Title
Uncovering Human Visual Priors

Author
Langlois, Thomas Alexandre

Publication Date
2018

Peer reviewed|Thesis/dissertation

Uncovering Human Visual Priors

By

Thomas A Langlois

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Psychology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Thomas Griffiths, Chair
Professor David Whitney
Professor Alyosha Efros

Summer 2018

Abstract

Uncovering Human Visual Priors

by

Thomas A Langlois

Doctor of Philosophy in Psychology

University of California, Berkeley

Professor Thomas Griffiths, Chair

Visual perception can be understood as an inferential process that combines noisy sensory information with internalized knowledge drawn from previous experience. In statistical Bayesian terms, internal representations of the visual environment can be understood as posterior estimates obtained by weighting imperfect sensory information (a likelihood) by internalized biases (a prior). Given limited perceptual resources, it is advantageous for the visual system to capitalize on predictable regularities of the visual world, and internalize them in the form of priors. This dissertation presents novel findings in the domain of spatial vision and visual memory, as well as some new work on memory for the 3D orientation of objects. In all cases, an unprecedented signal-to-noise ratio, achieved by employing serial reproduction chains (a "telephone game" procedure) combined with non-parametric kernel density estimation techniques, reveals a number of stunning intricacies in the prior for the first time. Methodological implications, as well as implications for amending prior empirical findings and revisiting past theoretical explanations are discussed.

To Steven Sabat, Gerald Wartofsky, and Cecilia Cho for their rare gifts as teachers.

Contents

Listing of figures

Acknowledgments

I have had the immense good fortune of becoming a part of a community of extraordinary people and scientists at UC Berkeley. Some of these people have since become friends, others have become mentors, some have become both. All have given me a sense of belonging, and common purpose. I would especially like to thank Tom Griffiths for his incredible mentorship and generosity. I cannot thank him enough for being part of such an exciting scientific journey. I will always be indebted to him for the incredible work we are doing, and the opportunity to become a member of his exceptional lab. I cannot think of a better place to grow as a thinker and as a scientist.

I would also like to thank Nori Jacoby, who has become a dear friend and mentor. His enthusiasm, talent, and energy push me to greater heights, and being his collaborator on the work presented in chapter 2 and chapter 3 has been, and continues to be, an extraordinary blessing. I look forward to a long and fruitful collaboration with him.

I have also been extremely lucky to have developed what I hope will be lifelong friendships with fellow students. Joshua Peterson, with whom I share a parallel path from the beginning. I will never forget the experiences we shared as we made our way through Berkeley, and look forward to having more stimulating conversations with him in the next chapter of our lives. I am also grateful for my burgeoning friendship with Ruairidh Battleday and Paul Krueger. They have enriched my life with their intelligence, insight, enthusiasm and dynamism. I would also like to thank many other members and collaborators of the lab, including Jordan Suchow, David Bourgin, Joshua Abbott, Stephan Meylan, Jess Hamrick, Aida Nematzadeh, Daniel Reichman, Sophia Sanborn, Rachel Jansen, and many others. They dazzle and inspire me, and make me proud to be a fellow colleague.

I would also like to thank my other Committee members, Alyosha Efros, and David Whitney. They exemplify scientific excellence, innovation, and creativity. Seeing their talent up close has been inspiring. Finally, becoming a part of the academic community at Berkeley was made possible when Steve Palmer took a chance with me and gave me the opportunity to get involved in visual perception research.

I would like to thank my parents Jean-Pierre and Catherine, and my brother Julien for their unconditional love, wisdom and unrelenting support. To Tiffany Ku, for her joyful presence at the time I started this journey. Finally, to Sylvia Blackmore, whose grace elevates even the most mundane moments.

