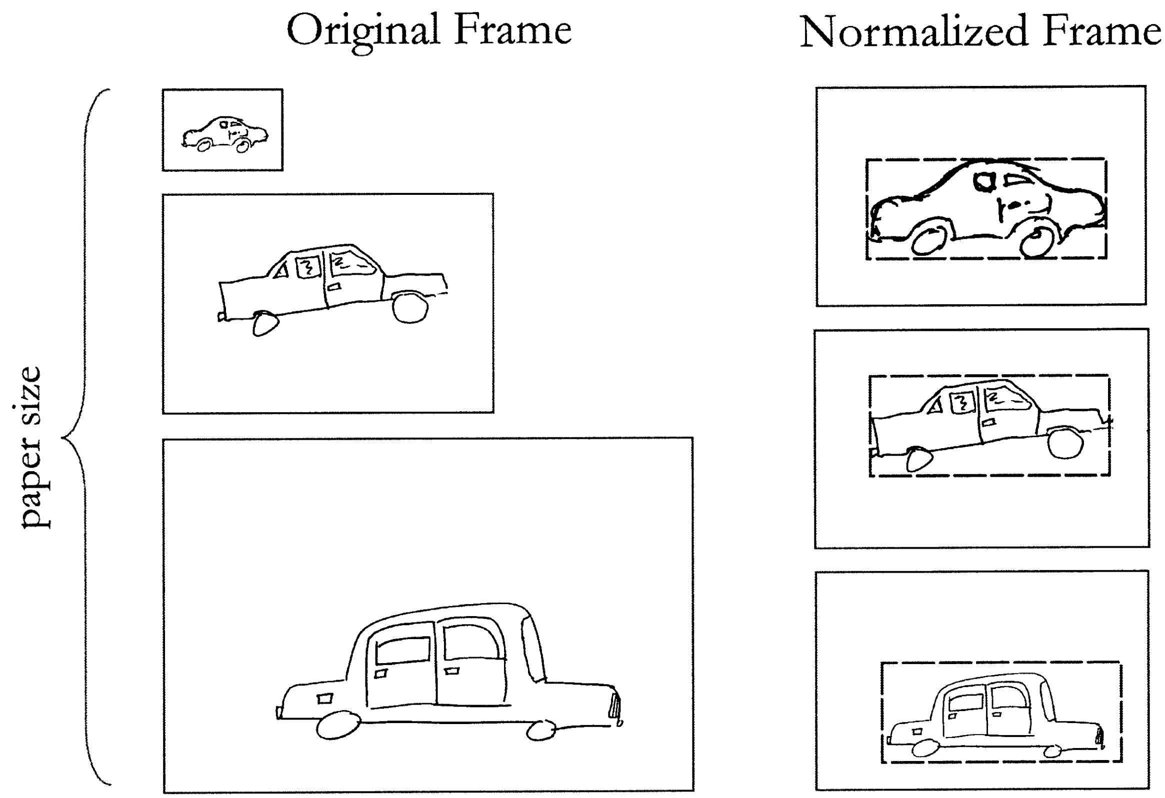


Figure 3: Example drawings of a car from three separate participants. Upper: small, medium, and large drawings, to scale. Lower: the same drawings, normalized to the size of the frame. The dashed bounding box is the same size in all three normalized drawings for reference.

size at which objects were drawn (and thus the visual angle which the drawn images subtended in an observer's visual field) were very different for the small, medium, and large paper sizes, whereas the ratio of the object within the frame was constant across paper sizes. This strongly suggests that when objects are reconstructed from memory, the drawn size is best characterized not by raw visual angle measurements but as a relative proportion between the object and a frame of space.

It is interesting that observers did not fill the page, even for the objects with the largest size, which were only about 40% of the scale of the frame. That is, observers preserved space around the edges of the objects, even on the smallest paper sizes. However, it is not the case that observers always leave blank space around all drawings–when observers drew a beach scene and a park scene, which do not necessarily have a clear edges as do objects, the average drawn size was 81% of the frame (SEM 1.8%)[5]. Further, this preserved ratio of the object and the frame is especially striking when considering the drawings of the small objects on the large paper sizes. In this condition, a paperclip was drawn on average 14.0 cm on the large paper (SEM 1.6 cm), which is dramatically larger than its actual size in the world (~3-5 cm). Thus, one intriguing possibility is that internal object representations contain information about the relative visual size of objects and a spatial envelope around them. For example, when drawing an object, the object is not scaled to the paper; rather, the object and its envelope are scaled to the paper. A representation of this kind would produce consistent ratios across different frame sizes.

An important open question is whether the observed relationship between the drawn size and the assumed size of objects reflects a conceptual (non-visual) bias or a perceptual (visual) bias. In other words, are these results driven by explicit knowledge that, for example, cars are typically 5 m long? Whereas semantic (non-visual) knowledge of an object's physical size likely plays a role, several points suggest that there is also a strong visual component. First, the relationship between assumed size and

[5]In these scenes, the calculated diagonal ratio was not 100% because observers typically drew a horizon line which extended across the entire horizontal axis, but did not necessarily make marks for grass/sand that touched the extreme bottom edge and for clouds/sun/trees that touched the extreme top edge of the paper.

drawn size is systematically logarithmic, which is a classic quantitative relationship between perceptual properties and physical stimulus properties (e.g. Weber-Fechner's law; see also Moyer, 1975). Second, this adjustment of drawing small objects smaller and large objects larger was not the same across paper sizes; the range of drawn sizes on the large paper was 11.2 cm, with only a 6.9 cm range for the medium paper and only a 2.5 cm range for the small paper. However, when normalized by the frame, the ratios of the object to the paper size were remarkably consistent. Although this does not rule out a purely conceptual (non-visual) representation driving these results, it is unclear why explicit knowledge of the physical size would be influenced by a frame, whereas it is known that perceptual tasks (e.g., adjusting the physical size of a line in a frame) are biased by framing ratios (Rock & Ebenholtz, 1959). Neither the current study, nor the subsequent studies can adequately answer the question about whether physical size information is represented visually or conceptually, but we believe that both are probably involved (see Hart, Lesser, & Gordon, 1992). The important points for the current study are that object information accessed from long-term memory representations contains visual size information that is consistent across observers, is related to real-world size, and is best characterized as a ratio with respect to the space or frame it occupies.

Experiment 2: Imagery

Here, we used an imagery paradigm to probe size information in existing long-term memory representations of objects. Specifically, we examined the visual size at which objects were imagined within the frame of a computer monitor. Imagery processes can be thought of as instantiating visual long-term memory representations (i.e., stored knowledge about the visual properties of an object or class of objects) in perceptual buffers (see Kosslyn, 1999). Thus mental imagery, like drawing, relies on accessing existing object representations. If observers imagine objects at a size within the frame of the computer screen that matches the size they drew objects relative to the page size, this would show converging evidence using an alternate method of probing existing visual object representations.

Previous work examining the imagined size of real-world objects is consistent with these predictions (Hubbard & Baird, 1988; Hubbard, Kall, & Baird, 1989; Kosslyn, 1978). For example, Kosslyn (1978) used a mental distance estimation procedure to calculate the visual angle at which animals of various sizes were spontaneously imagined. Interestingly, he found that small animals were spontaneously imagined at closer distances than larger animals, and that the visual angle subtended in the mind's eye was positively correlated with the size of the animal. In other words, small animals were imagined at smaller visual angles than large animals. He also noted that observers were not preserving an absolute scaling of the animals' physical size in their mental images, as the largest animals were imagined at less than twice the angle of the smallest animals despite being an order of magnitude bigger in size. Kosslyn's study was aimed at quantifying the extent of the mind's eye and not the relationship between object size and spontaneously imagined size; however, these results provide suggestive evidence that imagined size of objects might show convergent patterns with the drawn size ratio of objects we observed in Experiment 1.

Methods

A separate group of nine naïve observers were recruited from the MIT participant pool (age range 18-35), gave informed consent, and received 5 dollars for their participation. Stimuli were presented using MATLAB with the Psychophysics toolbox extensions (Brainard, 1997; Pelli, 1997).

At the start of each trial, the name of an object appeared at the center of the computer screen. Observers pressed a key to continue, and the screen blanked for 2 seconds. Observers were instructed to form a clear mental image of the object on the screen during that time. After two seconds, the mouse cursor appeared at the center of the screen. As observers moved the mouse, a rectangular box centered on the screen was drawn automatically: one corner of the rectangle was at the current mouse position and the opposite corner of the rectangle was at the same distance from the center of the screen in the opposite direction. Observers adjusted this rectangle by moving the mouse, and then clicked when the rectangle "formed a tight bounding box" around their mental image of that object. After the response, the screen blanked for 2 seconds and the name of the next object appeared. The names of the 100 objects in the object set were displayed in a random order. Observers were given a demonstration of how to adjust the size of the bounding rectangle before the experiment began.

Results

The left panel of Figure 4 shows the average size of imagined real world objects, plotted as a function of object size rank. Here, the imagined size was calculated as the visual angle subtended by the diagonal of the bounding box. The average imagined size for two sample objects—an egg and a refrigerator, is illustrated in the right panel of Figure 4. Averaging over the size rank of objects, observers imagined objects at 15.4 degrees visual angle (S.E.M.=4.5 degrees). Taking into account size rank, there was systematic positive relationship with imagined size (slope = 2.8 degrees/size rank, $r^2 = 0.98, p < 0.001$).

Importantly, across the 100 objects, this systematic variation in the imagined object size was again quite consistent across observers. The effective reliability was R=.96. Thus, despite the subjectivity of the task to simply imagine the object, some objects were consistently imagined smaller and others were consistently imagined larger.

The imagery data can be converted into a ratio between the imagined size and the size of the monitor. This allows for comparison between the drawing data (Experiment 1) and the imagery data. Collapsing across size rank, there was no significant difference in the average imagined ratio and average drawn ratio, (imagery: 34%, SEM 3.1%; drawing: 36%, SEM 1.2%; $t(71) = 0.6, n.s.$). However, the slope between size rank and imagined size was steeper than in the in the drawing study (6.3% per size rank in Experiment 2 vs. 2.7% per size rank in Experiment 1; $t(71) = 6.2, p < 0.0001$).

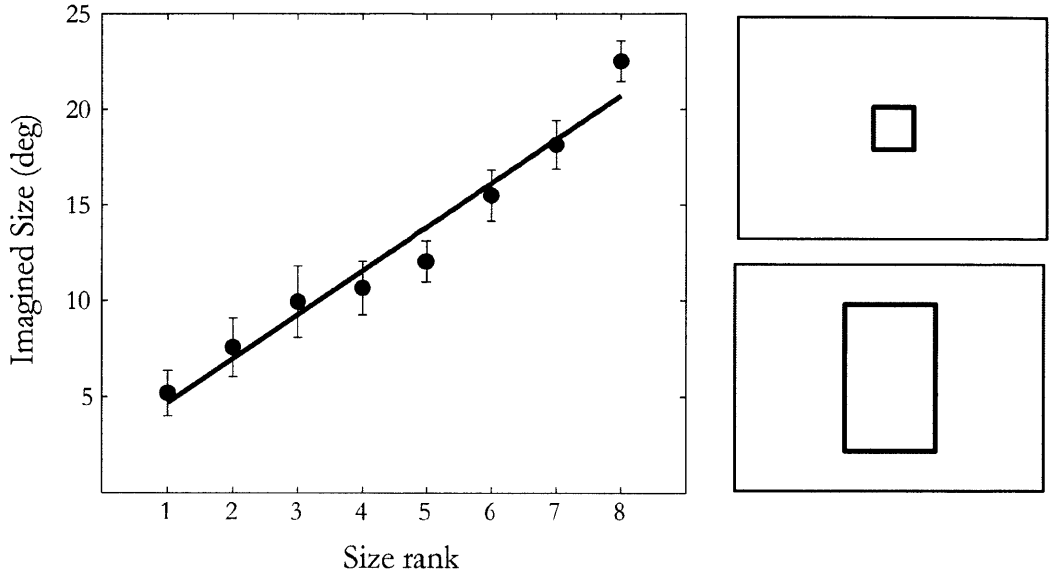


Figure 4: The average imagined size of objects on the computer screen is plotted as a function of the size rank of the object (black line). Error bars represent ±1 S.E.M. The average imagined sizes of two objects (upper: egg, lower: refrigerator) are shown on the right.

Discussion

The data show that the imagined size of real-world objects scales with the assumed size of the object: physically small objects are imagined at smaller visual angles than physically large objects. Further, when the imagined visual size was normalized by the monitor visual size, the resulting ratios were fairly compatible with those found in Experiment 1. Thus, the relative measure between the object and frame holds across different observers, different tasks, and different kinds of frames.

In the imagery experiment, the slope of the relationship between size rank and imagined size was actually steeper than the corresponding slope in Experiment 1. One speculative account of this finding is that during the drawing task, perception of the drawn objects constrains the dynamic range of the drawn size, and imagery processes are not constrained in the same way. Indeed, Kosslyn (1978) found that objects imagined from existing long-term memory were imagined at larger sizes than when pictures of those animals were shown and then subsequently imagined. Further, in his estimation of the "extent of the mind's eye", larger estimates ($\sim$50 degrees) were obtained using imagined objects from existing long-term memory than when the same method was used on images of objects ($\sim$20 degrees).

Hubbard and Baird (1988) extended Kosslyn's study by quantifying the relationship between the physical size of objects and the distance at which they are spontaneously imagined ("first-sight" distance). They found evidence for a power-law relationship between object size and first-sight distance (see also Hubbard, Kall, & Baird, 1989). For comparative purposes, this relationship between first-sight distance and object size can be transformed to reveal the corresponding relationship between the imagined visual angle and object size (as in Figure 4). Interestingly, this relationship is roughly linear with log object size, consistent with our findings.

To assess whether a visible frame is required to drive the relationship between imagined size and assumed size, we ran another imagery experiment in which there was no visual frame. Ten observers were blindfolded and asked to imagine an object (spoken aloud by the experimenter). After observers had formed a mental image, they remained blindfolded and traced a tight bounding box around the object in their mental image on a wall-sized blackboard in front of them. The 16 objects from Experiment 1 were used, and observers were guided to a new part of the blackboard for each object. Here, there was no visually present frame, but we again found a consistent linear relationship with the imagined size and size rank ($r^2 = 0.89, p < 0.001$), with an average slope of 3 deg/rank, though there was much more variability across individual's slopes (min: 1 deg/rank, max: 6.2 deg/rank). Thus, both imagery tasks on a monitor (with a frame) and blindfolded (without a frame) showed reliable and systematic influences of assumed size on the imagined size of real world objects.

Experiment 3: Perception

Experiment 1 and 2 used tasks that require observers to know what objects look like in order to draw and imagine them. In other words, they require retrieval of existing visual object representations. In Experiment 3, observers simply had to view images of real-world objects on the monitor and determine the size at which the objects "looked best." Similar tasks have been used on studies of viewpoint preferences (e.g. Palmer, Rosch, & Chase, 1981) and the aesthetics of spatial composition (Palmer, Gardner, & Wickens, 2008).

Because this is a perceptual task, one possibility is that the best visual size of the objects is driven by visual acuity constraints. One might predict that all objects will be sized at the fovea or parafovea (e.g. 2-8 degrees visual angle), perhaps modulated by the complexity of the image, without any systematic variation due to prior knowledge about the real-world size of the object. Alternatively, we might predict converging evidence with the results from Experiment 1 and 2. In this case, the visual size at which an object looks "best" might be systematically related to the logarithm of the real-world size of the object.

Methods

A separate group of ten naïve observers were recruited from the MIT participant pool (age range 18-35), gave informed consent, and received 5 dollars for their participation. 100 color pictures of real-world objects were used (see Figure 1). Larger versions of a few example images can be seen in the Appendix, and the image database can be downloaded from the first author's website. The experimental setup was the same as in Experiment 2.

At the start of each trial, the mouse position was set to the right side of the screen at a random height. Then, observers were presented with one picture of an object centered on a white background. The initial size of the object was determined by the height on the screen where the observer clicked to start the trial. Observers were told to select their preferred size to view the objects. Specifically, observers were shown a sample object at the smallest possible size of ~2 pixels ("intuitively, this size is too small or too far away") and at the largest size such that the edges of the object extended beyond the monitor edges ("intuitively, this is too large or too close"). Observers were shown that they could freely move the mouse up and down to adjust the size of the object, and clicked the mouse to select their preferred

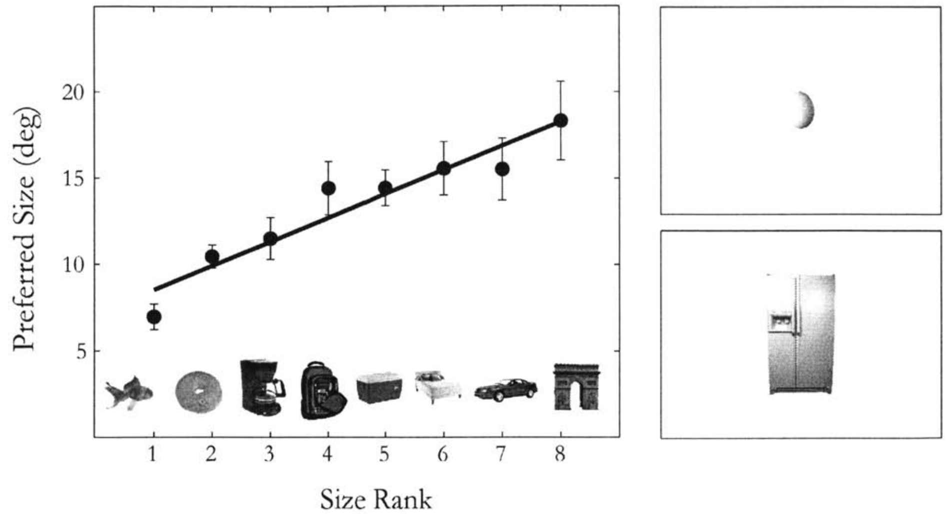


Figure 5: The average preferred size of objects, in degrees visual angle, is plotted as a function of the size rank of the objects (left). Error bars represent ± 1 S. E. M. The average preferred size of two objects from different size ranks are shown on the right.

view ("choose the view that's not too big or too small, but the one that looks best"). Each observer resized all 100 objects, with the order of objects randomized across observers.

Results

Data from one observer was excluded because they did not complete the task for all objects. The left panel of Figure 5 shows the average preferred visual size of the objects, plotted as a function of the size rank of those objects. As in previous experiments, the preferred visual size was calculated as the visual angle subtended by the diagonal of the bounding box. The average preferred size for two sample objects— an egg and a refrigerator, is illustrated in the right panel of Figure 5. The data show that as the assumed size of the objects increases, the preferred visual size at which to view them on the screen also increases systematically ($r^2 = 0.96, p < 0.001$). Thus, we again find a consistent relationship between the preferred visual size of the object and the size rank of that object in the world.

Across the 100 objects, the systematic variation in the preferred visual size was again very consistent across observers. The effective reliability was R=.84. Thus, despite the subjectivity of the task to select the "best view," smaller objects were consistently sized smaller and larger objects were consistently sized larger across observers.

These data can be converted into a ratio between the preferred visual size and the size of the monitor. Collapsing across the size ranks, the average preferred size ratio was 36% (SEM 3.6%), which was not significantly different from the average imagined size ratio or the average drawn size ratio (Experiment 1 - drawing: $t(71) = 0.17, n.s.$; Experiment 2 - imagery: $t(16) = 0.56, n.s.$). The slope of the regression line between size rank and preferred visual size was 3.9% per size rank (SEM 0.8%), which was significantly shallower than the imagery slope in Experiment 2 (mean: 6.3% per size rank, $t(16) = 2.25, p < 0.05$) with a trend toward being steeper than the drawn slope in Experiment 1 (mean: 2.7% per size rank, $t(71) = 1.99, p < 0.06$).

Discussion

These data show that when observers can freely resize objects on the screen, the preferred view of the object is proportional to the logarithm of their real-world sizes. These data rule out the simple account that acuity constraints drive visual preferences, because objects were not all resized to subtend equal visual angles at the maximal extent of the fovea or parafovea. Instead, we find that knowledge about the physical size of objects systematically influences the visual size at which objects are preferentially viewed. Similarly, the preferred visual sizes within the frame of the monitor match the ratios observed in the drawing experiment well. Thus, these data suggest that perceptual preferences about objects are related to the representations invoked by drawing and imagery tasks.

The current experiment required subjective judgments about the size at which pictures of objects "look best" (see also Palmer, Gardner, & Wickens, 2008). Despite the subjectivity of this task, observers were remarkably consistent in their preferred visual sizes, with high inter-rater reliability. One interpretation of what drives the preferred view of an object is the view with the best representational fit to existing long-term memory representations (Palmer, Schloss, and Gardner, in press). Specifically, the visual size at which a refrigerator looks best is the visual size (and the space around it) that matches with existing object representations, i.e., those that guided the drawing and imagery tasks. Perceptual preference tasks have also been conceptualized as a consequence of memory processes reflecting the output of the human inference system (Weber & Johnson, 2006). Akin to the previous literature on canonical perspective, we term this consistent visual size information the *canonical visual size.* This visual size depends on the assumed size of the object and is best specified not in terms of visual angle but in terms of visual size ratios between the object and a frame of space.

Experiment 4: Miniatures

In Experiment 4, we manipulated the size observers assumed an object to be in the world by presenting them with an image of a real-world object but telling them that it was a miniature version of that object, fit for a highly detailed architectural model. If the visual size ratio is truly a consequence of the physical size the observer believes the object to be in the world, then a "miniature" object should be preferentially viewed at a smaller size than its larger real-world counterpart.

An alternate account that predicts the data from Experiment 3 is that perhaps observers prefer to see all objects at a certain average visual size, but tend to modulate their settings around this size based on knowledge about the physical size of the object in the world. On this account, in this experiment observers who are viewing "miniature" objects should not show any difference in their preferred size ratios compared to observers who believed the objects to be typically sized real-world objects. They should have the same mean size setting, and should modulate around that preferred size by the same or perhaps a smaller dynamic range.

Further, the preferred visual sizes found in Experiment 3 could have been driven by image-level information solely (e.g. resolution, downward viewing angle, aspect ratio). The converging evidence from Experiment 1 and 2 make this unlikely. However, the miniature experiment serves as a control, as it uses exactly the same images and task as in Experiment 3, with only instructional variations. Thus, any differences in the preferred visual size between objects and miniature objects cannot be attributed to image-level effects.

Methods

A separate group of ten naïve observers were recruited from the MIT participant pool (age range 18-35), gave informed consent, and received 5 dollars for their participation. Stimuli and procedures were identical to those in Experiment 3, except for the instructions given. Here, the participants were told that they were looking at pictures of "toys" from a "highly detailed architectural model" (i.e., the kind of model that might have a toy cheese grater and a toy basketball). As before, participants were instructed to resize the objects on the screen so that they "looked the best".

Results

The left panel of Figure 6 shows the average preferred size of the objects that are thought of as "toys" (black line). For comparative purposes, these are plotted as a function of the same size rank used previously. The data from Experiment 3 is replotted for comparison (gray line). The average preferred size for two sample toy objects—a toy egg and a toy refrigerator, is illustrated in the right panel, along side the preferred size of the "typically-sized" egg and refrigerator from Experiment 3.

Overall, the average preferred size of toy objects on the screen was 5.5 degrees (S.E.M=2.13 deg), whereas the average preferred size of the same objects from Experiment 3 was 13.1 degrees (S.E.M.=3.3 deg; $t(18) = 4.65, p < 0.001$). As before, the preferred size of the objects, when seen "as toys" by the observers, still preserve the strong correlation with the size rank of the objects ($r^2 = 0.99, p < 0.001$). The slopes of the regression lines between Experiment 3 sizing regular objects and Experiment 4 sizing toy objects were not significantly different (Object: 3.9% per rank, Toy: 3.1% per rank, $t(18) = 0.9, n.s.$). Further, observers were very consistent in the relative

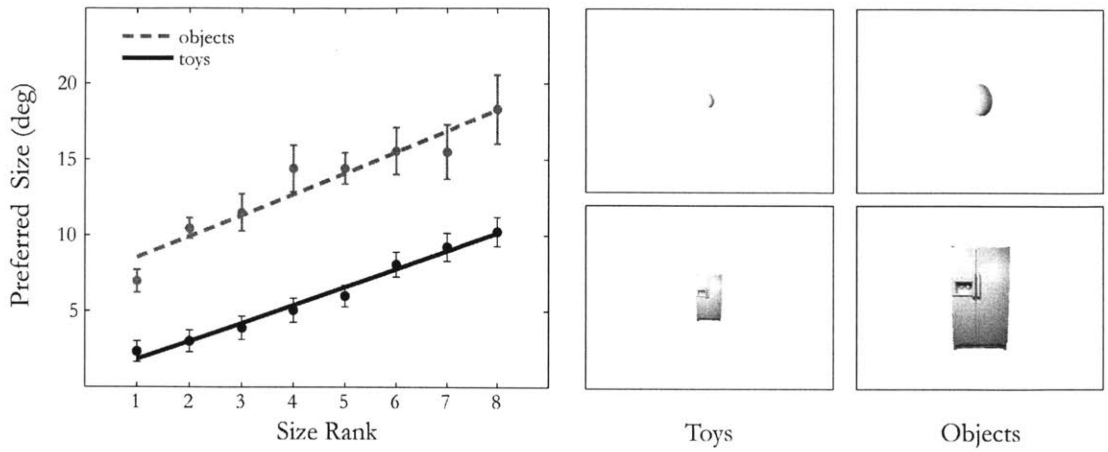


Figure 6: Left: The average preferred size to see images of "toy" objects on the screen is plotted as a function of the size rank of the object (black line). The data from Experiment 3 is replotted for comparison (gray line), in which a different set of observers resized the same images but thought of them as regular objects. Error bars represent ± 1 S. E. M. Right: The average preferred sizes of two toy objects are shown next to the average preferred sizes of those objects when assumed to be a typical real-world size.

sizes across all 100 objects, with an effective rater reliability of R=0.94.

Discussion

When observers think objects are smaller in the world, the preferred sizes of those objects are smaller on the screen. This is true even though separate groups of observers participated in Experiments 3 and 4. Further, this experiment demonstrates that preferred visual sizes are not driven solely by the image-level differences or the relationship between objects in the set, because the images in Experiment 3 and 4 were the same. Additionally, the relationship between preferred size and assumed size is preserved when observers think of the objects as miniatures. Likely this reflects the instructions that these objects were for a model, i.e., made "to scale" but at a smaller physical size. The largest miniature objects (e.g., houses, statue) were sized on the screen at around 27%. Thus we can estimate that observers likely thought of these images as having a physical size of around 30 − 60 cm (e.g. a coffeemaker or backpack), based on the Experiment 3 size ratios.

The current data also have interesting implications about how assumed real-world size influences preferred visual size, and what kind of information is stored in object representations. Likely, we don't have much visual experience with toy cheese graters, but we do have experience with cheese graters and with toys. It is also likely that learning from experience operates at multiple levels of abstraction (e.g., this specific cheese grater, all cheese graters, all kitchen appliances; and this toy, toys in general). Thus, such learned attributes can flexibly combine to generate a representation of, for example, a toy cheese grater, without ever having seen one before. As evidence that this is learned over experience, 18 to 30 month-old children sometimes make scale-errors, in which they attempt to get into a toy car or sit in a dollhouse chair, indicating that they can recognize a toy version of the object, but fail to incorporate its apparent physical size and instead carry out the associated actions with the typically-sized objects (DeLoache, 2004).

Another implication of this result is that assumed size modulates expectations about visual size. Put more strongly, a cheese grater on a white background will look more like a miniature cheese grater if it has a small ratio on the screen. Even though there were completely different observers between Experiment 3 and 4, the preferred visual size of miniatures was smaller than the preferred visual size of real-world objects. This further reinforces the main result that smaller objects in the world have smaller canonical visual sizes.

Experiment 5: Size Range

An additional factor that may be influencing the preferred size is the range of real-world object sizes in the image set. In all of the experiments reported here, participants were exposed to the whole range of real-world sizes (ranks 1-8, from very small to very large size). Here, we tested the impact of stimulus set in the perceptual preference task using a between-subjects design, where three groups of observers are exposed to a restricted range of objects sizes (e.g. only small objects in the world, only objects of medium size, or only large objects). If observers simply use a minimum small visual size for the smallest object and a maximum visual size for the largest objects, and scale the other objects between these two extremes, then the visual sizes we observe will be largely due to the stimulus set and not due to the absolute assumed size of the object. However, if there are reliable differences in the preferred visual sizes between the observer groups, even when the groups are exposed to a restricted range of real-world object sizes, then this would show that observers are guided by a common canonical visual size representation.

Methods

Three groups of 11 naïve observers were recruited from the MIT participant pool (age range 18-35), gave informed consent, and received 5 dollars for their participation. Observers completed the same procedure as in Experiment 3, but were exposed to only a subset of the items, with one group seeing only small items (ranks 1-4), another group seeing only medium items (ranks 3-6), and the final group seeing only large items (ranks 5-8).

Results

First, we examined if there were reliable differences between the three groups of observers on the averaged preferred size. We found a significant effect of group on the preferred size ratio ($F(2, 30) = 4.4, p < 0.05, \eta^2 = .22$), consistent with our predictions from Experiments 1-4: smaller visual sizes were preferred for the group seeing smaller real-world objects and larger visual sizes were preferred for the group seeing larger real-world objects.

We next compared the preferred visual sizes of each group with the original experiment in which observers were exposed to all size ranks 1 though 8. Three ANOVAs were conducted on the size ratios, one for each group of observers, with size rank as a within-subject factor, and stimulus set range as a between-subject factor (e.g. data from the observers seeing only the smallest objects was compared with data from

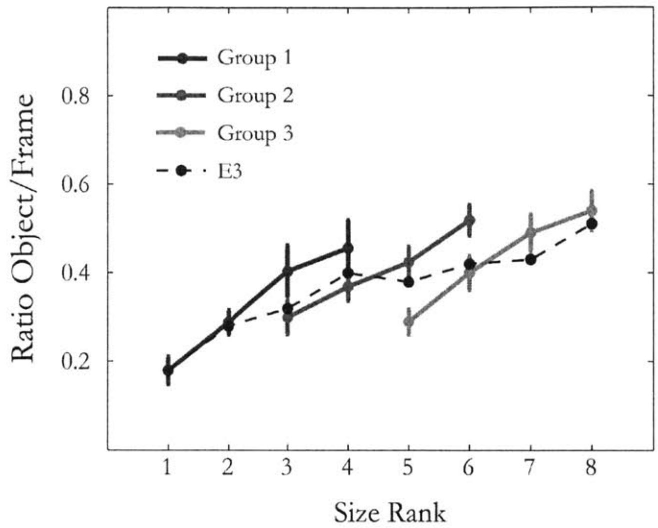


Figure 7: Results of size range experiment. The average preferred size of objects for the three groups of observers, expressed as a ratio between the size of the object and the size of the screen, is plotted for each group as a function of the size rank of the objects. Dashed line indicates the preferred sizes from Experiment 3. Error bars represent ± 1 S.E.M.

Experiment 3 for only the object size ranks of 1, 2, 3, and 4, and similarly for those seeing medium sized objects or large sized objects). The results are shown in Figure 7. Overall, the average preferred size for small objects was the same whether observers were only exposed to that range or the full physical size range (ranks 1-4; means: 32% and 29%; $F(1, 18) = 0.7, n.s.$). The same held for observers seeing only medium size objects (ranks 3-6; means: 39% and 38%; $F(1, 18) = 0.0, n.s.$) and for those seeing large objects only (ranks: 5-8; means: 45% and 45%; $F(1, 18) = 0.0, n.s.$). However, observers who saw only medium-sized objects or only large objects used a larger range of ratios on the screen than observers exposed to the whole range of objects physical sizes (medium ranks 3-6: experiment x size rank interaction: $F(1, 18) = 8.2, p < 0.01, \eta^2 = .31$; large ranks 5-8: experiment x size rank interaction: $F(1, 18) = 4.2, p = 0.055, \eta^2 = .19$).

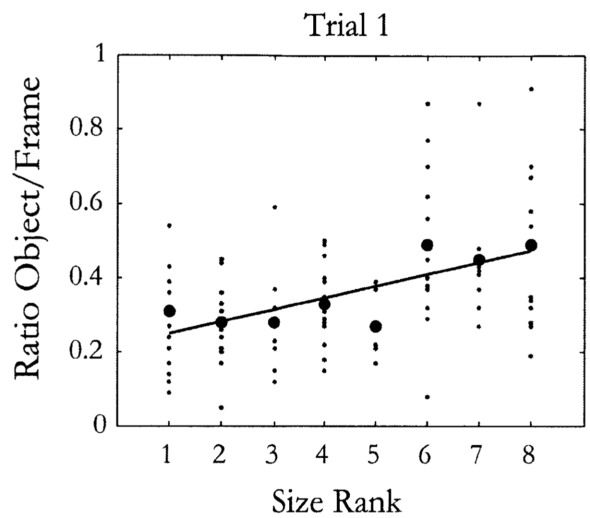


Figure 8: Trial 1 Analysis. Average drawn, imagined, or preferred visual size for the first trial only of all Experiments 1-5. These visual sizes are expressed as a ratio between the size of the object and the size of the frame, plotted as a function of the size rank of the objects. Each gray point represents a participant, with the average visual size per size rank shown in black points.

Discussion

These results show that the range of physical sizes does have an effect on the preferred visual sizes, by modulating the dynamic range between the smallest and largest item. Specifically, the preferred sizes found in the three observer groups have more size range than the preferred sizes found by participants exposed to the whole range in Experiment 3. This reveals that the object set, or the context in which a collection of objects is perceived, is another factor that modulates the preferred visual size. This result is interesting because it suggests that people have some flexibility in the scaling between assumed size and visual size. However, for the present purposes, it is also important to note that overall, the average visual size increased for each group exposed respectively only to small, medium or large objects and was consistent with the visual sizes from a different set of observers who were exposed to the whole object set. This demonstrates that the assumed size of objects influenced their preferred visual sizes, even across observers and stimuli ranges: smaller visual sizes were preferred for smaller objects and larger visual sizes were preferred for larger objects.

A related concern is that, over the course of multiple trials in the experiment, exposure to different objects with different real-world sizes may lead observers to adopt a systematic relationship between assumed size and preferred visual size over time. Thus, perhaps without this exposure to a variety of stimuli with different real world sizes, there would be no remaining effect of assumed size. If this were the case, then one would not expect to find an effect of real-world size on the very first trial. To examine this possibility, we conducted an analysis of the first trial completed for the 94 observers in Experiments 1 through 5 (excluding the first drawings of 21 observers in E1 which did not meet the criterion for inclusion). We again observe a positive relationship between the physical size rank the visual size ($r^2 = 0.18, df = 93, p < 0.001$; Figure 8). The slope of relationship is 3.2% per size rank. For reference, the average slope was 2.7% for drawing, 3.9% for perception, and 6.3% for imagery. Thus, despite the lack of power due to having only one trial per subject, this analysis suggests that, even on the first trial, the small objects were drawn, imagined and preferentially viewed at smaller size ratios than large objects. While there is likely a contribution of intertrial comparisons on the size effects found here, these analyses suggest that the consistency of the size ratios we have found in perceptual, imagery and memory tasks are not solely a consequence of intertrial comparisons or object set effects.

General Discussion

Evidence for canonical visual size

In the current studies, we asked whether accessing real world object knowledge yields consistent visual size representations across different mental processes. Using drawing from memory, imagery, and perceptual preference tasks we found that systematic visual size ratios were observed across different mental processes and across observers (See Figure 9). These results provide evidence for different canonical visual sizes for differently sized physical objects. Second, the data demonstrate that the canonical visual size of an object depends on the assumed real-world size of the object. Across

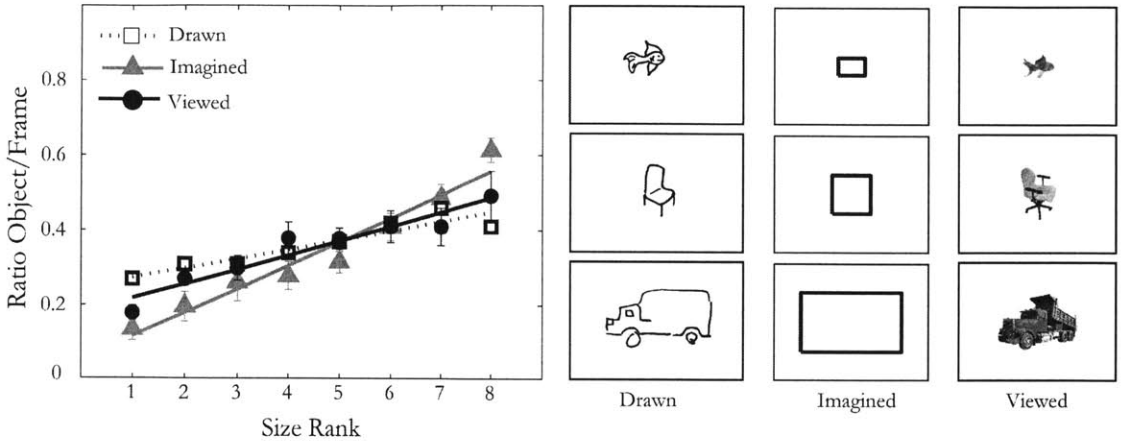


Figure 9: Left: Results of Experiments 1 (Drawn), 2 (Imagined), and 3 (Viewed), overlaid on one graph. The x-axis shows the size rank of the object; the y-axis shows the diagonal ratio of the object in the frame. Error bars represent ± 1 S.E M. Right: Example drawings of a fish, chair, and dump truck for a single observer. The average imagined size and preferred size across observers are shown for these same objects in the adjacent columns. Note that separate groups of observers participated in the Drawn, Imagined, and Viewed conditions.

all experiments and observers, there was a strong correlation with the size rank, and thus with the logarithm of the assumed size of the object in the world (Figure 9). This claim is further supported by the miniatures experiment in which we manipulated assumed size and showed corresponding changes in preferred visual size. Finally, these data argue that the canonical visual size is best characterized as a ratio between the object and the space around it. For instance, the canonical visual size of a chair is not a specific visual angle but rather is 38% of a surrounding spatial envelope (Figure 9). Experiment 1 most strongly supports specifying canonical visual size as a ratio, as the drawn size for any given object was equivalent across paper sizes when characterized as a ratio between the object and frame.

On a broader interpretation of these data, tasks which access object representations for visual size information are likely probing an underlying distribution of visual sizes, rather than just one specific canonical visual size. For example, while a strawberry may look best when presented at a size ratio of 18%, this may reflect only the most probable of a range of possible visual sizes. Exemplar-based models and view-centered models of object representation argue that observers store many instances of objects (e.g. Nosofsky, 1986; Edelman & Butlhoff, 1992; Ullman, 1989); if visual size information is also stored with these exemplars, this could give rise to a probability distribution over this dimension. The idea that object knowledge operates over probability distributions along various spatial and featural dimensions has received support from memory paradigms, in which systematic biases can be observed that reflect coding an episode with respect to a prior distribution (e.g. Huttenlocher, Hedges, & Duncan, 1991; Huttenlocher, Hedges, & Vevea, 2000; Konkle & Oliva, 2007; Hemmer & Steyvers, 2009; see also Baird, 1997). Broadly, accessing an existing object representation, e.g., for a drawing or imagery task, can be thought of as taking a sample from underlying distributions, of which visual size and perspective may be stored dimensions.

Framing effects

We found that observers were sensitive to the amount of space specified by a frame, drawing objects in such a way that across observers, a consistent ratio between the object and the paper size was preserved over a range of different frame sizes. These findings show converging evidence in support of a framing account of the "vista paradox," in which a large distant object viewed through a window (or through a naturally occurring corridor, e.g. in a cavern or street scene) appears to both shrink in physical size and recede in distance as the observer approaches it (Walker, Rupich, & Powell, 1989; see also the "coffee cup illusion", Senders, 1966). This notion that the framing ratio affects the perception of an object's physical size properties, beyond information from the object alone, has been documented in a number of other studies (e.g. Brigell et al, 1977, Kunnapas, 1955, Rock & Ebenholt, 1959). Further, it is interesting to note that under natural viewing conditions, objects are always seen in a space, maximally limited by the extent of the visual field. As such, any experienced view of an object has an implicit frame of space around it.

The relationship of the object with the space around it is only one simple statistic that may be stored from visual experience. More generally, these framing effects support the notion that object representations are inherently linked to contexts, both spatially and semantically (e.g. Bar, 2004; Oliva & Torralba, 2007). For example, reaction time benefits are found for identifying objects in semantically consistent versus inconsistent scenes (e.g., Palmer, 1975; Biederman, Mezzanotte, & Rabinowitz, 1982; Davenport & Potter, 2004), as well as for items appearing in a more likely position given the identity and position of a previous item (e.g. Grunau, Neta, & Bar, 2008). Combined, these results highlight the relative nature of object representations: in our accumulated visual experience with objects in the world, objects never appear in isolation. As such, pre-existing knowledge of object properties may be specified not only as item-specific information but also with more relative statistics, such as object-object and object-scene associations (e.g. keyboard and mouse; bed and bedroom). In the case of object size, for instance, we suggest that rather than simply storing visual angle information about objects, the relevant statistics may actually be relative measures between object angle and a visual frame of space.

Task-Demand Characteristics

One concern about these results is the issue of task-demands: are people showing effects of assumed object size because they are explicitly thinking about size while they make a size response? There are several pieces of data that speak to this issue. First, while both the imagery and perception studies (E2 and E3) directly involve making a resizing response, the drawing study (E1) does not. Here, the task instructions focus much more on object identity ("draw a cat"), while the drawn size is an indirect aspect of the task. Importantly, the results still show an effect of assumed object size. Second, demand characteristics might arise over the course of the experiment, as observers reference previous responses rather than treating each trial independently. Indeed, this is evident in our data in the restricted size range experiment (E5). However, even on observers' very first trial, the drawn/imagined/preferred visual size was still influenced by the assumed size of the object. Finally, the miniatures experiment (E4) also speaks to the issue of task-demands. Observers were told that the images were pictures of miniatures for an architectural model, i.e. very small in real-world size. Surely as a participant, one might feel as if they should select smaller sizes.

However, smaller than what? The observers were not the same as those who did E1. If there was no common understanding about the preferred visual size of a typically-sized car, they would not know how to make a toy car smaller. While none of these analyses and experiments perfectly address the issue of demand characteristics (and indeed, E5 points to the fact that other factors beyond assumed size and framing modulate the accessed visual size), the combined data from all the experiments strongly point to a role that the canonical visual size depends on the assumed size of the object in the world. The results of the drawing task (E1) are the strongest evidence of this point, as this experiment is least subject to task demand characteristics, and also provides the clearest support that canonical visual size is a relative statistic between the size of the object and its surrounding space.

Relationship between canonical visual size and real-world viewing

Experience typically arises in the real-world in which 3-dimensional geometry constrains the distributions of visual sizes that are likely for different sized objects. How do the canonical size for real-world objects compare to typical viewing distances? To explore this question, we first need to obtain typical viewing distances for real-world objects and thus what the corresponding visual angle is in one's visual field. Hubbard, Kall, & Baird (1989) obtained estimates of the typical distance of interaction for a range of everyday objects, which can be converted into visual angle measurements (from 1.5 degrees for a 3 cm object like a coin, to 25 degrees for a 4m object like a giraffe). Next, our data suggest that canonical sizes are not specified in absolute visual angles but are instead relative to a frame of space. Thus, in order to see if the visual size subtended by objects at their typical viewing distance is the same as the canonical visual size, one needs to specify what the "frame" is during real-world viewing. One intuitive possibility for the frame of real-world viewing is the whole visual field. However, with a 180-degree hemisphere as the frame, the corresponding visual size ratios at typical viewing distances are all much smaller that the canonical visual size ratios we observed in the present data. Another possibility is to use extent of the mind's eye as a proxy for the useable visual field and frame. The estimation varies between 20 to 60 degrees (Kosslyn, 1978; Hubbard & Baird, 1988; Hubbard, Kall, & Baird, 1989), with the larger estimates obtained when estimating over-flow distance of real-world objects. With a 60 degree estimate as the frame, typical visual size ratios would be between 3% for the coin to 42% for the giraffe. These estimated ratios are similar to the imagined ratios observed in the present data (see Figure 9). Of course, this speculation should be taken lightly as assumptions have been made about the size of the real-world frame and the accuracy of subjective reports of typical viewing distances. More work is required to integrate the canonical sizes found on the computer screen and drawn pages with the statistics of visual experience in the real world.

Finally, Hubbard, Kall, & Baird, (1989) have some evidence suggesting that there may be systematic differences between sizes arising from imagery vs. perceptual processes. For example, in their study, observers imagined bird's nests an average distance of $\sim$ 1 m while the average typical viewing distance was $\sim$ 6 m. In fact, when Hubbard, Kall, & Baird had observers imagine rods (unfamiliar objects) of a prespecified length, and then estimate their distance to the rod, they found that the relationship between size and viewing distance was less noisy than with familiar objects. These data suggest that canonical visual size may be derived not only from the distribution of visual experience, but also from structural or geometric properties of the object (e.g. bird's nests are rarely seen up close but the canonical visual size may be more similar to an object of similar size, such as a football, even though the distributions of visual experience with these objects are likely quite different). Future studies are required to distinguish between these hypotheses; as with canonical perspective, likely both visual experience and structural geometric factors are involved.

Familiar Size as a Depth Cue

Existing knowledge about the size of objects in the world can serve as a cue to depth—this is typically referred to as the familiar size cue. For example, in a classic study by Ittelson (1951), observers had to judge the distance to different monocularly viewed playing cards, where unbeknownst to the observers, all the cards were presented at the same distance but some playing cards were either three-quarters or one-and-a-half times the size of a normal playing card. The larger playing cards were estimated to be closer to the observer, such that a normal card at the reported distance would match the visual size of the card. Similar results were found for the smaller playing cards, estimated to be father away. These data show that when objects that have a familiar or known size, seeing them at a particular visual angle influences the perceived distance (see also Baird, 1962; Yonas et al., 1982).

Familiar size and canonical visual size are not the same: familiar size means that observers know the real-world size of objects in the world (e.g. expressed in meters). This is knowledge about an object property, i.e. object-centered information. In contrast, canonical visual size indicates that there is a privileged visual size for perceiving objects (where the visual size is expressed as a ratio between the visual angle of an object relative to a frame). Canonical perspectives and canonical visual sizes provide evidence that existing object representations contain specific information about perspective and visual size, which are viewer-centered properties. For example, at one extreme it might be argued that existing object representations are stored at one particular perspective and one particular scale. Alternatively, likely each object's representation has stored views from a range of perspectives and scales, with some perspectives and scales being more probable or preferred than others.

What, then, is the relationship between familiar size as a depth cue and canonical visual size? Both involve the observer having knowledge about the real-world size of the object. In the first case, observers can use the familiar size of an object to estimate its distance (e.g. Epstein 1963; 1965; Epstein & Baratz, 1964; Ittelson, 1951; Ono, 1969). In the case of canonical visual size, when observers access existing object representations in order to draw, imagine, or make a perceptual preference, the visual size that is outputted depends on the assumed size of the object. Combining these two ideas, one empirical prediction is that observers might be better able to use familiar size information to make accurate distance estimates if the familiar object at its canonical visual size within the real-world viewing frame.

Conclusion

Akin to studies on canonical perspective, we provide evidence that existing object representations also have canonical visual sizes, which depend on the assumed size of the object in the world relative to a frame of space. Both perspective and visual size are spatial dimensions that are under the control of an active observer—in this sense canonical views connect physical objects to a viewer in an environment. In fact, if one combines canonical perspective at the canonical visual size, this object knowledge specifies the optimal place in three-dimensional space from which to view an object. One intriguing possibility is that an active observer might use this information to reflexively navigate to a better view of objects in the world (e.g., Merleau-Ponty, 1962).

Appendix

Figure 10: Sample objects used in Experiments 3, 4 and 5 are shown here, so the level of object detail and resolution of the images is more apparent.

Chapter 3

A Familiar Size Stroop Effect: Real-world size is an automatic property of object representation

When we recognize an object, do we automatically know how big it is in the world? We employed a Stroop-like paradigm, in which two objects were presented at different visual sizes on the screen. Observers were faster to indicate which was bigger or smaller on the screen when the known size of the objects was congruent with the visual size than when it was incongruent—demonstrating a familiar-size Stroop effect. Critically, the known size of the objects was irrelevant for the task. In a second experiment, participants learned a rule to categorize novel objects as big or small, but these stimuli did not drive a Stroop effect, indicating that automatic known size processing is not cognitively penetrable. These results show that people access the familiar size of objects without the intention of doing so, demonstrating that real-world size is an automatic property of object representation.

Introduction

Every object in the world has a physical size which is intrinsic to how we interact with it (Gibson, 1979): we pick up small objects like coins and berries with our fingers, and we orient our body to bigger objects like chairs and tables. When we learn about objects, our experience is necessarily situated in a three-dimensional context. Thus, the real-world size of objects may be a basic and fundamental property of visual object representation (Konkle & Oliva, 2011) and of object concepts (Setti et al, 2009; Sereno & O'Donnel, 2009; Rubinsten & Henik, 2002).

One of the most fundamental properties of object representation is category information: we rapidly and obligatorily recognize objects and can name them at their basic-level category (Grill-Spector & Kanwisher, 2006; Thorpe, Fize, & Marlot, 1996). This indicates that when visual information about an object is processed, it automatically makes contact with category information. Here we examined whether our knowledge of an object's real-world size is also accessed automatically: as soon as you see a familiar object, do you also know how big it typically is the world?

We designed a Stroop-like paradigm to test whether the known size of the object is automatically accessed when you recognize a familiar object. In what is commonly referred to as the "Stroop effect" (Stroop, 1935; see MacLeod, 1999), observers are faster to name the ink color of a presented word when the word is congruent with the ink color than when it is incongruent (the word "pink" in pink ink or green ink). Even though the word itself is irrelevant to the task, fluent readers automatically and obligatorily read the word, even at a cost to performance. Stroop paradigms have been used as a tool to understand how we automatically draw meaning from words, and this has been extended to understand how we draw meaning from pictures (MacLeod, 1999).

In our familiar-size Stroop paradigm, we presented two objects at two different sizes on the screen and observers made a judgment about the visual size of the object ("which is smaller/bigger *on the screen?*"). Here, the identity of the objects and their known size are irrelevant to the task. Critically, the known size of the objects could be congruent or incongruent with the visual size (Figure 1a). If the known size of objects speeds or slows performance on this basic visual size task, this would be strong evidence that as soon as you recognize an object, you automatically access its known size as well.

Experiment 1: Familiar Object Stroop Task

Method

Participants

18 participants (age 18-35) gave informed consent and completed the experiment.

Apparatus

Observers sat 57 cm from a computer screen (29x39.5 cm) and viewed stimuli presented using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997).

Design

On each trial, two images of real-world objects were presented side by side, with one at a visually large size and the other at a visually small size. The task was to make a judgment about the visual size of the object ("which is visually smaller/bigger on the screen?") as fast as possible while maintaining high accuracy. In congruent trials, the known size of the familiar objects matched the visual size. For example, an alarm clock would be presented small while a horse was presented big. In incongruent trials, this was reversed and the horse was small on the screen while the alarm clock was big on the screen (Figure 1a).

Each trial started with a fixation cross for 700ms, then the two objects were presented until the observer responded. Correct responses were followed by a 900ms interval before the next trial began. Incorrect responses were followed by error feedback and a 5 second interval before the next trial began. The visual sizes were set so that the diagonal extent of the bounding box around each object was either 35% or 60% of the screen height, for visually small and big, respectively (11 and 18 degrees visual angle). This method of setting the visual size takes into account variations in aspect ratio across objects (Konkle & Oliva, 2011; Kosslyn, 1978).

Observers completed both tasks (which is visually smaller/larger?) with the order of the tasks counterbalanced across observers. The 36 big objects and 36 small objects were counterbalanced to appear in both congruent/incongruent trials with the correct answer on the left/right side of the screen, across smaller/larger visual size tasks (see also supporting information). There were 576 total trials (288 congruent / 288 incongruent).

Results

Incorrect trials and trials in which the reaction time (RT) was shorter than 200ms or longer than 1500ms were excluded, removing 3.8% of the trials. One participant was excluded due to a computer error.

Overall, reaction times for incongruent trials were significantly longer than for congruent trials (38ms, SEM=6ms; Cohen's d=1.5; 2x2 repeated measures ANOVA, main effect of congruency: $F(1,67) = 18.2, p = 0.001$; Figure 1b). All 17 observers showed an effect in the expected direction. This Stroop effect was also reliably observed in both visual size tasks, with paired t-tests showing significant Stroop effects (visually smaller task: 58ms, SEM=12ms; $t(16) = 4.98, p = 0.0001$; visually larger task: 17ms, SEM=4ms, $t(16) = 4.40, p = 0.0004$).

Additionally, people were faster to judge which was bigger on the screen than which was smaller (main effect of task: $F(1,67) = 14.6, p = 0.001$), consistent with the classic finding of faster RTs for visually larger items (Osaka, 1976; Payne 1967; Sperandio et al., 2009). The magnitude of the Stroop effect was significantly different across these tasks (task x congruency interaction: $F(1,67) = 8.9, p = 0.009$), with a larger effect when people were judging which object was smaller on the screen.

These results demonstrate a familiar-size Stroop effect: people are slower to complete a basic visual judgment when the familiar size of the object is incongruent with the visual size.

One account of these data is that our familiar size knowledge arises from extensive visual experience with objects in the world, and this expertise is required to automatically activate familiar size. However, an alternate explanation is that the interference between known size and visual size arises only at a very conceptual level. For example, if this effect is cognitively penetrable (Pylyshyn, 1999), then it would

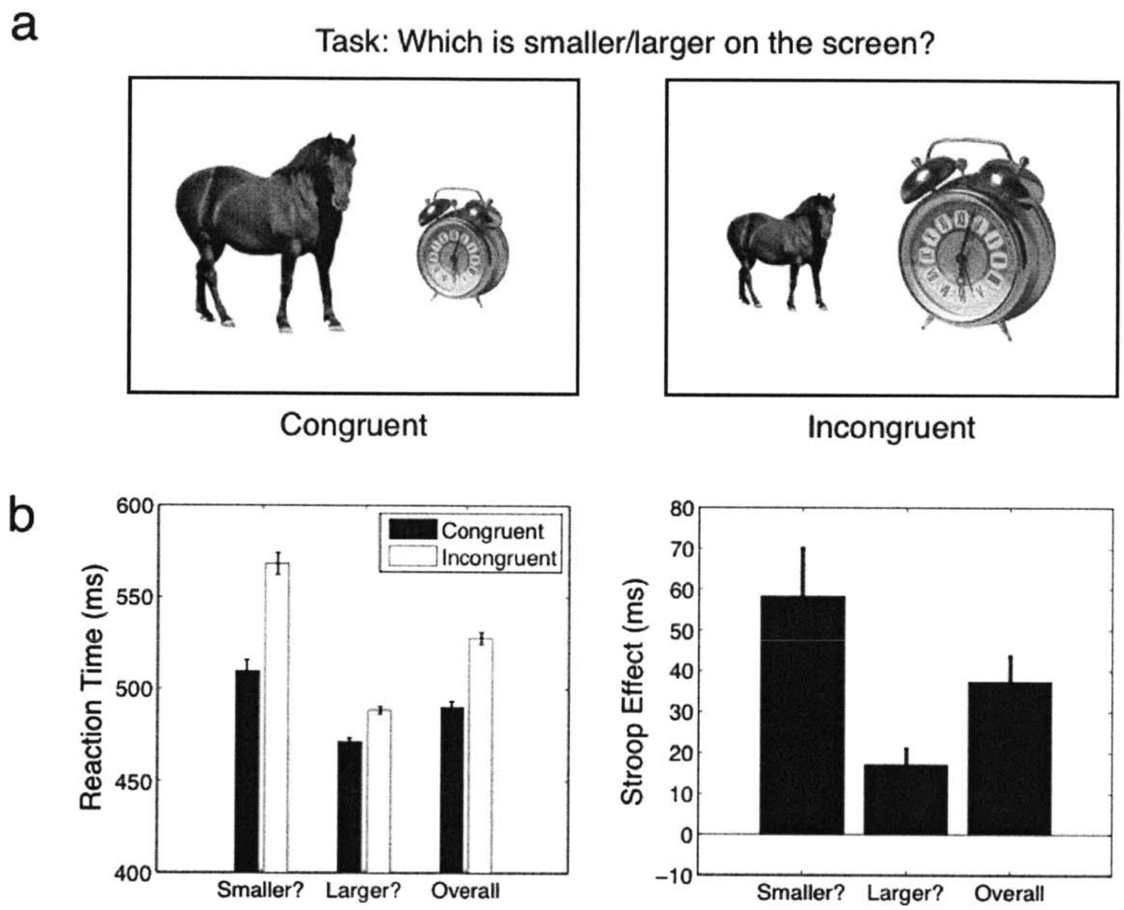


Figure 1: Known size Stroop task with familiar objects. A) Two familiar objects were presented side-by-side, and observers indicated which object was smaller or larger on the screen. The known size of the objects could either be congruent or incongruent with the presented size. Congruent and incongruent example displays are shown. B) The left panel shows overall reaction times for congruent trials (black bars) and incongruent trials (white bars), plotted for each task (smaller/larger visual size judgment) and combined across tasks. The right panel shows the difference between incongruent and congruent reaction times (Stroop effect). Error bars reflect ± 1 SEM.

be sufficient to simply instruct people that some objects are big and others are small in order to observe a Stroop effect. We test this possibility in Experiment 2.

Experiment 2: One-Shot Learning Stroop Task

We introduced participants to novel bi-color objects from a "block world", where observers were told that all objects in this world fall into two classes: Big objects made out of blue and red blocks, and small objects made of yellow and green blocks. Here we taught observers a simple rule with minimal experience, to see if this fact-based knowledge was sufficient to drive a Stroop effect. If observers show a Stroop effect on objects whose size is based on a rule, this would suggest that known size can be rapidly incorporated into our object knowledge. Alternatively, if observers do not show a Stroop effect, this would suggest that more experience with the objects is required for automatic known size processing.

Method

Participants

17 new participants (between ages 18-35) gave informed consent and completed the experiment.

Procedure

Observers were introduced to two example objects in the testing room, one small (approx. 120cm x 75cm) and one large (approx. 30cm x 18cm), depicted in Figure 2a. Participants were told that these were example objects from a block-world, where all blue-red block objects were big and all yellow-green block objects were small. Observers completed drawing and change-detection tasks to expose them to the example objects as well as the pictures of the bi-colored objects presented on the screen (see the supporting information). Following this familiarization phase, which lasted approximately 30 minutes, observers completed a block-world Stroop task.

Design

The trial design was as in Experiment 1, except that pictures of bi-color block objects were presented (Figure 2b). Each observer saw 16 unique yellow-green objects and 16 unique blue-red objects. These were presented each 32 times (congruent/incongruent trial type x correct answer on left/right side of screen x smaller/larger task x 4 repetitions), yielding 512 total trials (256 congruent / 256 incongruent).

The color-size rule was counterbalanced across observers (see Figure S1 in the supporting information). Across observers each item appeared in both blue-red or yellow-green, as both an implied big and implied small object, ensuring that object shapes were fully counterbalanced across conditions.

Results

Incorrect trials and trials in which the reaction time (RT) was shorter than 200ms or longer than 1500ms were excluded, removing 3.7% of the trials.

Overall, we observed no difference in reaction time between congruent and incongruent trials (-4ms, SEM=3ms; Cohen's d=-0.3; 2x2 ANOVA, main effect of congruency: $F(1,67) = 1.7, p = 0.21$; Figure 2c). The Stroop effect was not present in either task (smaller task: -12ms, SEM=6ms; larger task: 4ms, SEM=7ms; task x congruency interaction: $F(1,67) = 1.9, p = 0.19$). Of the 17 observers, 7 showed an effect in the expected direction and 10 showed an effect in the opposite direction. A power analysis indicated this study had very high power (>99%) to detect a stroop effect of similar magnitude as in Experiment 1, and high power (83%) to detect an effect of half the size. At a power of 99% this study could detect an effect size of d=1.0 ($\sim$12 ms Stroop effect).

To compare the Stroop effect between experiments, we conducted 2x2 ANOVA with familiar/bi-color objects as a between-subject factor, and congruency as a within-subject factor. There was a significant main effect of congruency ($F(1,67) = 25.1, p < 0.001$) and a significant interaction between experiments and congruency ($F(1,67) = 36.7, p < 0.001$). That is, people in the familiar object experiment showed a Stroop

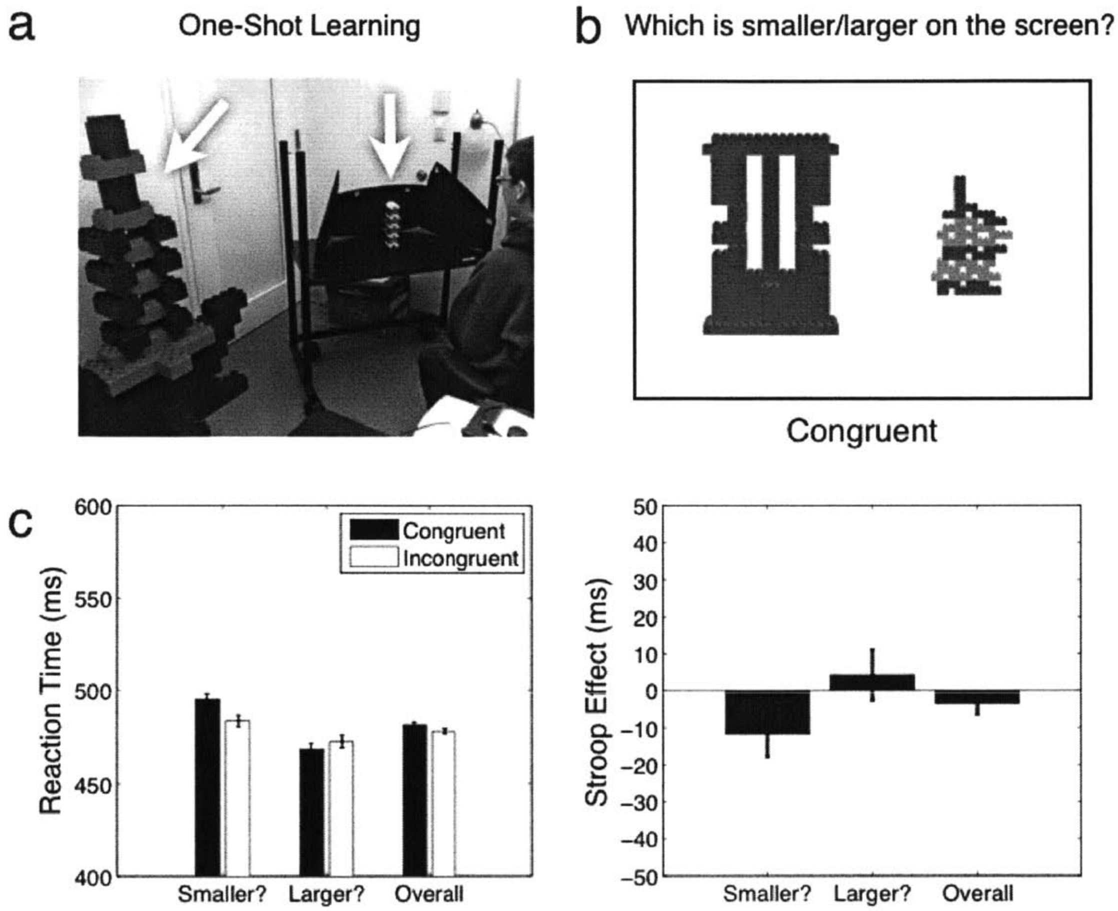


Figure 2: One-shot learning Stroop task. A) Observers were familiarized with one bi-color big object and one bi-color small object, and were told that all big objects were made of blue and red blocks, and all small objects were made of yellow-and green blocks (counterbalanced across observers). B) During the Stroop task, two bi-color objects were presented side-by-side, and observers indicated which object was smaller or larger on the screen. The implied real-world size of the objects could either be congruent or incongruent with the presented size. An example display that is congruent with the familiarization in (A) is shown. C) The left panel shows overall reaction times for congruent trials (black bars) and incongruent trials (white bars), plotted for each task (smaller/larger visual size judgment) and combined across tasks. The right panel shows the difference between incongruent and congruent reaction times (Stroop effect). Error bars reflect ± 1 SEM.

effect (38ms, SEM=6) while people in the one-shot learning experiment did not (-4ms, SEM=3). Two-sample t-tests confirmed this result ($t(32) = 6.06, p < 0.0001$). Further, there was no difference in overall reaction time between the two experiments ($F(1, 67) = 0.4, p = 0.53$), indicating that across experiments participants were not any faster or slower overall to make visual size judgments.

These across-experiment comparisons show that there is a robust Stroop effect with familiar objects that was not detected for stimuli whose real-world size is implied based on an explicit rule. Even though observers know this rule with certainty, the data show that this fact-based knowledge was not sufficient to generate a detectable Stroop effect within the reasonable power of the current design. This suggests that in order for known size to have a strong and automatic impact performance, more extensive experience and learning is required.

General Discussion

A hallmark of our object recognition system is that object processing automatically connects with stored knowledge, allowing for rapid recognition (Thorpe, Fize, & Marlot, 1996). Nearly as soon as we are able to detect an object, we can also name it at the basic-category level (Grill-Spector & Kanwisher, 2006; Mack et al., 2008). Here we show that even when object information is completely task-irrelevant, familiar size gives rise to a Stroop effect. These results suggest that we not only identify objects automatically, but we also access their real-world size automatically.

A previous study used similar displays as in Figure 1a, but observers judged which object was bigger *in the world* (Srinivas, 1996). They also found a Stroop effect, with observers faster to make a known size judgment when the visual size was congruent. Thus, Srinivas demonstrated that a perceptual feature (visual size) influenced the speed of a conceptual/semantic judgment (which is bigger in the world), which makes sense as visual size may be a route to accessing known size more quickly. The current study demonstrates the complementary effect: known size facilitates/interferes with a visual size judgment, revealing the speed and automaticity with which task-irrelevant semantic information is brought to bear on a very basic perceptual task. Together these results speak to the integral nature of perceptual and semantic features, demonstrating a direct and automatic association between known size and visual size.

What is the underlying relationship between known size and visual size that gives rise to a Stroop effect? One possibility is that interference occurs at a relatively high-level concept, arising from a common abstract concept of size. However, the data from the one-shot learning experiment are not wholly consistent with pure conceptual interference. Had we found that teaching people a simple size rule led to a Stroop effect, this would be strong evidence supporting a more abstract locus of interference. However, we observed that simply being able to state whether something is big or small with minimal experience did not lead to strong interference with visual size judgments. Instead, the data imply that the association between an object and its known size has to be learned with repeated experience before it can automatically interfere with a visual size task.

A second possibility is that interference in this task arises in more perceptual stages of processing. Consistent with this idea, a number of researchers have claimed that stored information about real-world size is represented in a perceptual or analog format (Moyer, 1973; Pavio, 1975; Rubinsten & Henik, 2002). Further, objects have a canonical visual size, proportional to the log of their familiar size, where smaller objects like alarm clocks are preferred at smaller visual sizes, and larger objects like horses are preferred at larger visual sizes (Konkle & Oliva, 2011; Linsen et al., in press). On this more perceptual account of interference, in the congruent condition both objects have a better match to stored representations which include visual size information, facilitating and/or interfering with visual size judgments.

Certainly, these two accounts of the familiar-size Stroop effect are not mutually exclusive. It is likely that there is interference at multiple levels of representation, from more perceptual ones (realized in visual-size biases) to more conceptual ones (e.g. semantic facts that a horse is big). The present data do suggest, however, that strong interference effects are not granted in one shot by learning a rule, but instead must be grounded in repeated perceptual experience. This mirrors the results for the classic stroop effect, where intermediate or fluent reading ability is required to show interference with color naming (Comalli 1962; MacLeod, 1991). Importantly, regardless of the sources of interference between known size and visual size, the present data clearly show that the known size of objects is automatically activated when an object is recognized.

Supporting Information

Experiment 1: Aspect Ratio Counterbalancing

Pictures of real-world objects vary in their aspect ratio, which influences the apparent visual size of the object (e.g. a wide object tends to look smaller than a tall object, even if they are equated on other dimensions such as the diagonal extent or fit inside similar bounding circles). To account for this factor, the 36 big and 36 small object stimuli were paired by aspect ratio in advance. Matching the aspect ratio of the two objects removes uncertainty about how to compare the size of objects with different aspect ratios.

In the first half of the experiment, all trials contained pairs of objects that were matched in aspect ratio ("matched pairs"). Thus all observers saw these same pairs. Each pair was fully counterbalanced, appearing in congruent/incongruent trials, with the correct answer on the left/right of the screen, in both bigger/smaller visual size tasks.

To double the number of trials to increase power, we pseudo-randomly paired the big and small objects, and this pairing was different for each observer. Here, we ensured that if a tall big object was paired with a wide small object (e.g. A door and harmonica), then there would also be a tall small object paired with a wide big object (e.g. A bottle and a train). Thus, the aspect ratio of two items on any given display can be different, but across trials the ratio of aspect rations was balanced across congruent and incongruent condition. Again each of these "pseudo-pairs" was fully counterbalanced.

Both of these methods of stimulus selection ensure that the aspect ratio (or ratio of aspect ratios) was completed balanced across congruent and incongruent trial types. Thus, any differences in reaction times between these two conditions cannot be driven by effects of aspect ratio on perceived visual size. Within each part of the experiment (matched trials, paired trials), observers completed both visual judgment tasks (which is visually smaller/larger?) with the order of the tasks counterbalanced across observers.

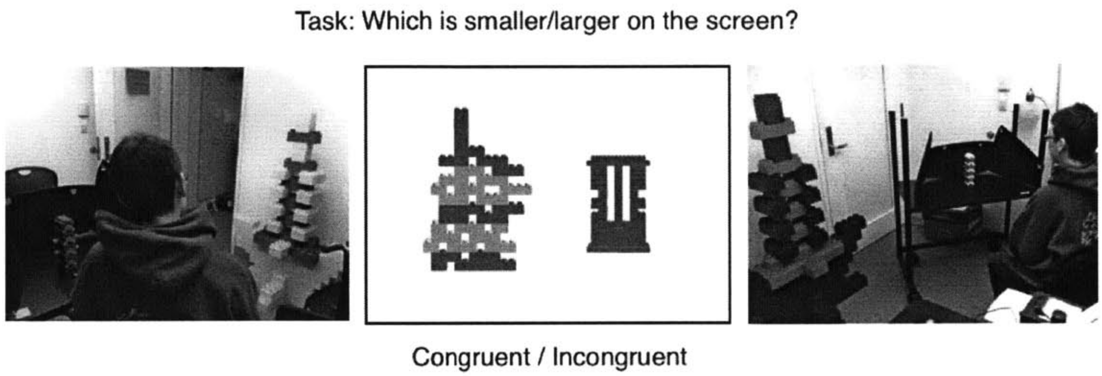
Results. We analyzed the reaction times for matched trials in the first half of the experiment and paired-trials in the second half of the experiment, and found consistent convergent results. For the matched pairs, reaction times were again significantly longer on incongruent than congruent trials (mean RT difference=34.0, SEM=8.7; $t(16) = 3.94, p = 0.001$), and errors were more frequent in incongruent vs. congruent trials ($t(16) = 3.84, p = 0.001$). Similarly, Stroop effects both in RT and errors were observed in the pseudo-pairs (mean RT difference=28.6, SEM=8.7; RT difference: $t(16) = 3.29, p = 0.0046$; Error difference: $t(16) = 3.61, p = 0.0023$). There was no difference in the overall reaction time ($t(16) = 0.32, p = 0.75$) or magnitude of the Stroop effect ($t(16) = 0.53, p = 0.60$) across these stimulus pairing conditions.

Experiment 2: Familiarization Methods

To ensure that observers were familiarized with the reference objects, participants completed two tasks. For the first task, observers were given a sheet of paper, a pencil, and markers, and had 3 minutes to create a colored drawing of each reference object. For the second task, participants were given 10 seconds to study the reference object, after which they closed their eyes and one or more blocks were added to the object. They then had three guesses to indicate which blocks were added and were given feedback. Added blocks were always of the same colors as the reference objects. Observers completed three change-detection trials per reference object.

We next familiarized the participants to pictures of other objects from block world, presented on the screen. Each trial, the participant was instructed to turn and look at the big or small reference object, after which they looked back at the screen and a picture of a new object from block world presented. This picture was always of the same size class (e.g. a "small" object also made out of green and yellow blocks). Observers' task was to determine if the depicted object was slightly taller or slightly shorter than the reference object, were it in the world along side the reference object. Each depicted object was presented for 500ms, which did not allow sufficient time for observers to count the number of rows of blocks. The 32 novel bi-color object pictures were presented in random order. This task ensured that participants were familiarized with each depicted object on the screen, and further required them to conceive of its physical size in the world. Combined, these familiarization tasks took 30 minutes to complete.

Experiment 2: Counterbalancing



Supplementary Figure 1. To counterbalance for the pairing of color and known size, some observers learned that big objects were yellow-green (left image) while others learned that big objects were red-blue (right image). Thus depending on the familiarization, the same display fell into either the congruent or incongruent condition (center image).

Chapter 4

The representation of objects in ventral temporal cortex depends on real-world size[1]

While previous work has shown systematic organization for specific object categories with functional selectivity (faces, houses, bodies), the large-scale organization of other object categories with more distributed cortical representations remains unknown. Here we find that object representations can be differentiated by their real-world size in adult human occipito-temporal cortex. In a functional neuroimaging experiment, observers were shown pictures of known big and small objects presented at the same retinal size. A region in the parahippocampal gyrus was preferentially active for big objects, while an adjacent region in inferior temporal cortex was more active for small objects, with a mirrored organization along the lateral surface. These regions show known-size preferences across different object categories and retinal sizes, and during mental imagery. The systematic medial to lateral organization of big and small objects suggests that real-world object size is an organizing dimension of distributed object representation.

Introduction

One of the most robust results in visual neuroscience is the systematic response of a large section of ventral temporal cortex to objects and shapes (Grill-Spector &

[1]This chapter is under review as Konkle, T. & Oliva, A. (2011). The representation of objects in ventral temporal cortex depends on real-world size.

Malach, 2004; Milner & Goodale, 1995; Ungerleider et al., 1982). How object representations are organized within this cortex is an active research area, with evidence for category-selective regions for only a few object categories (faces, bodies, letter strings; Cohen et al., 2000; Downing et al., 2001; Kanwisher et al., 1997; McCarthy et al., 1997), and for partially or fully distributed representations for all other objects, reflected in distinct and reliable multi-voxel patterns of activity (Carlson et al., 2003; Cox & Savoy, 2003; Haxby et al., 2001; Norman et al., 2006; O'toole et al., 2005). While it has been suggested that multi-voxel patterns of activity reflect inhomogeneities of category selectivity at the sub-voxel level (e.g. for orientation in V1 see Kamitani & Tong, 2005), more recent analyses have shown that patterns of activity are reliable even with increased spatial smoothing (Op de Beeck, 2010), can generalize across subjects(Shinkareva et al., 2008), and likely are capitalizing on large-scale organization (Freeman et al., 2011). These results indicate that there is a systematic organization to these distributed object activation patterns that we do not yet understand.

One intrinsic and often overlooked property of an object is its physical size in the world (Cate et al., 2011; Konkle & Oliva, 2011; Mullally & Maguire, 2011): Small objects, like blueberries or paperclips, are held in the fingers or hands, are manipulable, tend to have rounder shapes, are typically experienced with a steep head angle at small retinal sizes (<3 degrees at arms length). Big objects like chairs, tables, or even houses, are interacted with by body movement, are more navigationally relevant, tend to be boxier, and are typically experienced with a head angle more parallel with the horizon and subtend large retinal sizes (~30 degrees for a car at 10 m; see also Haldane, 1928). Thus, the size of objects in the world influences how an observer interacts with objects and leads to systematic biases in visual experience with each object category, influencing distributions of experienced shape, features, and eccentricity.

In the current study, we investigated if the known size of real-world objects was an organizing dimension of object representation. Here we focused on the representations of everyday objects excluding faces, animals, and classically-defined tools. Using a large stimulus set of specific real-world objects, we examined if any neural regions showed differential activity to big versus small objects, independent of retinal size. Our findings showed a consistent organization of big and small objects in medial to lateral regions of ventral temporal cortex in the left-hemisphere. A similar mirrored pattern of organization was also observed along the lateral surface. Our results suggest that the real-world size of objects is a systematic property that predicts large-scale patterns of activity for objects in ventral visual cortex.

Results

Differential responses to big and small objects

To examine if objects of big and small real-world sizes were systematically represented in different regions of cortex, we conducted a functional magnetic resonance imaging experiment in which images of isolated big objects (e.g. car, piano) and isolated small objects (e.g. strawberry, safety pin) were presented at the same retinal size on the screen to 12 observers. (Fig. 1a; Experiment 1). We conducted a whole-brain random effects analysis to identify regions preferentially active to known small objects or to known big objects ($P < 0.001$, cluster threshold $> 100mm^3$). Along the ventral surface of the brain, a bilateral region of the parahippocampal gyrus was more active to big relative to small objects (henceforth labeled as "BigV" or big-ventral region), while a left-lateralized adjacent region in the inferior temporal sulcus extending to the fusiform gyrus was more active to small relative to big objects (henceforth "SmallA" or small-anterior region, Fig. 1b). Along the lateral surface, a more posterior small-preference region was observed ("SmallP" or small-posterior), with a big-preference region in the right transoccipital sulcus ("BigD", or big-dorsal; Table 1).

These regions were observed reliably in single subjects (Fig. 1b; Table 1), with a SmallA region present in 9 of 12 participants, a SmallP region present in all 12 participants, and a bilateral BigV region present in 10 of 12 participants (FDR < 0.05, cluster threshold$> 100mm^3$). The SmallP region was bilateral in half of the participants. The Big-Dorsal region was less reliably observed at the single-subject level (5

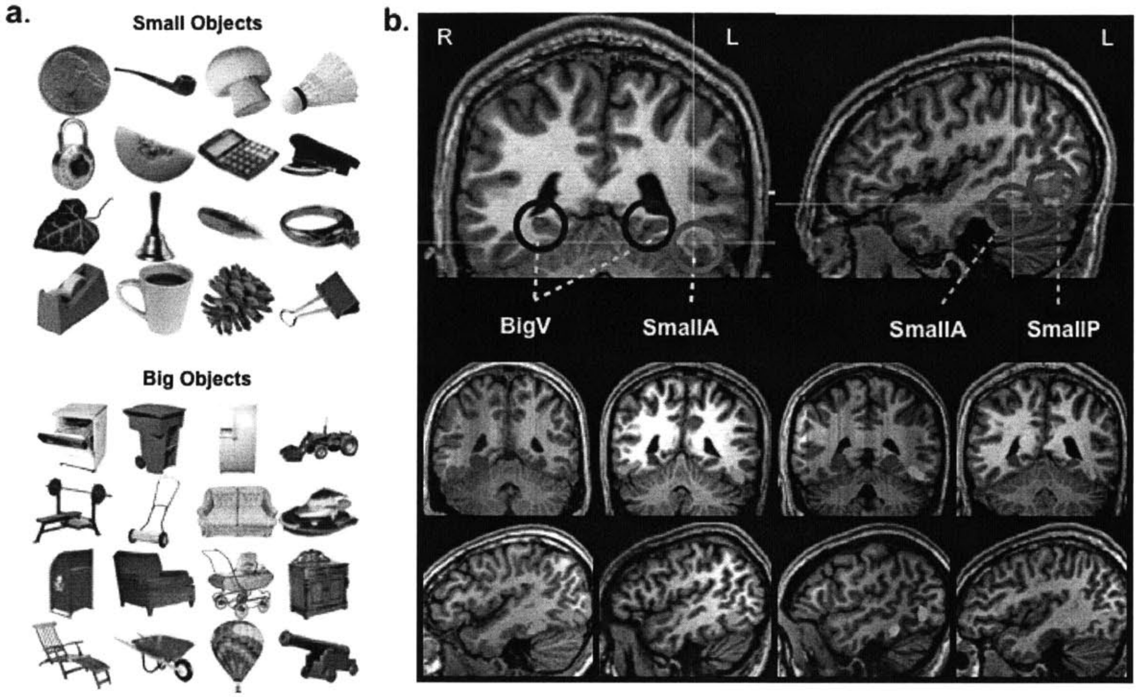


Figure 1: (A) Example objects of known small and big sizes. All images were presented at the same retinal size. The stimulus set contained 200 small and 200 big objects. (B) Contrasts of known big vs. small objects. Upper Panel: Results of a random-effects analysis, small¿big contrast, ($n = 12, P < 0.002$, cluster threshold=10, see Online Methods), plotted on the brain of one sample subject (sagital section, x=-42, coronal section, y=-42). The bilateral region with preference for big objects on the ventral surface is shown (BigV). Two small-preference regions were found, one anterior (SmallA) and one posterior (SmallP). Lower panel: These regions of interest are shown for 4 participants (12 of 12 showed a SmallP region; 10 of 12 showed a SmallA region; 10 of 12 showed a bilateral Big region).

Region		n	X	Y	Z	Peak T
L Inferior Temporal	SmallA	9/12	-42	-46	-11	6.0
L Lateral Occipital	SmallP	12/12	-42	-61	-2	5.9
L Parahippocampal	BigV	10/12	-30	-40	-5	-6.5
R Parahippocampal	BigV	10/12	30	-40	-2	-6.4
R Transoccipital	Big-Dorsal	5/12	12	-73	40	-7.0

Table 1. Talairach coordinates of the peak voxel from the group random effects analysis.

of 12) with a more variable position across subjects, and it was thus not included for further analysis. For all subsequent experiments and for additional participants, this experiment was used as a localizer to independently define these regions, and only responses in the left-hemisphere are reported as there were no differences between the responses in the left and right hemisphere.

To estimate the effect size within these regions, 8 new participants were shown two runs of the big and small object localizer. Regions of interest were estimated from the first run for each subject and the magnitude of activation to big and small objects was computed in these regions using data from the second run using an ROI GLM. All 8 participants showed a SmallP and SmallA region in the left hemisphere, and 7 of 8 showed a BigV region in the left-hemisphere, providing a replication of Experiment 1. These regions showed differential responses that were 1.5 to 1.7 times higher for objects of the preferred size relative to objects of the non-preferred size (see Supplementary Fig. 1; Supplementary Table 1). Both regions also showed a significant above-baseline response to objects of the non-preferred size (Supplementary Table 1). Thus, these objects generally activate a large swath along the ventral surface of cortex, and here we show big objects have a peak of activity on the medial aspect while small objects have a peak of activity on the lateral aspect.

Tolerance to retinal size changes

Ventral temporal cortex has been well-characterized as having object-selective responses that are tolerant to changes in retinal size, position, and viewpoint–a hallmark of high-level object representations (DiCarlo & Cox, 2007; Grill-Spector et al., 1999; Sawamura et al., 2005; Vuilleumier et al., 2002). Consistent with this, in Experiment 2, we manipulated the retinal size at which the objects were presented, and found that all of these regions showed more activity to objects of the preferred known size independent of retinal size (Fig. 2; main effect of known size: BigV: $F(1,23) = 51.5, p = 0.001$; SmallA: $F(1,23) = 85.8, p < 0.001$; SmallP: $F(1,31) = 317.7, p < 0.001$). In the BigV region, there was also an effect of the retinal size, with a stronger response to stimuli presented at retinally large compared

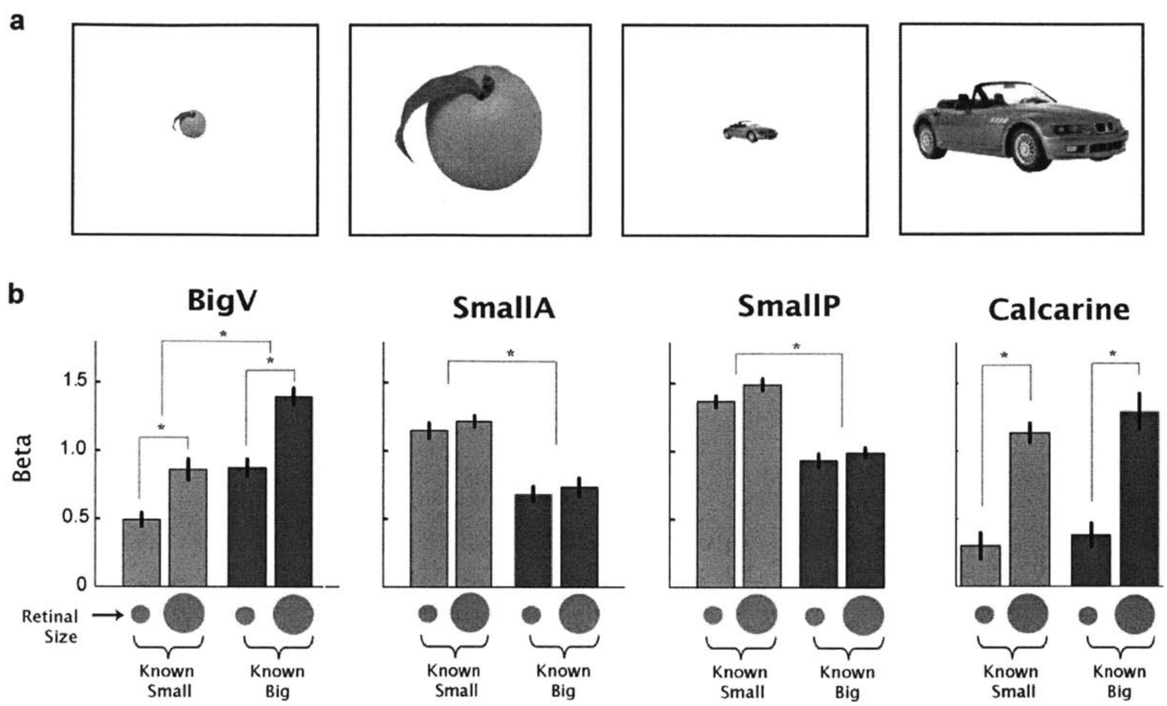


Figure 2: Retinal Size Manipulation Results. A) Objects of known small and big sizes were presented at small and large retinal sizes on the screen. B) Activations in independently-localized BigV, SmallA, and SmallP, and anatomically-defined early visual cortex regions (Calcarine) in left hemisphere were measured with ROI GLMs and the beta weights for the four conditions are shown. Error bars reflect $\pm$ 1. S.E.M. The BigV region showed effects of both the known size and the retinal size, while the small regions showed only preference for the known size of objects with no modulation to retinal size. The early visual control region showed modulation by the retinal size, with no effect of known size of the object.

to retinally small sizes (main effect of retinal size: Big-L: $F(1,23) = 14.8, p = 0.012$; SmallA: $F(1,23) = 0.6, p = 0.46$; SmallP: $F(1,31) = 5.0, p = 0.06$). Thus, the BigV region shows a large retinal size preference for both known small and known big objects, suggesting that the features it represents are not fully scale-invariant and are enhanced by peripheral input (Arcaro et al., 2009; Levy et al., 2001; Levy et al., 2004). As a control region, we examined the response in an anatomically-defined region of early visual cortex along the calcarine sulcus, and found more activity for retinally larger images than retinally smaller images, with no effects of known size (Calcarine: retinal size: $F(1,27) = 22.8, p = 0.003$; known size: $F(1,27) = 2.5, p = 0.16$).

The retinal size manipulation also serves as a control for the distribution of low-level features in early retinotopic areas (Levy et al., 2001). One potential concern is that pictures of big and small objects presented at the same retinal size may give rise to an uneven feature distribution presented to early foveal and peripheral retinotopic cortex. However, both big and small regions showed tolerance to retinal size variation, which varies the features presented to early areas. Thus, any uneven feature distribution stimulating foveal versus peripheral retinotopic cortex cannot explain away the activity in the big and small regions.

Mental imagery of big and small objects

Mental imagery requires observers to draw on stored knowledge about the visual form of an object (Mechelli et al., 2004). For example, mental imagery of faces and places have been shown to activate face- and place-selective visually responsive regions (O'Craven & Kanwisher, 2000). More recently it has been demonstrated that even the multi-voxel pattern of activation across ventral visual cortex for tools, foods, faces, and buildings is similar from perception to imagery (Reddy et al., 2010; Stokes et al., 2009). These results suggest that ventral temporal cortex contains stored information about the visual appearance of different kinds of objects, which is activated similarly during mental imagery and perception. Thus, when imagining different real-world objects, we predicted that big objects would preferentially activate medial ventral cortex while small objects would preferentially activate lateral ventral cortex.

To test this, names of objects were presented aurally to a new set of observers, whose task was to form a mental image of each object (Experiment 3). Afterwards they were presented with known big and small objects visually (as in Experiment 1), to independently localize the big and small regions of interest in each subject. Consistent with our predictions, when these participants imagined big and small objects, the big and small regions showed more activity to objects with the preferred known size (Fig. 2; BigV : $t(6) = 4.0, p = 0.007$; SmallA: $t(7) = 2.4, p = 0.048$; SmallP marginal: $t(7) = 1.8, p = 0.107$). These results also serve as a control for any concerns that our previous results were driven by pictorial artifacts of the stimuli: here, any perceptual features instantiated via imagery processes are meaningfully tied to object concepts and are not driven by unintentional feed-forward stimulus artifacts.

Size Processing versus Stored Knowledge

One potential interpretation of these data is that the magnitude of activity in these regions is related to the size the observer thinks the object is in the world. For example, the bigger one conceives of an object, the more it will drive activity in the big region and the less it will drive activity in the small regions, independent of the object's identity. To test this, observers in Experiment 3 also imagined the big and small objects at an atypical size, e.g. a tiny piano (the size of a matchbox), or a giant apple (the size of car). If activity in these regions is driven by the conceived size of the object, then in these size-violation conditions the response in these regions should be reversed. Alternatively, if activity is driven by activating stored representations of the visual form of big and small objects, then the big and small regions should respond similarly between the typical size conditions and the size-violation conditions.

When observers imagined big and small objects in the size-violation conditions, we still observed that the big and small regions showed more activity to objects with the preferred known size. That is, imagining an apple the size of a car still activated the small-preference regions more than imagining a tiny piano (see Fig. 3; SmallA: $t(7) = 2.6, p = 0.036$; SmallP: $t(7) = 2.4, p = 0.048$; BigV trending: $t(6) = 1.7, p = 0.136$). Thus activity in these regions do not reflect the conceived size of the imagined object; these regions are not performing a size computation independent of object identity.

The big region had a less pronounced preference for big relative to small objects when those objects were imagined at the wrong sizes (marginally significant interaction: $F(1, 27) = 5.9, p = 0.051$). This suggests that this region may in part be reflecting the physical size an observer imagines the object to be. However, a more parsimonious account of this data is that the big region also has a peripheral preference, as observed in our retinal size manipulation experiment (Fig. 2). If observers imagine giant apples at a large retinal size and tiny pianos at a small retinal size (see Konkle & Oliva, 2011), then this would give rise to the results observed here. Consis-

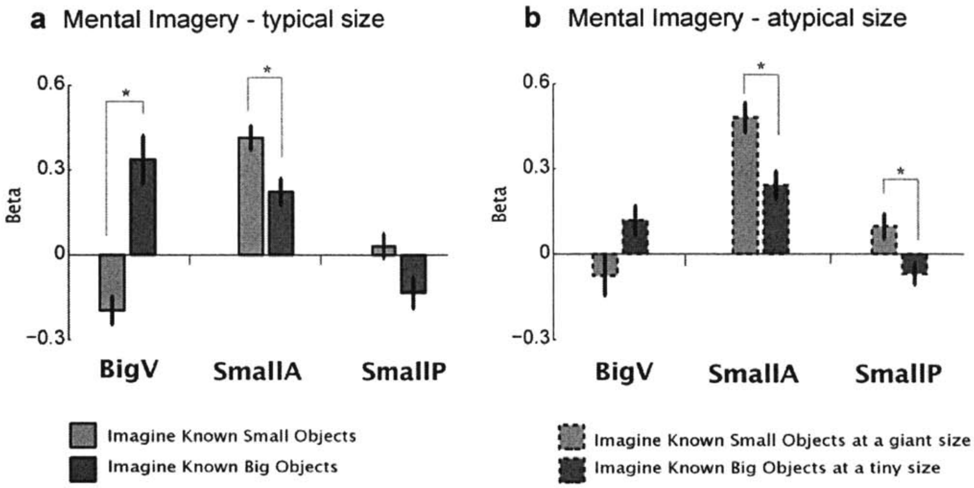


Figure 3: Mental Imagery Results. Activations in independently-localized BigV, SmallA, and SmallP regions in left hemisphere are shown. Orange bars show data for imagined objects known to be small (e.g. strawberries) and blue bars show data for imagined objects known to be big (e.g. pianos). (A) Bars with solid borders reflect conditions where observers imagined typically-sized objects. (B) Bars with dashed borders reflect conditions where the objects were imagined at atypical sizes. These results show activity in these regions is driven more by the known size of the typical object than by the conceived size of the object.

tent with this interpretation, the small regions did not have any strong modulations by retinal size, and did not show an interaction in the size-violation conditions.

Organization of Category Information

In all experiments reported so far, objects were presented in blocks by the known size of the object. Thus one potential concern is that only a few of the objects (e.g. monuments or tools; Aguirre et al., 1998; Chao et al., 1999) drive the response in these big and small preference regions. In Experiment 3 we presented observers with object exemplars blocked by category (e.g. 16 backpacks, 16 grills), allowing us to estimate the response in these regions to each object category independently. These object categories parametrically varied in known size on a log scale (Fig. 4a, see Supplementary Methods and Supplementary Fig. 2; see also Konkle & Oliva, 2011).

The big and small regions showed a systematic modulation of overall activity based on the real-world size of the object (Fig. 4b; BigV: $F(15, 127) = 8.1, p < 0.001$; SmallA: $F(15, 79) = 3.6, p < 0.001$; SmallP: $F(15, 127) = 3.9, p < 0.001$). In the

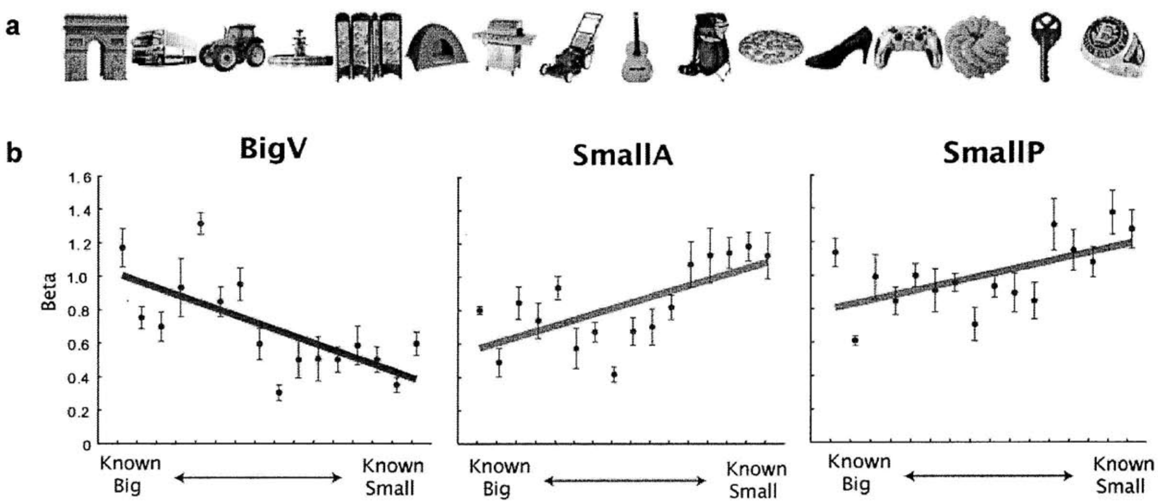


Figure 4: Stimuli and results of the category blocks experiment. A) 16 different categories of objects were presented in a blocked-design. An example stimulus from each object category is shown, arranged by known size. These object categories follow a logarithmic scaling along this dimension. B) GLMs were conducted in the independently-defined regions-of-interest in each subject, with data for BigV, SmallA, and SmallP regions in left-hemisphere shown. Each plot shows the average beta values along the y-axis, computed for each category across subjects. The categories are arranged along the x-axis by real-world size, as shown in (a), and the regression line over the group data is shown. Error bars reflect ± 1. S.E.M. Multiple object categories showed high responses in each region, with systematically more activity with increasing or decreasing object size in the big or small regions, respectively.

big region, multiple categories of objects with a large real-world size significantly activated this region, indicating the responses we observed in the first experiment were not driven a select set of stimuli from particular categories. Similarly, the small regions showed a high response to a variety of small object categories that are not considered classical tools (Fig. 4a). We observed a relatively parametric modulation of response in these regions as a function of known size, both when calculated on the group averaged betas and also within single subjects (BigV: mean $r = 0.51, t(7) = 8.1, p < 0.001$; SmallP: mean $r = -0.36, t(7) = -4.0, p < 0.005$; SmallA: mean $r = -0.46, t(4) = -4.7, p = 0.01$; see Methods, see Supplementary Table 2 for single subject results). However, it is not the case that there is a clear linear mapping between the log of known object size and the response of these regions. Nevertheless, the data show that a variety of big object categories activate the big region, ruling out any concerns that these regions were solely driven by specific categories in our previous sets such as houses or monuments (Aguirre et al., 1998) or places (Epstein & Kanwisher, 1998).

Discussion

Nearly all object categories besides faces, bodies, places, and letter strings, do not have a spatially-contiguous and highly-selective cortical representation, but instead activate a swath of ventral and lateral temporal cortex to varying degrees (Carlson et al., 2003; Cox & Savoy, 2003; Haxby et al., 2001; Norman et al., 2006; O'toole et al., 2005). The organizing dimensions of object representations within this cortex have yet to be discovered. Here we show that object-responsive cortex has reliably differential responses to known big and small objects, suggesting that the real-world size of objects is a large-scale organizing dimension of object representation.

On the ventral surface, we observed a medial region responsive to big versus small objects (BigV) and an adjacent lateral region more responsive to small versus big objects (SmallA; see Fig. 5). Along the lateral surface of cortex, we observed another region of cortex with preferential responses to small objects (SmallP) and a dorsal

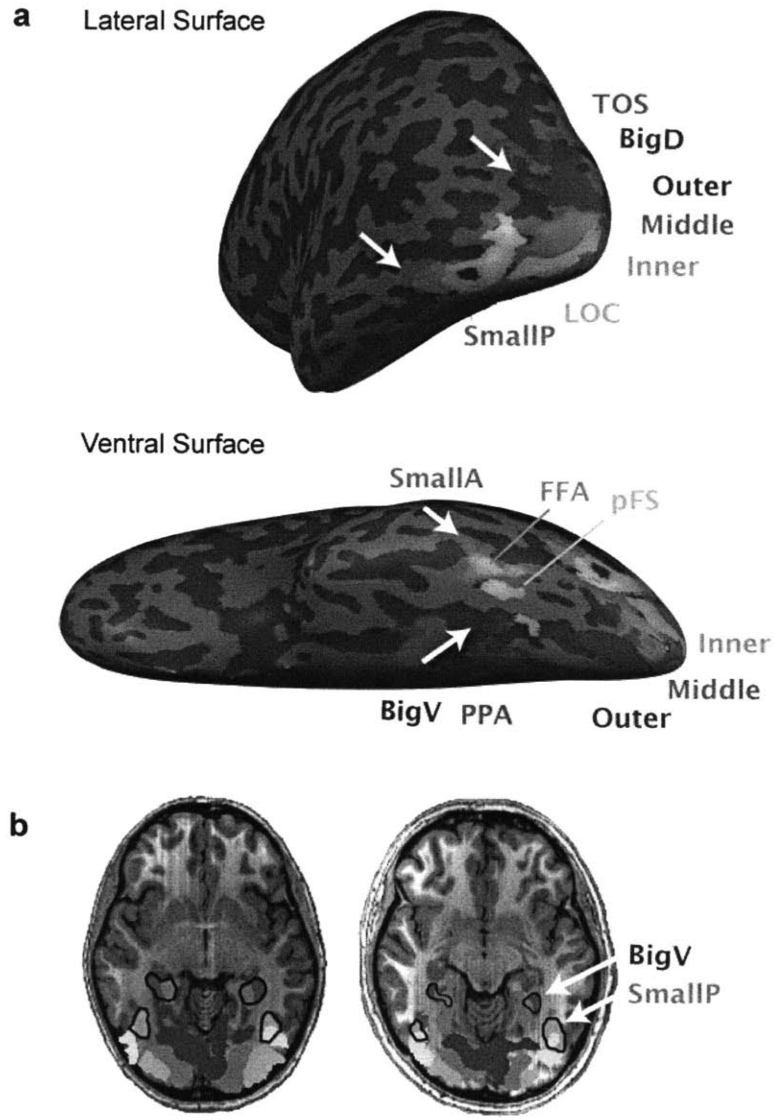


Figure 5: Relationship of these regions to other well-characterized regions. A) Functionally-localized regions from a single representative subject are shown on an inflated brain. Inner, middle, and outer eccentricity rings are shown in light, medium, and dark blue respectively (see Supplementary Methods). LOC and pFS (objects>scrambled) are shown in yellow, FFA (faces>objects) is shown in pink, and PPA and TOS (scenes>objects) are shown in green. The SmallA and SmallP regions are shown in orange and the BigV and BigD region is shown in blue, also indicated with white arrows. B) Axial slices from two different subjects (z=-3; z=-6). BigV and SmallP regions are outlined in black, with voxels that overlapped with PPA and LOC shown in the green and yellow, respectively.

region with preferential responses to big objects (BigD). The ventral activations were reliable within and across observers and were localized with less than 10 minutes of scanning. In subsequent experiments, we demonstrated that (i) these regions are more responsive to objects of the preferred size independent of changes in retinal size, (ii) these regions are activated during mental imagery of objects, both when imagining typically-sized objects and objects at a non-normative size, and (iii) these regions are not driven by a few specific categories but instead respond to many different object categories, with a relatively parametric response to objects spanning the range of real-world sizes.

Notably, objects of the non-preferred size also have a response well above baseline in all of these regions (See Supplementary Information; see also Fig. 2, Fig. 4, and Supplementary Fig. 1). Thus, the big and small regions described here should not be considered highly-selective modules, as has been argued about other category-selective regions along occipito-temporal cortex (Kanwisher, 2010). Rather, these data are consistent with a distributed account of the representation of everyday objects, where a large swath of cortex is active during the processing of such objects, but to different degrees (Haxby et al., 2001; Ishai et al., 1999). We demonstrate that the size of the object in the world predicts more medial or lateral ventral cortex engagement.

Relationship to surrounding characterized regions

Previous studies characterizing category-selective regions along the ventral and lateral surface of visual cortex have found that these regions come in pairs, e.g for faces (FFA-fusiform face area, OFA-occipital face area), bodies (FBA-fusiform body area, EBA-extrastritate body area), general shape-selectivity (pFS-posterior fusiform, LOC-lateral occipital complex), and scenes (PPA-parahippocampal place area, TOS-transoccipital sulcus; see Schwarzlose et al., 2008). These regions are arranged in a mirrored fashion from medial-ventral wrapping around the lateral surface to medial-dorsal regions (Hasson et al., 2003), with the ventral surface showing more overall visual form information and the lateral surface showing more location-, motion-, and local shape information (Beauchamp et al., 2002; Drucker & Aguirre, 2009; Haushofer et al., 2008; Schwarzlose et al., 2008). Consistent with this pattern, we also observed big and small object regions fitting into this large-scale mirrored organization, with BigV and SmallA on the ventral surface and SmallP and BigD on the lateral surface (Fig. 5a). Thus the properties that have been discovered for the highly-selective regions may also apply to the nearby regions of less-selective object-responsive cortex, where object size predicts the large scale topographic organization of objects that have a more distributed representation.

The SmallP and BigV ROIs show proximity to and overlap with the well-known functionally localized regions of the lateral occipital complex (LOC: objects>scrambled; Grill-Spector et al., 1999) and the parahippocampal place area (PPA: scenes>objects; Epstein & Kanwisher, 1998, Fig. 5b, see Supplementary Methods). SmallP is just anterior to LOC, which is nearby and partially overlapped with other related regions such as the extrastriate body area (EBA), motion area MT, and the medial temporal gyrus tool area (MGT-TA; Beauchamp et al., 2002; Chao et al., 1999; Downing et al., 2001; Valyear & Culham, 2010). BigV is partially overlapped with PPA, a scene-selective region which has also been shown to have a reliable response to objects, particularly large objects such as buildings (Aguirre et al., 1998; Diana et al., 2008; Downing et al., 2006; Epstein, 2005; Litman et al., 2009), as well as to strongly contextual objects (Bar, 2004), which tend to be larger than non-contextual objects (Mullally & Maguire, 2011). Interestingly, both LOC and PPA have recently been showed to modulate their responsiveness with the implied size/distance of the depicted object (Amit et al., 2008; Cate et al., 2011), with object size predicting parametric activity in parahippocampal cortex (Mullally & Maguire, 2011).

Given the relationship between the PPA and the BigV region, one concern is that responses in the BigV region may be driven by scene imagery (O'Craven & Kanwisher, 2000), where bigger-sized objects are more likely to cause observers to imagine a scene than smaller objects. However, such an account would predict that imagining giant apples would lead to more scene imagery than tiny pianos, and this pattern of data was not observed in the BigV region. Additionally, given the rapid presentation rates of the visually-presented objects, we think scene-imagery is unlikely to account for the full range of data reported here (Epstein & Ward, 2010; see also Mullally & Maguire, 2011)

While it is possible to test whether these overlapping regions along lateral and ventral cortex are driven more by shape statistics, contextual associations, implied distance, or objects of different sizes, likely there is no single "high-level" description that will perfectly capture the response properties of these large regions of cortex. Under the proposal here, these regions of cortex are more likely to be involved in the processing and storage of small or large objects, where the exact nature of those stored representations across the ventral surface remains an open question.

Proposals for the topography of object knowledge

Existing proposals for large-scale organization of object knowledge in ventral cortex argued for spatial clustering by conceptual superordinate categories (e.g. food, animals, tools; Chao et al., 1999; Mahon & Caramazza, 2011; see also Weber et al., 2009), by some perceptual shape space (Ishai et al., 1999; Op de Beeck et al., 2008b) or by eccentricity biases that stem from local or global visual processing needs (Hasson et al., 2002; Levy et al., 2001; Malach et al., 2002) or high-level cognitive processes (Gauthier, 2000). Here we make the proposal that a physical parameter (the size of objects in the world) may capture some of the variance of organization and can refine and extend these featural and processing-based proposals. Why might the real-world size of objects be a natural parameterization of object knowledge?

One possibility is that the size of objects in the world gives rise to systematic biases in retinal visual experience which are extracted in early visual areas and ultimately dictate where high-level object representations will be in more anterior cortex. This idea most closely dovetails with the eccentricity-bias hypothesis of Levy, Hasson, and Malach, which proposes that high-level object representations extend anteriorly from particular eccentricities bands in early retinotopic cortex (Hasson et al., 2003; Levy et al., 2001; Levy et al., 2004; Malach et al., 2002). In support of this proposal, Levy et al., (2001) found face-selective areas tend to have a foveal-bias while scene-selective areas have a peripheral bias, and these extend from center and peripheral early visual areas respectively. This organization is consistent with recently discovered full retinotopic maps, where the fovea to periphery maps from lateral to medial ventral cortex (Arcaro et al., 2009). To apply the center-periphery organization to other categories, Malach et al. (2002) proposed that the location of different object categories might be driven by different processing resolution needs, whether it be analysis of fine-detail (foveal) or more holistic integration (peripheral), similar to arguments for process-dependent organization (Gauthier, 2000). However, this proposal does not easily lead to testable predictions about other objects until it is first determined what kinds of processing resolution an object requires (Tyler et al., 2005). Further, it has also recently been shown that scene-selective areas actually show a greater response to high spatial frequency (Rajimehr et al., 2008), inconsistent with the idea of coarser spatial processing.

Here we make a more explicit prediction that the experienced retinal size, as well as other correlated dimensions such as object curvature and spatial frequency content, are systematically related to the real-world size of the objects (with smaller rounder objects subtending smaller visual angles at typical distances than larger boxier objects). Unsupervised efficient learning mechanisms in the early visual hierarchy may extract these regularities in eccentricity and shape (Attneave, 1954; Carlson et al., 2011; Field, 1987); by more anterior stages along the visual hierarchy, the more scale- and position-tolerant visual representation in these areas may be naturally arrayed along the cortical sheet in a continuous manner, with small-object features extended from foveal cortex and large-object features extended from more peripheral cortex (see Hasson et al., 2003).

In addition to biases in incoming visual experience, requirements for subsequent processing for action and navigation might also predict that real-world size would be a valuable organizing dimension of object representation. This is similar to the connectivity-hypothesis proposed by Mahon and Caramazza, in which the organization of object representation is driven by long-range network connectivity (Mahon & Caramazza, 2011; Mahon et al., 2007): manipulable objects like tools may require different "down-stream" processing requirements than animate objects like animals.

Thus, the ventral stream organization may be driven not by biases in experience, but instead by functional-connectivity with dorsal regions subserving different action representations. The real-world size of objects naturally constrains the kinds of actions and effectors that will be used when an observer interacts with an object, and thus extends this proposal beyond animals and tools to the large range of other biological and manmade artifacts, which often get grouped together as "other objects" (Hasson et al., 2003; Op de Beeck et al., 2008a). Real-world object size is a natural proxy for the underlying continuum between manipulable and navigationally-relevant features. Finally, while the eccentricity-bias and connectivity-driven hypotheses have been discussed as competing alternatives, the real-world size of objects unifies these proposals, suggesting both bottom-up experience-driven learning and top-down requirements for subsequent actions provide convergent pressures for object knowledge to be topographically organized by real-world size.

Methods

Participants

22 healthy observers with normal or corrected-to-normal vision participated in one or more of the experiments in a 2 hr fMRI session (age 19-36, 13 female, 21 right-handed). Informed consent was obtained according to procedures approved by the MIT Internal Review Board.

MRI Acquisition

Imaging data were collected on a 3T Siemens fMRI Scanner at the Martinos Center at the McGovern Institute for Brain Research at MIT. Experiments 1 and 3 used a 12-channel phased-array head coil and Experiment 2 used a 32-channel phased-array head coil. Blood oxygenation level dependent (BOLD) contrast was obtained with a gradient echo-planar T2* sequence (33 oblique axial slices acquired parallel to the anterior commissure - posterior commissure line; 64 x 64 matrix; FoV = 256 x 256 mm; 3.1 x 3.1 x 3.1 mm voxel resolution; Gap thickness = 0.62 mm; TR = 2000 ms; TE = 30 ms; Flip angle = 90 degrees).

Experiment 1: Big and Small Object Localizer

12 observers were shown images of big real-world objects and small real-world objects in a standard blocked design. All objects were shown at the same visual angle (9x9 degrees). Each block was 16s during which 20 images were shown per block for 500 ms each with a 300 ms blank following each item. Fixation periods of 10s intervened between each stimulus block. Ten blocks per condition were shown in a single run of 8.8 min (265 volumes). A total of 200 big and 200 small distinct object images were presented. Observers were instructed to pay attention to the objects and to press a button when a red frame appeared around an item, which happened once per block.

Experiment 2: Retinal Size Manipulation

8 observers were shown blocks of big and small objects at big and small retinal sizes. The big and small objects stimuli were the same as in Experiment 1, and the retinal sizes were 11x11 degrees visual angle and 4x4 degrees visual angle for the small and big visual sizes, respectively. The blocked design and stimuli were the same as in Experiment 1: each block was 16s during which 20 images were shown for 500 ms each with a 300 ms blank following each item. Blocks were separated by fixation periods of 10s. There were four conditions (2 real-world sizes x 2 retinal sizes), presented in a pseudorandom order, such that all conditions appeared in a shuffled order 5 times per run (8.8 min, 265 volumes). Two runs were conducted in this experiment, yielding 10 blocks per condition. Observers were instructed to pay attention to the objects and to press a button when a red frame appeared around an item, which happened once per block.

Experiment 3: Mental Imagery

The names of different objects were presented aurally to 8 naïve observers, and observers were instructed form a mental image of each object. Observers' eyes were closed for the entire duration of each run. In 16s blocks, observers heard 5 object names (3.2 s per object), followed by the word "blank" signifying the beginning of each 10s blank interval. Runs always began with a 10 second blank interval. The two main conditions were blocks of small object names (e.g. "peach") and big object names (e.g. "lawn chair"). In two additional conditions, observers imagined these small objects at giant sizes (e.g. hearing the words "giant peach") and the big objects at tiny sizes (e.g. hearing the words "tiny lawn chair"). There were 30 small objects and 30 big objects, divided into two sets. Each run used the stimuli from one set and contained 3 blocks of each condition, lasting for 5.4 min (161 volumes). Six runs were conducted in the experiment, three for each object set, yielding 12 total blocks per condition. All imagery runs were conducted first, prior to the presentation of any experiments with visual stimuli, including the Big Small Object localizer.

Sounds were presented through Sensimetric MRI Compatible Insert Earphones. To set the volume levels in the scanner, a functional run was started and the volume of the stimuli was slowly increased until the participant pressed a button indicating they could hear the stimuli clearly.

Before the experiment, observers were given detailed instructions that they should imagine only isolated objects, and that "giant" versions of small objects should be imagined "as having the same size as a car or piano" while tiny versions of large objects should be imagined "as having the same size as a matchbox or something that could fit in your hand." Observers then were given a short practice run outside the scanner in which they were presented with one block each of small objects, big objects, tiny versions of big objects, and giant versions of small objects, with intervening blank periods. None of these practice object stimuli were used in the main experiment.

Experiment 4: Object Category Blocks

8 observers were shown blocks of objects by category. Each block was 16s, during which 16 object exemplars from a category were presented for 700 ms with a 300 ms blank following each item at 9x9 degrees visual angle. Fixation periods of 10s preceded and followed each stimulus block. The visual stimuli consisted of 16 different object categories, each with 16 different exemplars per category, which ranged systematically in real-world size. The 16 object categories were divided into two groups of 8, both of which contained the whole range of real-world sizes (see Supplemental Materials). In each run, 8 categories from a set were shown in a random order and repeated in the second half of the run again in a random order. This ensured that each category was seen more than once per run and occurred in both the first and second half of the run. Each run was 7.1 min (213 volumes). Four runs were conducted in the experiment, two for each image set, yielding four stimulus blocks per category. Observers were instructed to pay attention to the objects and to press a button when a red frame appeared around an item, which happened once per block.

Data Analysis

Functional data were preprocessed using Brain Voyager QX software (Brain Innovation, Maastricht, Netherlands). Preprocessing included slice scan-time correction, 3D motion correction, linear trend removal, temporal high pass filtering (0.01Hz cutoff), spatial smoothing (4-mm FWHM kernel), and transformation into Talairach coordinates. For the ROI overlap computations, analyses were performed on unsmoothed functional data in ACPC space (no talairach transform).

Statistical analyses were based on the general linear model. All GLM analyses included regressors for each experimental condition, defined as square-wave regressors for each stimulus presentation time convolved with a gamma-function to approximate the idealized hemodynamic response. A whole-brain, random affects group average analysis was conducted on data from the Big Small Object localizer (Experiment 1). A contrast was performed at an uncorrected threshold of $P < 0.001$ (cluster threshold of $100mm^3$) to test for regions more active to small vs. big objects and visa vera.

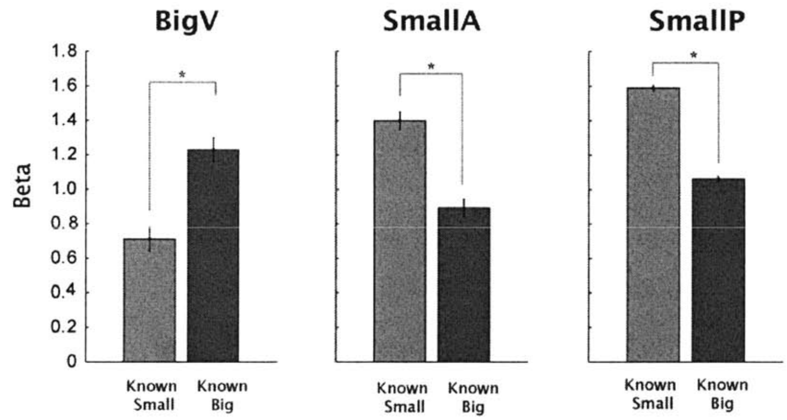
To obtain regions-of-interest from the Big Small Localizer, whole-brain GLMs were conducted for each individual. The SmallA and SmallP regions were defined from contrasts of Small>Big, and the BigV regions were defined from the opposite contrast of Big>Small. All ROIs were taken from t-maps corrected at FDR<0.05, with a cluster threshold of 10. In some cases, the threshold was made more conservative, e.g. when the SmallA and SmallP regions, which each have distinct peaks, were connected by voxels with lower t-values. If any of the targeted ROIs were not present at FDR<0.05, the threshold was lowered to FDR<0.02. If no clear ROI was present at that threshold, then that ROI was not defined for that participant. ROIs were defined as the set of contiguous voxels that were significantly activated around the peak voxel identified from within a restricted part of cortex based on the anatomical position.

For all ROI analyses, all ROIs were defined from the Big Small Localizer (independent dataset), and the response of these regions to different experimental conditions was assessed in subsequent experiments. For each subject and each ROI, GLMs were run on the average time series of the voxels in the ROI to obtain regression coefficients (betas) for the experimental conditions. For experiments with 2x2 designs (Experiment 2: Retinal size manipulation; Experiment 3: Mental Imagery), to evaluate the effects of each factor across observers, repeated-measures ANOVAs were run on the betas across observers for each ROI. For the parametric ROI analysis (Experiment 3: Category Blocks), the correlation between the real-world size rank order and the beta weights for those object categories was computed for each subject and each ROI. These r-values were Fisher's-z transformed (to be distributed normally), and t-tests were subsequently conducted to test whether these single-subject correlations were reliably different than zero.

Supporting Information

Supplementary Methods: Effect size in big and small regions

In 8 participants, we presented two runs of the big small object localizer (see Experiment 1 Online Methods). Regions of interest were defined from the first run of the big small object localizer, and the magnitude of response in each region was estimated using data from the second run of the same experiment. Each region showed a reliable differential response to objects of big and small known sizes, while also showing an above-baseline response to objects with the non-preferred size. The results are depicted in Supplementary Figure 1 and the statistics are reported in Supplementary Table 1.



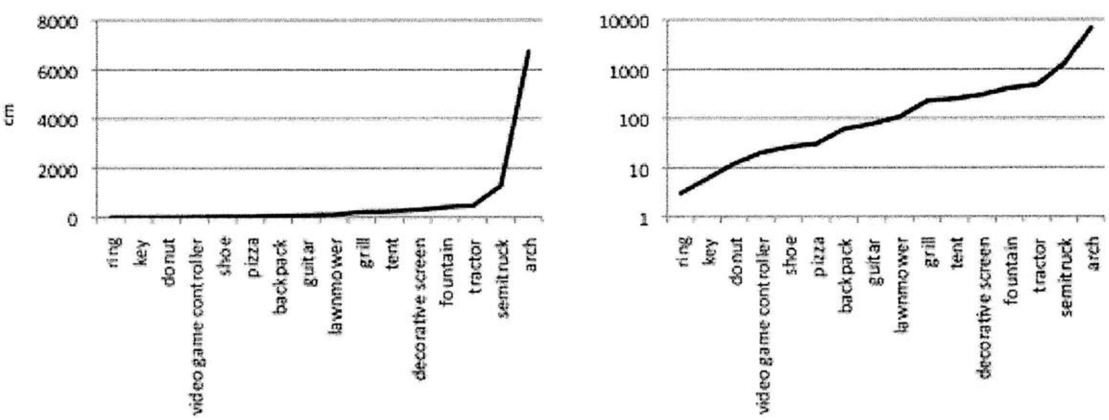
Supplementary Figure 1. Effect Size estimation from independent data. Beta weights are shown for the big- and small-preference regions in the left-hemisphere. Error bars reflect ± 1. S.E.M.

	Small Objects Mean Beta (sem)	**Big Objects** Mean Beta (sem)	Small Objects vs. Big Objects	Non-preferred vs. Baseline
BigV	.71 (.23)	1.23 (.18)	t(6)=3.54, p<0.05	t(6)=3.11, p<0.05
SmallA	1.4 (.18)	.89 (.11)	t(6)=4.61, p<0.005	t(6)=7.83, p<0.001
SmallP	1.59 (.14)	1.06 (.14)	t(7)=16.63, p<0.001	t(7)=7.57, p<0.001

Supplementary Table 1. Average beta weights for big and small objects in independently localized big and small regions in the left hemisphere.

Supplementary Methods: Category Blocks Stimuli

In the category blocks experiment, object categories were chosen based on size ranking results from Konkle & Oliva, 2011. In that study, observers sorted 100 objects into 8 bins according to their known size. The 8 size ranks reflected a logarithmic scaling of actual size in the world. In the current study, the 16 object categories used in the category blocks experiment were chosen to have 2 objects from each size rank. These 16 categories were divided into two image groups, where a single run of the experiment used images from either group 1 or group 2. Image group 1 contained keys, donuts, shoes, backpacks, grills decorative screens, tractors, and arches. Image group 2 contained rings, video game controllers, pizzas, guitars, lawnmowers, tends, fountains, and semi trucks. A graph of the actual typical size of these 16 object categories is shown in Supplementary Figure 2.



Supplementary Figure 2. Known object size of stimuli in Experiment 2: Category Blocks. The typical real-world size of 16 different objects categories was measured or estimated based on internet searches. Here actual size was characterized as the linear distance between opposing corners of the 3D bounding box (i.e sqrt$(height^2 + width^2 + depth^2)$). Left: Plot of the actual size of objects with a linear scale. Right: Plot of the actual size of objects with a log scale. See Konkle & Oliva, 2011.

Supplementary Methods: Single Subject Correlations in Category Blocks Experiment

The big and small regions showed a parametric modulation of overall activity based on the real-world size of the object category. This effect was reliable both at the group level and single subject level.

At the group level, beta values were computed for each category from an ROI GLM for each subject, the beta values were averaged across subjects, and then the correlation was computed. Activation increased with increasing known size in the big region (BigV: $r = 0.69, p = 0.003$), and decreased in the small regions, as predicted (SmallA: $r = -0.67, p = 0.005$; SmallP: $r = -0.58, p = 0.017$).

At the single subject level, Pearson's r was computed for each subject, and Fischer-z transformed to follow a normal distribution, and then t-tests of these single subject correlation coefficients were conducted to test for difference from zero. We observed a significant modulation of activity in the predicted directions for the big and small regions (BigV: mean $R = 0.51, t(7) = 8.1, p < 0.001$; SmallA: mean $R = -0.46, t(4) = -4.7, p = 0.009$; SmallP: mean $R = -0.36, t(7) = -4.0, p = 0.005$). This also held when the Spearman rank-order correlation was computed for each subject. Supplementary Table 2 indicates for each region the number of subjects who showed modulation of activity in the predicted direction and in the unpredicted direction.

	# subs with correlation in predicted direction	# subs with significant correlation Predicted direction	#subs with significant correlation in unpredicted direction
BigV-L	8/8	6/8	0/8
BigV-R	8/8	6/8	0/8
SmallP-L	7/8	2/8	0/8
SmallP-R	5/5	3/5	0/5
SmallA-L	5/5	2/5	0/5

Supplementary Table 2. Summary of parametric modulations across subjects for Experiment 4: category blocks. We predicted increasing activation with increasing size for the big region, and decreasing activation for the small regions. Nearly all subjects showed modulations in the predicted direction with most showing a significant correlation at the single subject level. No subjects showed significant modulations in these regions in the opposite direction.

Supplementary Methods: Characterizing Reliability and Overlap of ROIs

The Small and Big ROIs showed proximity to and overlap with the well-known functionally localized regions of the lateral occipital complex (LOC: objects>scrambled; Grill-Spector et al., 1999) and the parahippocampal place area (PPA: scenes>objects; Epstein & Kanwisher, 1998). To characterize the overlap between these areas, we first gathered data from a new set of 7 observers on two runs of the Big and Small ROI localizer and two runs of a PPA and LOC localizer described below. We then computed overlap between two ROIs, e.g. LOC and SmallP, and compared it to the overlap of the ROI to itself across runs. Overlap was characterized using a procedure modified from previous methods (Scholz et al., 2009).

PPA and LOC were localized from a standard localizer experiment in which stimulus blocks of scenes, objects, faces, and scrambled objects were shown, with each block lasting 16s during which 20 images were shown for 500ms each with a 300 ms blank (images shown at 9x9 degrees visual angle). Fixation periods of 10s preceded and followed each stimulus block. The conditions were presented in a pseudorandom order, such that all 4 conditions appeared in a shuffled order 4 times per run. A run was 7.1 min (213 volumes). Observers were instructed to press a button when a red frame appeared around an item, which happened once per block.

Overlap between two target regions was computed over a range of t-value thresholds then averaged, allowing for regions to be different sizes. For any two regions being compared, the range of t-values started at the maximum t-value of the two region's FDR<0.05 threshold, and increased by steps of 0.02 to the lowest of the two peak t-values. Additionally, we required a minimum of 10 voxels and a maximum of 500 voxels from both ROIs at any given threshold. The analysis proceeded by getting the contiguous set of voxels around the peak voxel that were above the specific threshold within an anatomically defined mask. At each threshold, degree of overlap was quantified as the percent of voxels of the smaller region were contained in the larger region, for left hemisphere ROIs only. This measure can be conceptualized as what percent the smaller region is contained in the bigger region without relying on a specific arbitrary t-threshold.

The overlap analysis showed that the LOC region across two runs was 72% (SEM=6%) contained with itself, and the SmallP region was 81% contained with itself (SEM=11%). Given these numbers as reference of within-ROI reliability, the SmallP was on average 35% contained in the LOC region (SEM 6) (See Supplementary Table 3). Thus while there is some overlap between LOC and SmallP, the regions which show a preference for small objects are not capturing the same region as is localized with objects>scrambled.

The PPA region across two runs was 85% contained (SEM=8%) and the Big region across two runs was 90% contained (SEM = 3%). Comparing these two regions together, the PPA and Big regions were on average 58% contained (SEM=11%) (See Supplementary Table 3). On average there was relatively more overlap between PPA and Big regions then there was with LOC and SmallP regions.

	BigV1 BigV2	PPA1 PPA2	BigV1 PPA2	PPA1 BigV2	SmP1 SmP2	LOC1 LOC2	SmP1 LOC2	LOC1 SmP2
Sub1	0.91	0.81	0.21	0.01	0.95	0.87	0.52	0.80
Sub2	-	0.98	-	-	1.00	0.93	0.00	0.27
Sub3	0.81	0.86	0.38	0.95	0.75	0.57	0.26	0.46
Sub4	-	0.92	0.19	-	0.32	0.59	0.36	0.19
Sub5	0.98	0.93	-	0.91	-	0.69	0.61	-
Sub6	0.89	0.93	0.68	0.97	0.99	0.78	0.14	-
Sub7	0.93	0.36	0.84	0.44	0.86	0.58	0.27	0.35
Mean	**0.90**	**0.83**	**0.46**	**0.66**	**0.81**	**0.72**	**0.31**	**0.41**
SEM	0.03	0.08	0.56	0.11	0.11	0.06	0.08	0.11

Supplementary Table 3: Average % containment between a region with itself across runs and with the comparison region across runs. For example, BigV1 indicates the big-preference region (big>small) defined from the first functional run; BigV2 indicates the big-preference region defined from the second functional run. SmP1 and SmP2 indicates the SmallP region (small>big) defined in the first and second runs.

Chapter 5

Conclusions

In this thesis, I proposed that the real-world size is an intrinsic part of object representation, with behavioral consequences for perceptual tasks, and neural consequences for the spatial topography of object knowledge. Here, I summarize the empirical findings and suggest that they can be coherently explained by thinking about the experience of an observer situated in a three-dimensional world. I introduce this *situated observer framework*, and highlight the implications for models of object representation and the organization of cortex.

Summary of Empirical Findings

Real-world objects have a canonical visual size

In Chapter 2, we observed that when people draw or imagine objects, requiring them to access their perceptual object representations, the visual size at which objects are elicited depends systematically on known size. Additionally, when people resize pictures of objects to look best (putatively to provide the best "representational fit", Palmer, Schloss, and Gardner, in press), they scale objects to be proportional to the log of the known size. Thus, when accessing existing object representations to perform a variety of tasks, there is a systematic relationship between the visual size of the objects and their real-world size. Akin to the previous literature showing objects

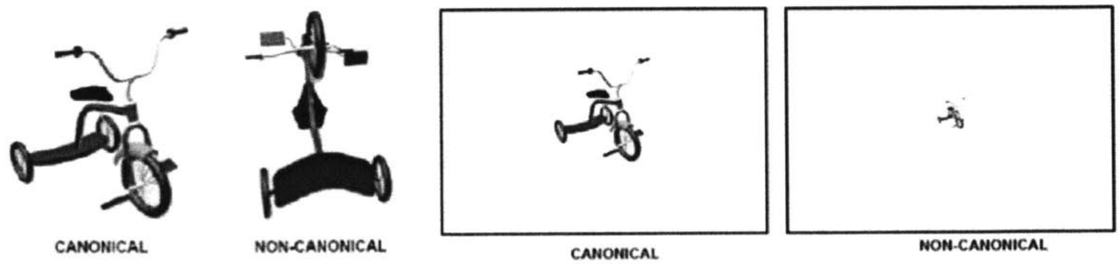


Figure 1: Canonical and non-canonical perspective of a tricycle, from Palmer, Rosch, & Chase, 1981. Canonical and non-canonical visual size of the tricycle, from Konkle & Oliva, 2011.

have a canonical perspective, this suggests that objects also have a *canonical visual size* (Figure 1).

How does canonical visual size influence models of object representation? Objects can be recognized across a range of visual sizes, and this visual-size invariance has led to the inference that object representations do not contain visual size information (e.g. Biederman & Cooper, 1992). However, we suggest that visual size information *is* stored for each object, and this visual size information is systematically related to the known size of objects in the world.

Real-world size processing is automatic

When we are presented with an object, we can automatically and rapidly recognize it and name its category. In Chapter 3, we found that our knowledge of real-world size is also accessed automatically during object processing: when you see a familiar object, you not only know what it is, you also know how big it is in the world.

In our familiar-size Stroop task, two real-world objects of big and small known sizes (e.g. A piano and an apple) were presented side by side on the screen at two different *visual (or retinal)* sizes. Observers' task was to make judgments about the visual size (e.g. Which is bigger on the screen?). Object identity and real-world size was irrelevant to the task. Nevertheless, we observed a familiar-size stroop effect: for example, when the piano was presented bigger than the apple (congruent trials), observers were faster to make visual size judgments than when the known size

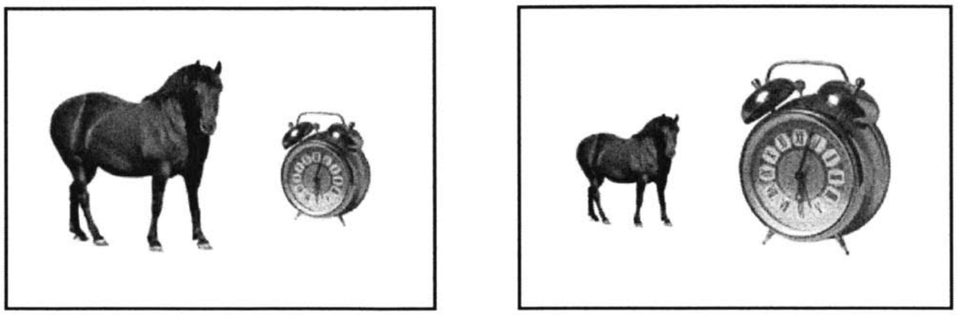


Figure 2: Example congruent and incongruent displays for the familiar-size Stroop task

mismatched with the retinal size (incongruent trials; Figure 2). We showed that this effect is not cognitively penetrable, but instead likely arises from richer knowledge about real-world size acquired over time.

One of the defining characteristics of object recognition is that we can recognize the category of an object over different positions, viewpoints, and visual sizes. However, even though we can recognize an apple whether it's depicted on a stamp or a billboard, these data suggest that in doing so we also automatically access its typical size in the world. The results suggest that our perceptual representations of objects not only give rise to automatic categorization, they also give rise to automatic real-world size information. This provides support for the claim that the real-world size of objects is a basic and fundamental property of object representation.

Real-world size predicts the spatial organization of object representation

In Chapter 4, we examined the impact of real-world size on the neural organization of object knowledge. Specifically, we examined the representations of everyday objects excluding faces, animals, and classically-defined tools, and asked whether there were any regions that were differentially selective to objects of big and small known size. Our findings showed a consistent organization of big and small objects in medial to lateral regions of ventral temporal cortex in the left-hemisphere. This was mirrored

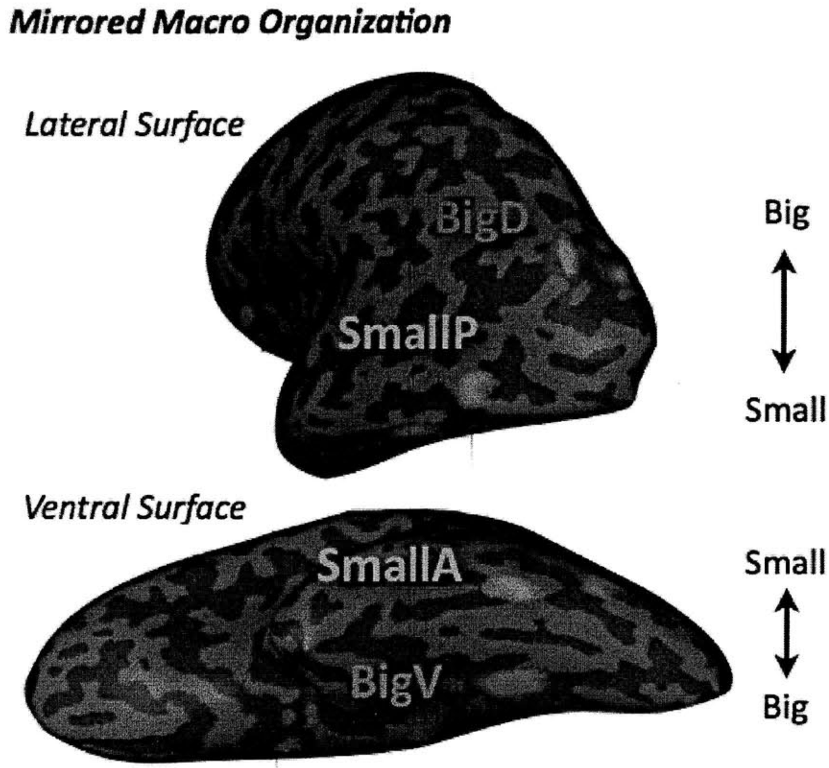


Figure 3: Evidence for a mirrored macro organization. Along the ventral surface of the temporal lobe, there was medial to lateral organization of preference for big to small objects (BigV to SmallA). This organization was mirrored on the lateral surface (SmallP to BigD).

along the lateral surface, with small to big organization moving from lateral to medial. Thus, consistent with ideas of Hasson et al. (2003), we find support for a *mirrored macro organization* of object knowledge (Figure 3).

Several follow-up experiments characterized the responses of these big and small preference regions, showing the underlying representations are tolerant to changes in visual size, are activated during mental imagery, and are active for a range of different object categories of the preferred size, and thus not solely driven by objects like monuments or tools. It is important to note that in general, these objects drive a large above-baseline responses across ventral cortex. Thus these regions are not highly-selective modules for big objects only or small objects only, but instead show a systematic differential response.

The current state of understanding about the organization of object representation across cortex is that there are islands of category selectivity amongst a sea of heterogenous (but systematic) responses whose organization is unknown. The results of Chapter 4 show evidence that real-world size is a coarse organizing dimension underlying these distributed activation patterns of different object categories. It will be exciting to more closely examine whether previously established category-selective regions are arranged within this broader size-based organizational scheme, as the areas for faces and bodies typically fall in between the Small and Big regions of selectivity (see also Chapter 4, Figure 5).

A Situated-Observer Framework

The empirical results of this thesis argue for a critical role of real-world size for object representation. This novel approach arises from the observation that objects are fundamentally physical entities—our experience with objects, both in our lifetime and over evolutionary time, arises as situated observers in a three-dimensional world. Here I outline the critical points of this *situated-observer framework* and discuss the implications of the empirical findings from within this unifying context.

The three-dimensional world structures the shape, size, and various features of objects (Haldane, 1928; see Appendix): natural objects in the world have optimized their shape and sizes to be coordinated with their biological infrastructure and functions. This implies that there are systematic characteristics of objects based on their physical size. Additionally, the three-dimensional world structures how we experience these objects: we observe and interact with the world as observers standing from a height above the horizon, with arms of a certain length, and bodies of a certain size. The distributions of object information projected on our retina are naturally structured based on these geometric constraints, giving rise to biases in experienced retinal size, eccentricity, elevation, spatial frequency, and orientation based on object size.

Furthermore, we adopt the prominent theory that the visual system is tuned to the statistics of the natural world (Gilbert et al. 2001, Simoncelli & Olshausen

2001, see also Kourtzi & Conner, 2011). Cortical circuits extract the covariance of features in their input in order to form efficient representations of incoming visual information (Attneave, 1954; Carlson et al., 2011; Field, 1987). Supporting evidence for this comes from the tuning of V1, whose Gabor-like features can be derived from a computationally optimal representation of natural scene statistics (Olshausen & Field, 1996).

Combined, these assertions can provide a coherent framework explaining why we have a canonical visual size for different sized objects, and how the cortex comes to be organized by real-world size, and why some categories of objects may have clustered regions while others have more distributed regions. It is important to point out that experience was not manipulated in any of the current studies; rather we examined the representations of objects an adult human observers, which have been acquired and established over a lifetime of visual experience, with visual tuning priors established over evolutionary time. However, this experience-based framework allows us to make testable predictions about why object representation is organized this way and may subsequently provide new insights for models of object representation.

Why do objects have a canonical visual size?

Within the situated observer framework, canonical visual size is most naturally explained by the most experienced retinal size. If the visual system naturally learns and stores the regularities between the real-world size of objects and experienced retinal size, then a strong prediction is that the canonical visual size reflects the mode of experienced retinal size. While geometric inferences can be made that small retinotopic envelope sizes are typical for small objects in typical viewing conditions and large retinotopic envelope sizes are typical for very large objects, characterizing the actual distributions has not yet been done.

What drives the real-world size topography of occipito-temporal cortex?

A prominent proposal for what drives the organization of occipito-temporal cortex is the eccentricity-bias hypothesis (Hasson et al., 2003; Levy et al., 2001; Malach et al., 2002). On this account, eccentricity mapping is observed even in higher level regions, with some regions showing peripheral biases and other regions showing foveal biases. They propose that some object categories require global processing and will be represented in peripherally-biased regions, whereas other object categories require local processing and will be represented in foveally-biased regions. Thus the organization of object representation is driven by eccentricity biases of different object categories. However, one of the major criticisms of this account is that it is not clear what kind of processing is required for different objects (e.g. if faces are processed holistically, that might predict a more peripheral, global processing location; See Tyler et al., 2005).

In the situated-observer framework, however, the real-world size of objects is correlated with experienced eccentricity: small objects are more likely to appear within the fovea, whereas large objects tend to extend into the periphery and to appear peripherally. Based on these retinal size statistics, small objects will drive more foveal visual cortex while big objects will drive more peripheral visual cortex. Thus, real-world size topography is a natural consequence of size-induced retinal biases in visual experience. One strength of the account proposed here is that it makes specific predictions about the distribution of activity for any object category, based on systematic biases in retinal size, and not processing-demands per se. For example, future work can test whether the spatial distribution of activation for any arbitrary object category can be predicted from the natural distribution of experienced retinal sizes for that category, or whether experimentally manipulating the distribution of experienced retinal sizes (e.g., for novel objects) would affect the spatial distribution of activation.

Additionally, real-world size is correlated with other low-level features, e.g. spatial frequency, orientation and elevation (see Appendix). This suggests that the kind of early shape features extracted in foveal cortex will be more suited for efficiently representing small objects, and similarly for big objects in more peripheral cortex. As a consequence of the large-scale eccentricity organization, these pre-cursor object representations may naturally be arrayed along the cortical sheet by real-world size.

Why do only some object categories have focal regions?

One potential answer to this question is that some objects may have a tighter range of experienced visual sizes than others. Additionally, in a specific band of eccentricity, visual input may be dominated by one (or a few) kinds of objects (Figure 4). Indeed, the frequency of experience, and in particular early experience, has been argued to affect the degree of spatial clustering of category-specific features, for example in the visual-word form area (McCandliss, Cohen, & Dehaene, 2003), and a similar symbol-selective area in monkeys (Srihasam & Livingstone, 2011). On such an account, faces, bodies, and word strings are some of the most frequent categories in our visual experience; an efficient representational system might devote more representational space to these categories.

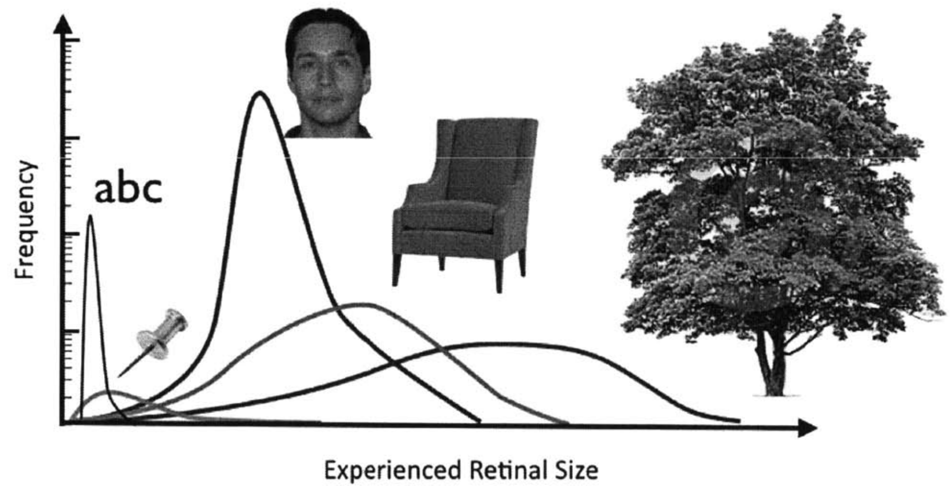


Figure 4: Hypothetical distributions of for the frequency of different experienced retinal sizes, for letters, thumbtacks, faces, chairs, and trees.

Interestingly, the proposed framework here also provides a hypothesis about where these patches of more selective cortex should be located. For example, faces have a particular range of experienced visual angle (larger than the smallest objects like coins and berries, smaller than large objects like chairs, and tables), at which they are likely the dominantly experienced object category. This may predict why the face-selective regions of cortex are extended from more medium eccentricities (rather than requiring a deformation step proposed in Hasson, 2003; see also Tyler et al., 2005). A similar argument can be made for letter strings. Further data is needed to examine these specific predictions.

Real-world size as a proxy for the underlying representations

What do these results tell us about the underlying representations in these regions? This organization raises a fundamental question that has not been addressed in previous research on object representation: what size-specific visual properties are used to represent objects? Future research can address this question by investigating what is being represented in the big and small regions. For example, a focus on this dimension makes sense of a number of low-level biases that have been observed in high-level areas (see Appendix), and future work examining how shape changes as a function of real-world size may be a valuable new approach to thinking about shape parameterization. In other words, the work presented in this thesis suggests that the real world size of objects is a proxy for certain features underlying object representations, and there future research focusing on these size-specific properties can provide new insight into the nature of these underlying object representations.

Conclusion

The contributions of this thesis are to establish that real-world size is a core parameter of object representation, as evidenced by its role in automatic object recognition processes and the large-scale spatial topography of object knowledge in the visual cortex. I have proposed a situated observer framework within which to understand the role of real-world size in object representation and the organization of object knowledge in the cortex. The visual experience of an active observer in the three-dimensional world is systematically biased according to the real-world size of objects. An efficient representational system must take advantage of such systematic biases for optimal encoding. This work provides testable predictions about retinal size biases in visual experience, the role of real-world size in object representation, and an approach in which to understand the neural representation of any object in the world.

Appendix: Consequences of vision in a three-dimensional world

Characterizing the statistics of visual experience is a challenging endeavor, usually requiring head-mounted cameras and advanced image processing techniques. However, some biases in the retinal experience for different sized objects can be reasonably inferred from the geometric constraints of an observer situated in a three-dimensional world.

Here I discuss several dimensions of visual experience which likely have systematic differences for big and small objects. These dimensions include (i) retinal size and eccentricity, (iii) spatial frequency and orientation, and (iv) upper/lower visual field position.

Next, I review neuroimaging results that show low-level visual biases in relatively high-level object areas beyond extrastriate cortex. In general these different low-level visual features tend to co-occur and drive object-responsive regions. I discuss these in the context of eccentricity-bias proposals of object representation, which fail to correctly predict the convergence of these low-level biases.

Retinal Size and Eccentricity

For this thesis, the most critical dimension of natural visual experience that needs to be characterized is the relationship between the physical size of objects in the world, viewing distance, and the subsequent distribution of experienced retinal sizes (e.g. See Chapter 1, Figure 2). While this has not been characterized broadly across object categories, we can make a reasonable inference about typical visual sizes of experience for the two extremes of real-world size—tiny and giant objects.

In these edge cases, it is likely that very small objects are typically experienced at smaller visual angles than larger objects. A paperclip subtends 3 degrees at arm's length–at distances beyond that it is typically occluded and not the focus of attention, and paperclips are rarely experienced closer than that. For a paperclip to subtend 30 degrees, for example, it would have to be held a couple inches from the eye. Big objects like cars and houses, typically subtend 30+ degrees visual angle. At farther distances they too are often occluded by intervening objects (to see a car at 3 degrees it has to have a vast unblocked expanse in front of it). At closer distances you enter the part-hierarchy of the object, where now what you are actually seeing is a car door or handle, and the overall car becomes the context rather than the object of attention.

It is an open question if this relationship between known size and visual size holds systematically for entire range of real-world size. A strong prediction from canonical visual size is that on average across objects, there is a general logarithmic relationship between the physical size of the objects in the world and their most typical viewing distance and thus their most experienced retinal size (Figure 1). Based on the reasoning above, this function would arise because of two different limitations: at far distances objects of any size become occluded, this happens at closer distances for a paperclip than it does for a car; at close distances the entire object itself ceases to be the focus of attention and instead parts of the object are the focus.

We can recognize objects at a large range of visual sizes, assuming they are centrally-fixated. However, because acuity falls off with eccentricity, in order to rec-

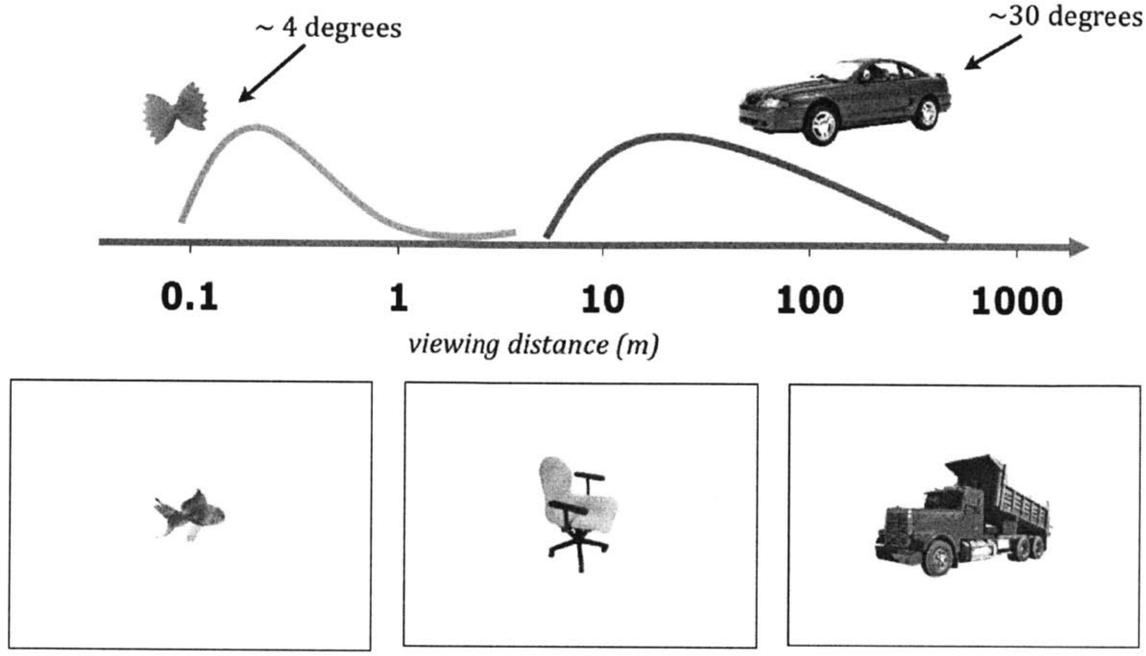


Figure 1: Top: hypothetical distributions of viewing distance for small and large objects. Bottom: examples for three how visual size might scale with log real-world size.

ognize something in the periphery, it has to subtend a larger visual size that increases proportional to the eccentricity (e.g. Rovamo & Virsu, 1978). In natural experience, when we are fixating an object, there will be other object information in the periphery. Assuming the retinal size biases described above, it follows that bigger objects in the world are more likely to be recognized in the periphery than smaller objects because they tend to subtend larger visual sizes. Thus retinal size biases can also produce eccentricity biases in object recognition experience for objects of different real-world sizes.

Shape Biases, Spatial Frequency, and Orientation

Other correlated visual features regarding the shape of objects are evident when considering that objects have to withstand gravity in the world. In Haldane's 1928 essay "On being the right size", he elegantly explains that there are structural reasons why living things look the way they do: you cannot simply scale up a mouse to be

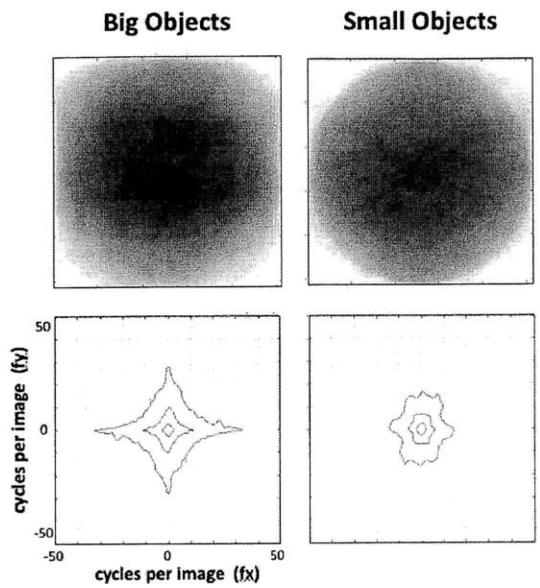


Figure 2: Analysis of small and big object image statistics. Top: Average image from 200 images of big objects (left) and 200 images of small images (right). Bottom: Average Fourier spectrum across these image sets. Sections represent 60, 80, and90% of the image energy. The Fourier representation is a polar plot: Spatial frequency is reflected in the radial dimension with low spatial frequencies near the origin and high spatial frequencies moving outward from the center; Orientation is mapped by polar angle, with vertical at 0 degrees and horizontal at 90 degrees.

the size of an elephant because the bones would break, the lungs would not deliver enough oxygen, and the arteries would not withstand the increased pressure needed to pump blood. In general, if an object is scaled up in physical size keeping the same material and construction, the mass of the object increases by the cube of the scaling factor while the surface area only increases by the square of the scaling factor (the square-cube law). Natural objects tend to have maximized this tradeoff, achieving a shape that's not overly strong (or weak) for their size (Gordon, 1981). This implies that there is systematic covariation between shape and real-world size.

What are the shape features that covary with real-world size? A very coarse answer to this question comes from a inspecting the average Fourier spectra of big and small objects (Figure 2). In general, big objects have more power at high spatial frequencies than small objects. Further, big objects have relatively more power at meridian orientations than oblique orientations, while small objects have relatively equal power at meridian and oblique orientations. Put simply, big objects are boxier and small objects are rounder.

Upper-Lower visual field biases

A well known challenge of vision is to inferring the size (and shape) of an object from an impoverished 2-dimensional projection: a small object at a close distance can subtend the same visual angle as a large object of the same shape, at a far distance. However, it is quite challenging to construct a real-world demo of this problem of infinite solutions, in which there was an atypically-sized object (e.g. an extra-large chair) hidden amongst a scene of regular objects, where the entire scene that looked normal in the two-dimensional projection. Figure 3 shows a simple scene with a normal sized desk and chair, as well as a chair that is twice the size and farther away from the camera in order to subtend the same visual angle. Notice that the two projected views of the scene are not the same. In fact, the large-far away chair is higher in the picture. It looks either like it is, in fact, large and far away, or like it is typically-sized but floating.

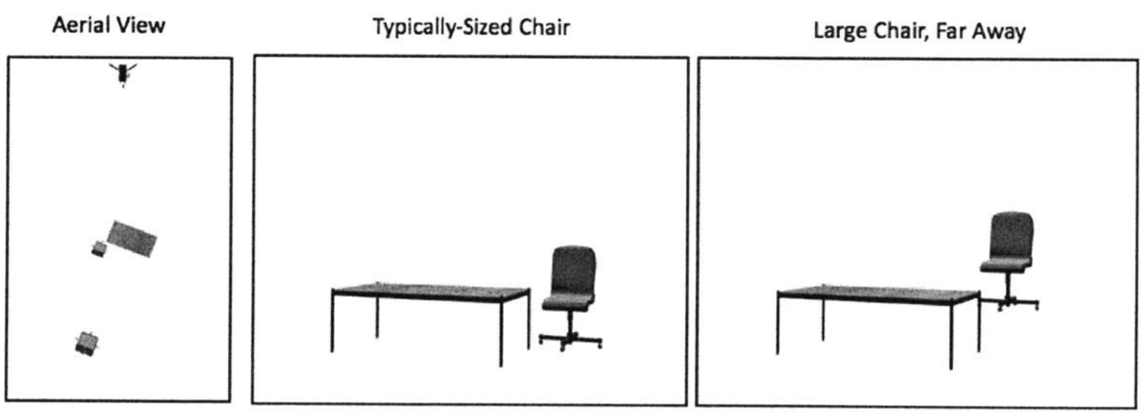


Figure 3: Left: Aerial view with a normal chair and a large chair placed father away from the camera so as to subtend the same visual angle. Middle & Right: Views of the typical and large chair.

What is going on here? The answer is that when we view the world, we view it from an average height of 5.5 ft (1.67 m). It is in fact possible to make the typically-sized chair and desk scene project the same view as the large chair far from the desk. To do so, you need to adopt a "crocodile" view, in which the center of your eye is at exactly ground level The same scene with a crocodile view is shown in Figure 4.

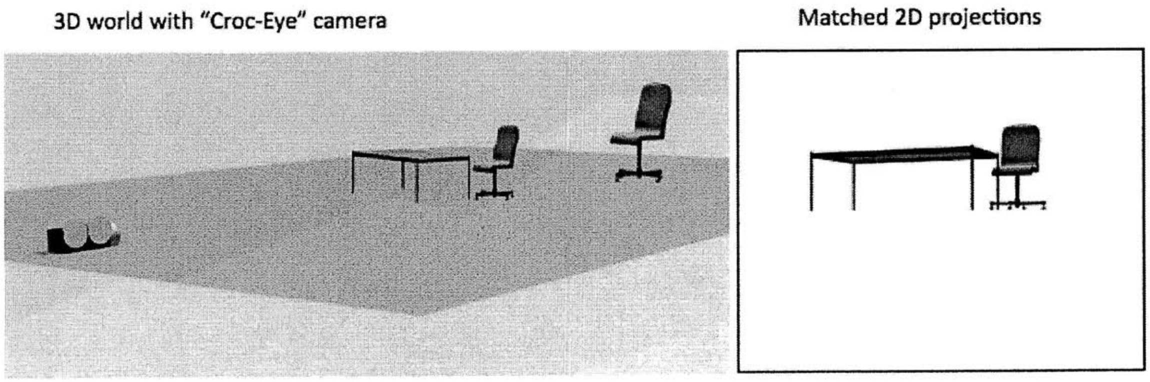


Figure 4: Croc-eye view. In order for the two chairs and desk to cast the same retinal projection, the camera needs to be placed at ground level. The matched 2D projection from this view is shown on the right.

A direct consequence of viewing the world from a height above the croc-eye view is that the farther objects are away from us, the higher in the visual field they appear. Recall that bigger objects tend to be experienced at larger distances than smaller objects. It follows that big objects tend to be experienced higher in the visual field than small objects.

Correlated low-level features drive high-level regions

Interestingly, these correlated low-level visual biases described above are evidenced in the response properties of high-level object-responsive areas.

The visual biases for *small objects* have been reported to drive lateral regions of visual cortex. For example, in area LOC, which overlaps slightly with the small posterior region, there is a lower-visual field bias (Carlson, et al., 2010; McKyton & Zohary, 2007; Sayres & Grill-Spector, 2008; Niemeier, et al., 2005). Lower field biases are also present in the extrastriate body area and fusiform face area (Schwarzlose et al, 2008). LOC and surrounding cortex also has a foveal bias (Sayres & Grill-Spector, 2008; Arcaro et al, 2009; Levy et al., 2001; Hasson et al., 2003). LOC also shows a preference for concave rather than convex shapes (Haushofer et al, 2008), which is consistent with the simple rounder/boxier shape statistics suggested here.

The systematic visual biases of *big objects* tend to drive responses in medial ventral and medial occipito-parietal areas. For example, both the PPA and TOS show an upper-visual field preference (Schwarzlose et al, 2008). PPA and BigV have a peripheral eccentricity bias (Arcaro et al, 2009; Levy et al., 2001; Hasson et al., 2003), and a larger retinal size bias (Chapter 4). While there is less work specifically examining shape representation in PPA, it has been recently been shown that PPA responds more to cubes than spheres, and more to high spatial frequency than low spatial frequency (Rajimehr, 2011).

Overall these data suggest that in object-selective cortex, along the ventral and lateral surface, there are a number of consistent low-level biases that co-occur and can be explained by systematic differences in visual experience for big and small objects. Interestingly, in most of the paper reporting these biases, their co-occurrence is not highlighted or explained, though most appeal to natural statistics of experience to account for them. In none of them is the real-world size of objects, eye-height, or viewing distance mentioned.

The main framework in which these visual biases are interpreted is the eccentricity-bias proposal of Malach, Levy, and Hasson (Hasson et al., 2002; Levy et al., 2001; Malach et al., 2002). In this framework, they suggest that some object categories may require more local processing (faces) and other categories may require more global/holistic processing (houses). They propose that the location of these category selective areas can be explained by extensions of early foveal and peripheral cortex. However, this account does not predict upper/lower visual field biases, and it reverses the spatial frequency predictions. This failure of the eccentricity organization to account for the particular co-occurrences of low-level visual biases has been pointed to as a weakness of this account (Schwarzlose, 2008; see also Tyler et al., 2005). Here we suggest that the geometric constraints of an observer in the world unifies these low-level biases and is consistent with the responses in occipito-temporal cortex. High-level areas show responses to low-level features because these are generally correlated during natural visual experience, and there is no gain in efficiency of representation to de-correlate the high-level from the low-level responses.

Summary

There are a number of low-level visual statistics that are highly correlated with objects of different sizes in natural experience. These include, but are not likely limited to, retinal size and eccentricity, spatial frequency and orientation, shape curvature, and upper/lower visual field position. These low-level features drive responses in more anterior visual cortex including the category-selective areas. While this makes it difficult to put a simple label that fully characterizes the response properties of a high-level region, it does resonate with the prominent assumption that the visual system is tuned to the statistics of natural experience. Moving forward it will be valuable to empirically characterize the distribution of experience along various visual dimensions, as this can provide leverage to understand covarying structure extracted by different stages of the visual system.

References

Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1998). An area within human ventral cortex sensitive to "building" stimuli: evidence and implication. *Neuron 21*(2):373-83.

Amit, E., Trope, Y., & Yovel, G. (2008). A distance principle of organization of the ventral visual stream. *Journal of Vision, 8*(6):329.

Arcaro, M. J., McMains, S. A., Singer, B. D., & Kastner, S. (2009) Retinotopic organization of human ventral visual cortex. *Journal of Neuroscience, 29*(34):10638-52.

Ashbridge, E. & Perrett, D.I. (1998). Generalizing across object orientation and size. In V. Walsh and J. Kulikowski, (Eds.), Perceptual Constancy, (pp. 192-209). Cambridge University Press.

Attneave F. (1954). Some informational aspects of visual perception. *Psychological Review, 61*(3):183-93.

Baird, J. C. (1963). Retinal and assumed size cues as determinants of size and distance perception. *Journal of Experimental Psychology, 66*(2), 155-162.

Baird, J. C. (1997). Sensation and judgment: Complementary theory of psychophysics. Mahwah, NJ: Erlbaum.

Bar, M. (2001). Viewpoint Dependency in Visual Object Recognition Does Not Necessarily Imply Viewer-Centered Representation. *Journal of Cognitive Neuroscience, 13*(6), 793-799.

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience, 5,* 617-629.

Beauchamp, M. S., Lee, K. E., Haxby, J. V., & Martin, A. (2002). Parallel visual motion processing streams for manipulable objects and human movements. *Neuron 34*, 149 -159.

Biederman, I., Mezzanotte, R.J., & Rabinowitz, J.C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology, 14,* 143-177.

Biderman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94*(2), 115-147.

Biederman, I., & Cooper, E. E. (1992). Size invariance in visual object priming. *Journal of Experimental Psychology: Human Perception and Performance, 18*(1): 121-133.

Blanz, V., Tarr, M. J., & Bulthoff, H. H. (1996). What object attributes determine canonical views? *Perception, 28,* 575-599.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 433-436.

Brigell, M., Uhlarik, J., & Goldhorn, P. (1977). Contextual influences on judgments of linear extent. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 105-118.

Blthoff, H. H., & Edelman, S. (1992). Psychophysical support for a two-dimensional view interpolation theory of object recognition. *Proceedings of the National Academy of Sciences, USA, 89,* 60-64.

Carlson, E. T., Rasquinha, R. J., Zhang, K., & Connor, C. E. (2011). A Sparse Object Coding Scheme in Area V4. *Current Biology, 21*(4):288-93.

Carlson, T. A., Schrater, P., & He, S. (2003). Patterns of activity in the categorical representations of objects. *Journal of Cognitive Neuroscience, 15*(5):704-17.

Carlson, T., Hogendoorn, H., Fonteijn, H., Verstraten, F. A. J., (2010). Spatial coding and invariance in object-selective cortex. *Cortex, 47,* 14-22.

Carmichael, L. Hogan, H. P., & Walter, A. A. (1932). An experimental study of the effect of language on the reproduction of visual perceived form. *Journal of Experimental Psychology, 15,* 73-86.

Cate, A. D., Goodale, M. A, & Kohler, S. (2011). The role of apparent size in building-and object-specific regions of ventral visual cortex. *Brain Research, 1388,* 109-22.

Cave C. B., Squire, & L. R. (1992). Intact and long-lasting repetition priming in amnesia. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 18,* 509-20.

Cavanagh, P. (2011). Visual Cognition. *Vision Research.* Article in press.

Chao, L. L., Haxby, J. V., & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature Neurosciene, 2*, 913-19.

Cohen L, et al. (2000). The visual word form area: Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain, 123*(2), 291-307.

Cox DD, Savoy RL (2003) Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. Neuroimage 19(2):261-270.

Davenport, J. L., & Potter, M. C. (2004). Scene consistency in Object and Background Perception. Psychological Science, 15(8), 559-564.

DeLoache, J. S. (2004). Scale Errors Offer Evidence for a Perception-Action Dissociation Early in Life. Science, 204, 1027-1029.

Diana RA, Yonelinas AP, Ranganath C (2008) High-resolution multi-voxel pattern analysis of category selectivity in the medial temporal lobes. Hippocampus 18(6):536-41.

DiCarlo JJ, Cox DD (2007) Untangling invariant object recognition. Trends Cog Sci 11(8):333-41.

Downing P, Chan AWY, Peelen M, Dodds C, Kanwisher N (2006) Domain specificity in visual cortex. Cereb Cortex16(10):1453-61.

Downing PE, Jiang Y, Shuman M, Kanwisher N (2001) A cortical area selective for visual processing of the human body. Science 293:2470 -2473.

Drucker DM, Aguirre GK (2009) Different spatial scales of shape similarity representation in lateral and ventral LOC. Cereb Cortex 19(10):2269-80.

Edelman, S., & Blthoff, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. Vision Research, 32(12), 2385-2400.

Epshtein, B., & Ullman, S. (2005). Feature hierarchies for object classification. IEEE Computer Society, 1, 15505499.

Epstein, R. & Kanwisher, N. (1998). A cortical representation of the local visual environment. Nature, 392(6676), 598-601.

Epstein R. (2005) The cortical basis of visual scene processing. Vis Cogn 12(6):954-78.

Epstein RA, Ward EJ (2010) How reliable are visual context effects in the parahippocampal place area? Cereb Cortex 20(2):294-303.

Epstein, W. (1963). The Influence of Assumed Size on Apparent Distance. The American Journal of Psychology, 76(2), 257-265.

Epstein, W. & Baratz, S. S. (1964). Relative Size in Isolation as a Stimulus for Relative Perceived Distance. Journal of Experimental Pscyhology, 67(6), 507-513.

Esptein, W. (1965). Nonrelational Judgments of Size and Distance. The American Journal of Pscyhology, 78(1), 120-123.

Field DJ (1987) Relations between the statistics of natural images and the response properties of cortical cells. J Opt Soc Am A. 4(12):2379-94.

Fiser J, Biederman I (1995) Size invariance in visual object priming of gray-scale images. Perception, 24(7): 741-748.

Freeman J, Brouwer GJ, Heeger DJ, Merriam EP (2011) Orientation Decoding Depends on Maps, Not Columns. J Neurosci 31(13):4792-804.

Furmanski, C. S., & Engel, S. A. (2000). Perceptual learning in object recognition: object specificity and size invariance. Vision Research, 40(5), 473-484.

Gauthier I (2000) What constrains the organization of the ventral temporal cortex? Trends Cogn Sci 4(1):1-2.

Gilbert CD, Sigman M, Crist RE. 2001. The neural basis of perceptual learning. Neuron 31: 68197.

Goldstone, R. L., Lippa, Y., & Shiffrin, R. M. (2001). Altering object representations through category learning. Cognition, 78, 2743.

Gordan, J. E. (1981). Structures: Or Why Things Don't Fall Down. Cambridge MA: Da Capo Press.

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y. & Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. Neuron, 24 (1), 187-203.

Grill-Spector K, Malach R (2004). The human visual cortex. Annu Rev Neuroscience 27:649-77.

Grill-Spector, K. & Kanwisher, N. (2006). Visual Recognition: As Soon as you known it is there, you know what it is. Psychological Science, 16(2), 152-159.

Gronau, N., Neta, M., & Bar, M. (2008). Integrated Contextual Representation for Objects Identities and Their Locations. Journal of Cognitive Neuroscience, 20(3), 371-388.

Haldane JBS (1928) On being the right size. The Oxford Book of Modern Science

Writing: Oxford.

Hart, J. H. Jr., Lesser, R. P., & Gordon, B. (1992). Selective Interference with the Representation of Size in the Human by Direct Cortical Electrical Stimulation. Journal of Cognitive Neuroscience, 4(4), 337-344.

Hasson U, Harel M, Levy I, Malach R (2003) Large-scale mirror-symmetry organization of human occipito-temporal object areas. Neuron 37:1027-1041.

Hasson U, Levy I, Behrmann M, Hendler T, Malach R (2002) Eccentricity bias as an organizing principle for human high-order object areas. Neuron 34(3):479-490.

Haushofer J, Livingstone MS, Kanwisher N (2008) Multivariate patterns in object-selective cortex dissociate perceptual and physical shape similarity. PLoS Biol 6(7):e187.

Haxby JV, et al (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293:2425-2430.

Hemmer, P. & Steyvers, M. (2009). A Bayesian Account of Reconstructive Memory. Topics in Cognitive Science, 1, 189-202.

Hubbard, T. L., & Baird, J. C. (1988). Overflow, first-sight, and vanishing point distances in visual imagery. Journal of Experimental Psychology: Learning, Memory, and Cognition, 14, 641-649.

Hubbard, T. L., Kall, D., & Baird, J. C. (1989). Imagery, memory, and size-distance invariance. Memory & Cognition, 17, 87-94.

Huttenlocher, J. , Hedges, L. V., & Vevea, J. (2000). What do categories affect stimulus judgments? Journal of Experimental Psychology: General, 129, 220-241.

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in establishing spatial location. Psychological Review, 98, 352-376.

Intraub, H. & Richardson, M. (1989). Wide-angle memories of close-up scenes. Journal of Experimental Psychology: Learning, Memory, & Cognition, 15(2), 179-187.

Ishai A, Ungerleider LG, Martin A, Schouten JL, Haxby JV (1999) Distributed representation of objects in the human ventral visual pathway. Proc Natl Acad Sci USA 96(16):9379-84.

Ittleson, W. H. (1951). Size as a cue to distance: Static localization. The American Journal of Pscyhology, 64(1), 54-67.

Jolicoeur, P. (1987). A size-congruency effect in memory for visual shape. Memory & Cognition, 15(6), 531-543.

Kamitani Y, Tong F. (2005) Decoding the visual and subjective contents of the human brain. Nat Neurosci 8(5):679-685.

Kanwisher N (2010) Functional specificity in the human brain: A window into the functional architecture of the mind. Proc Natl Acad Sci USA 107(25):11163-70.

Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. J Neurosci 17(11):4302-11.

Kass, J. (1997). Topographic Maps are Fundamental to Sensory Processing. Brain Research Bulletin, 44(2), 107-112/

Konkle, T. & Oliva, A. (2011). Canonical visual size for real-world objects. Journal of Experimental Psychology: Human Perception and Performance. 37(1):23-37.

Konkle, T., & Oliva, A. (2007). Normative representation of objects: Evidence for an ecological bias in perception and memory. In D. S. McNamara & J. G. Trafton (Eds.), Proceedings of the 29th Annual Cognitive Science Society, (pp. 407-413), Austin, TX: Cognitive Science Society.

Kosslyn, S. (1978). Measuring the Visual Angle of the Minds Eye. Cognitive Psychology, 10, 356-389.

Kosslyn, S. (1978). Measuring the visual angle of the minds eye. Cognitive Psychology, 10, 356-389.

Kosslyn, S. (1999). Image and Brain: The Resolution of the Imagery Debate. Cambridge: The MIT Press.

Kourtzi, Z., & Kanwisher, N. (2001). Representation of Perceived Object Shape by the Human Lateral Occipital Complex. Science, 293(5543), 1506-1509.

Kourtzi, Z., & Conner, C. E. (2011). Neural Representations for Object Perception: Structure, Category, and Adaptive Coding. Annual Review of Neuroscience, 34, 45-67.

Kriegeskorte et al (2008) Matching Categorical Object representations in Inferior Temporal Cortex of Man and Monkey. Neuron, 60, 1126-1141.

Kunnapas, T. M. (1955). Influence of Frame Size on Apparent Length of a Line. Journal of Experimental Psychology, 50(3), 168-170.

Levy I, Hasson U, Avidan G, Hendler T, Malach R (2001) Center-peripheryorganization of human object areas. Nat Neurosci 4:533-539.

Levy I, Hasson U, Harel M, Malach R (2004) Functional analysis of the periphery effect in human building related areas. Hum Brain Mapp 22:15-26.

Linsen, S., Leyssen, M., Gardner, J. S., & Palmer, S. (in press). Aesthetic preferences in the Size of Images of Real-world Objects. Perception.

Litman L, Awipi T, Davachi L (2009) Category-specificity in the human medial temporal lobe cortex. Hippocampus 19(3):308-19.

Logothetis, N. K. & Sheinburg, D. L. (1996). Visual Object Recognition. Annual Review of Neuroscience, 19, 577-621.

Mack, M. L., Gauthier, I., Sadr, J., Palmeri, T. J. (2008). Object detection and basic-level categorization: Sometimes you know it is there before you know what it is.

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

Mahon BZ et al (2007) Action-related properties shape object representations in the ventral stream. Neuron 55(3):507-520.

Mahon BZ, Schwarzbach J, Caramazza A (2010). The representation of tools in left parietal cortex is independent of visual experience. Psychological Science, 21(6), 764-71.

Mahon BZ, Anzellotti S, Schwarzbach J, Zampini M, Caramazza A. (2009). Category-specific organization in the human brain does not require visual experience. Neuron, 63(3), 397-405.

Mahon BZ, Caramazza A (2011) What drives the organization of object knowledge in the brain? Trends Cogn Sci 15(3):97-103.

Mahon BZ, Caramazza A (2009) Concepts and Categories: A Cognitive Neurosphychological Perspective. Annual Review of Psychology, 60, 27-51.

Malach R, Levy I, Hasson U (2002) The topography of high-order human object areas. Trends Cogn Sci 6(4):176-184.

McCarthy G, Puce A, Gore JC, Allison T (1997) Face-specific processing in the human fusiform gyrus. J Cog Neurosci 9(5):605-610.

McKyton A, Zohary E. Beyond retinotopic mapping: the spatial representation of objects in the human lateral occipital complex. Cereb Cortex 17: 11641172, 2007.

Mechelli A, Price CJ, Friston KJ, Ishai A (2004) Where bottom-up meets top-down: neuronal interactions during perception and imagery. Cereb Cortex 14(11):1256-65.

Merleau-Ponty, M. (1962). Phenomenology of perception. London: Routledge & Kegan Paul.

McCandliss, B. D., Cohen, L., Dehaene, S. (2003). The visual word form area: expertise for reading in the fusiform gyrus. Trends in Cognitive Sciences, 7(7), 293-299.

Milliken, B., Jolicoeur, P. (1992). Size effects in visual recognition memory are determined by perceived size. Memory & Cognition, 20(1), 83-95.

Milner AD, Goodale MA (1995) The visual brain in action. Oxford University Press, Oxford

Moyer, R. S. (1973). Comparing objects in memory; Evidence suggesting an internal psychophysics. Perception & Psychophysics, 13(2), 180-184.

Mullally SL, Maguire EA (2011) A New Role for the Parahippocampal Cortex in Representing Space. J Neurosci 31(20):7441-9.

Niemeier M, Goltz HC, Kuchinad A, Tweed DB, and Vilis T. A contralateral preference in the lateral occipital area: Sensory and attentional mechanisms. Cerebral Cortex, 15: 325331, 2005.

Norman KA, Polyn SM, Detre GJ, Haxby JV (2006) Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. Trends Cogn Sci 10:424-430.

Nosofksy, R. M. (1986). Attention, similarity, and the identification-categorization relationship. Journal of Experimental Psycholoyg: General, 115, 39-57.

O'Craven KM, Kanwisher N (2000) Mental imagery of faces and places activates corresponding stimulus-specific brain regions J Cog Neurosci12(6):1013-1023.

O'Toole AJ, Jiang F, Abdi H, Haxby JV (2005) Partially distributed representations of objects and faces in ventral temporal cortex. J Cogn Neurosci 17(4):580-90.

Olshausen, B. A., & Field, D. J. (1996) Emergence of simple-cell receptive field properties by learning a sparce code for natural images. Nature, 381, 607-609.

Oliva, A. & Torralba, A. (2007). The Role of Context in Object Recognition. Trends in Cognitive Sciences, 11(12), 520-527.

Oliva, A., & Torralba, A. (2001). Modeling the Shape of the Scene: a Holistic Representation of the Spatial Envelope. International Journal of Computer Vision, 42, 145-175.

Ommer, B., & Buhmann, J. M. (2010). Learning the compositional nature of visual object categories for recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32, 501516.

Ono, H. (1969). Apparent Distance as a Function of Familiar Size. Journal of Experimental Psychology, 79(1), 109-115.

Op de Beeck HP, Haushofer J, Kanwisher NG. (2008a) Interpreting fMRI data: maps, modules and dimensions. Nat Rev Neurosci 9(2):123-35.

Op de Beeck HP, Torfs K, Wagemans J (2008b) Perceived shape similarity among unfamiliar objects and the organization of the human object vision pathway. J Neurosci 28(40):10111-23.

Op de Beeck HP. (2010) Against hyperacuity in brain reading: Spatial smoothing does not hurt multivariate fMRI analyses? Neuroimage, 49(3):1943-1948.

Osaka, N. (1976). Reaction time as a function of peripheral retinal locus around fovea: effect of stimulus size. Perceptual and Motor Skills, 42, 603-606

Paivio, A. (1975). Perceptual comparisons through the minds eye. Memory & Cognition, 3(6), 635-647.

Palmer, S. E. Gardner, J. S. & Wickens, T. D. (2008). Aesthetic issues in spatial composition: effects of position and direction on framing single objects. Spatial Vision, 21(3-5), 421-449.

Palmer, S. E., (1975). The effects of contextual scenes on the identification of objects. Memory and Cognition, 3, 519-526.

Palmer, S. E., Rosch, E., & Chase, P. (1981). Canonical perspective and the perception of objects. In J. Long & A. Baddeley (Eds.) Attention and performance IX (pp. 135-151). Hillsdale, NJ: Erlbaum.

Palmer, S. E., Schloss, K. B., & Gardner (In press). Hidden knowledge in aesthetic preferences: Color and spatial composition". In Shimamura, A. P., & Palmer, S. E., Aesthetic science: Connecting minds, brains, and experience. Oxford University Press.

Palatucci, M., Pomerleau, D., Hinton, G., & Mitchell, T. (2009) Zero-shot learning with semantic output codes. NIPS.

Payne, W. H. (1967). Visual reaction time on a circle about the fovea. Science, 155, 481-482

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. Spatial Vision, 10, 437-442.

Pylyshyn, Z. (1999) Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. Behavioral and Brain Sciences, 22, 341-365.

Rajimehr R, Devaney K, Young J, Postelnicu G, Tootell R (2011) The parahippocampal place area responds selectively to high spatial frequencies in humans and monkeys. PloS Biol 9(4):e1000608.

Reddy L, Tsuchiya N, Serre T (2010) Reading the mind's eye: Decoding category information during mental imagery. NeuroImage 50(2):818-825.

Reich, L., Szwed, M., Cohen, L., Amedi, A. (2011) A ventral visual stream reading center indpendent of visual experience. Current Biology, 21(8), 383-8.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. Nature Neuro- science, 2, 10191025.

Rock, I., & Ebenholtz, S. (1959). The relational determination of perceived size. Psychological Review, 66, 387-401.

Rosenthal, R., & Rosnow, R. L. (1991). Essentials of Behavioral Research: Methods and Data Analysis. Boston, MA: McGraw Hill.

Rovamo, F., Virsu, V., & Nasanen, R. (1978). Cortical magnification factor predicts the photopic contrast sensitivity of peripheral vision. Nature, 217, 54-56.

Rubinsten, O., & Henik, A. (2002). Is an ant larger than a lion? Acta Psychologica, 111, 141-154.

Sayres R and Grill-Spector K. Relating retinotopic and object- selective responses in human lateral occipital cortex. Journal of Neurophysiology, 100: 249267, 2008.

Sawamura H, Georgieva S, Vogels R, Vanduffel W, Orban G. (2005) Using functional magnetic resonance imaging to assess adaptation and size invariance of shape processing by humans and monkeys. J Neurosci 25(17):4294-306.

Scholz, J., Triantafyllou, C., Whitfield-Gabrieli, S., Brown, E.N. & Saxe, R. (2009). Distinct regions of right temporo-parietal junction are selective for theory of mind and exogenous attention. PLoS One, 4(3), e4869.

Schwarzlose RF, Swisher JD, Dang S, Kanwisher N (2008) The distribution of category and location information across object-selective regions in human visual cortex. Proc Natl Acad Sci USA 105(11):4447-52.

Schyns, P. G., Goldstone, R. L., & Thibaut, J. P. (1998). The development of features in object concepts. Behavioral and Brain Sciences, 21, 117.

Schyns, P. G., & Rodet, L. (1997). Categorization creates functional features. Journal of Experimental Psychology: Learning, Memory, and Cognition, 23, 681696.

Senders, J. W. (1966). The coffee cup illusion. American Journal of Psychology, 79, 143-145.

Sereno, S., & ODonnel, P. J. (2009). Size matters: Bigger is faster. The Quarterly Journal of Experimental Psychology, 62(6), 1115-1122.

Setti, A., Caramelli, N., Borghi, A.M. (2009). Conceptual information about size of objects in nouns. European Journal of Cognitive Psychology, 21(7), 1022-1044.

Schacter DL, Tulving E (1994) What are the memory systems of 1994? In: Memory systems of 1994? (Schacter DL, Tulving E, eds), pp 1-38. Cambridge, MA: MIT Press.

Shinkareva SV et al. (2008) Using fMRI brain activation to identify cognitive states associated with perception of tools and dwellings. PLoS One. 3(1):e1394.

Simoncelli EP, Olshausen BA. 2001. Natural image statistics and neural representation. Annu. Rev. Neurosci. 24: 1193216.

Sperandio, I., Savazzi, S., Gregory, R. L, Marzi, C. A. (2009), Visual reaction time and size constancy. Perception, 38, 1601-1609.

Srihasam, K. & Livingstone, M (2011) Anatomical correlates of early vs late symbol learning. Poser presented at the Vision Sciences Society, Naples Florida.

Srinivas, K. (1996). Size and reflection effects in priming: A test of transfer-appropriate processing. Memory & Cognition, 24(4), 441-452.

Stevens, S. S. (1957). On the pscyhophysical law. Psychological Review, 64(3), 153-181.

Stokes M, Thompson R, Cusack R, Duncan J (2009) Top-down activation of shape-specific population codes in visual cortex during mental imagery. J Neurosci 29(5):1565-72.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. Journal of Experimental Psychology, 18, 643-662.

Tanaka, K. (1996). Inferotemporal Cortex and Object Vision. Annual Review of Neuroscience, 19, 109-139.

Tarr, M. J. (1995). Rotating objects to recognize them: A case study of the role of viewpoint dependency in the recognition of three-dimensional objects. Psychonomic Bulletin and Review, 2(1), 55-82.

Tarr, M. J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. Cognitive Psychology, 21(28), 233-282.

Tarr, M. J., Williams, P., Hayward, W. G., & Gauthier, I. (1998). Three-dimensional object recognition is viewpoint dependent. Nature Neuroscience, 1(4), 275-277.

Torralba, A., Murphy, K. P., & Freeman, W. T. (2004). Sharing features: Efficient boosting procedures for multiclass object detection. Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2, 762769.

Tyler CW et al (2005) Extended concepts of occipital retinotopy. Curr Med Imaging Rev 1(3):319-329.

Ullman, S. (1989). Aligning pictoral description: An approach to object recognition. Cognition, 32, 193-254.

Ullman, S. (2007). Object recognition and segmentation by a fragment-based hierarchy. Trends in Cognitive Sciences, 11(2), 58-64.

Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. Nature Neuroscience, 5, 16.

Underleider LG, Mishkin M (1982) Two cortical visual systems. Analysis of Visual Behavior, eds Ingle MA, Goodale MI, Masfield RJW (MIT Press, Cambridge, MA), pp 549 -586.

Valyear KF, Culham JC (2010) Observing learned object-specific functional grasps preferentially activates the ventral stream. J Cog Neurosci 22(5):970-984.

Vuilleumier P, Henson R, Driver J, Dolan R. (2002) Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. Nat Neurosci 5(5):491-499.

Walker, J. T., Rupick, R. C. & Powell, J. L. (1989). The vista paradox: A natural visual illusion. Perception & Psychophysics, 45(1), 43-48.

Wandell, B. A., Dumoulin, S., & Brewer, A. A. (2007). Visual Field Maps in Human Cortex. Neuron, 56, 366-383.

Weber M et al. (2009) Predicting judged similarity of natural categories from their neural representations. Neuropsychologia 47(3):859-868.

Weber, E. U. & Johnson, E. J. (2006). Constructing preferences from memory. In: Lichtenstein, S. & Slovic, P., (Eds.), The Construction of Preference (pp. 397-410). New York NY: Cambridge University Press.

Yonas, A., Pettersen, L., & Granrud, E. (1982). Infants Sensitivity to Familiar Size as Information for Distance. Child Development, 53(5), 1285-1290.