1
Introduction

Forming accurate internal memory representations of the visual environment poses a difficult problem for the human visual system. It must create useful representations of sensory information that is often noisy and extremely complex, while keeping pace with a relentless stream of visual input. As if understanding the sheer mechanics of such a feat were not enough, computational-level questions (Marr, 1982) arise, such as: what is the purpose of an information processing system that retains visual information from past experiences, and how does it accomplish that? Studying visual memory representations places scientists at the junction where two enduring areas of psychological research meet: the study of memory on the one hand, and the study of visual perception on the other. As early as the 1890s, William James theorized that there is a distinction between memory for information that is held in the "conscious present", and memory for information that is stored and only later remembered (not held in conscious awareness) (James, 2013), a distinction that is closely related to the more modern notions of short-term "working memory" and long-term "episodic memory" (Atkinson & Shiffrin, 1968; Brady, Konkle, & Alvarez, 2011; James, 2013). An important difference between these two memory processes has to do with storage capacity—while short-term memory has a very limited capacity (Miller, 1956), long-term memory has a far greater capacity, and can store many more items in different modalities, such as visual images, words, or associations (Atkinson & Shiffrin, 1968; Brady, Konkle, Alvarez, & Oliva, 2008). However, constraints remain on the nature of memory representations—far from possessing eidetic memories of

our experiences, our long-term memory systems do not form "veridic" copies of our experiences, in spite of its enormous storage capacities. In fact, it has been shown to be sensitive to significant distortions over time (Loftus, 2005), and generally privileges meaningful items that are linked to stored knowledge (Wiseman & Neisser, 1974). Biases like these remind us of the intimate functional relationship between our visual and memory systems, and the judicious retrospective record-keeping that it performs in the service of establishing a sound basis for future action and perception, a sentiment summarized by J.J Gibson in the statement "We must perceive in order to move, but we must also move in order to perceive" (Gibson, 2014).

Visual long-term memory refers to a person's ability to remember an image that was seen in the past, but not held in working memory (Brady et al., 2011). In other words, it refers to our ability to store visual episodic information for future retrieval, below conscious awareness. It may be distinguished from "stored knowledge" which Alvarez describes as "the preexisting visual representations that underlie our ability to perceive and recognize visual input." (Brady et al., 2011) These distinctions have a clear statistical interpretation: from a Bayesian perspective, the visual system integrates sensory information (the "likelihood") with information acquired from previous experience (the "prior"). The visual information that we experience and then store into long-term memory corresponds to the "likelihood," while the "stored knowledge" is closely related to the "prior." As a result, information that we retrieve from memory can be thought of as the "posterior" which is a function of both the "likelihood" and the "prior." Thus, when information is retrieved, it is colored by our preexisting visual representations, and usually does not correspond to the original sensory percept. While it may seem at first glance that information that is stored should not undergo transformations when it is retrieved from memory, forming priors is actually advantageous, as it allows perceptual systems to capitalize on the statistical regularities of the environment that are important to us.

Visual memory research encompasses a very wide range of phenomena (Brady et al., 2011), but this dissertation explores biases in spatial memory, which refers to the remembered physical position of objects, or their 3D orientation. In what follows, we describe new findings from a series of experiments investigating both the remembered 2D position of objects inside background frames (including aesthetic biases in spatial composition), and spatial memory biases in the remembered 2D position of a simple dot inside an image are discussed. The final chapter explores systematic biases in the remembered 3D orientation of objects. In all cases, a novel experimental design is adopted (using a "telephone game" experimental procedure), followed by non-parametric kernel density estimation techniques, which capture these biases in unprecedented detail—often revealing hitherto unknown patterns.

In a typical spatial memory experiment, a participant is presented with a dot positioned somewhere inside a simple geometric shape (such as a circle). Following a delay, the participant is shown the shape, and must indicate the exact position in which the dot was originally presented. Previous work discovered that memory for the exact position of the dot inside a circle is biased towards "categories" or "prototypes" centered in each of the four quadrants of the circle (Huttenlocher, Hedges, & Duncan, 1991) . Later work examined whether similar "categorical" effects in spatial memory are present when the dot is presented inside different geometric shapes (curved shapes, such as vertical and horizontal ellipses, as well as angular shapes—regular polygons: an equilateral triangle, square, and pentagon). Using a parametric model to estimate the locations of the prototypes for all the shapes, their results confirmed that memory for the position of dots presented inside these new shapes shows substantial "categorical" effects (Wedell, Fitting, & Allen, 2007). Later work reported that memory for the position of dots placed randomly over natural images also yielded significant "categorical" effects—suggesting that spatial memory is biased towards the geometric centers of mass inside regions that correspond to segmentations obtained from human observers (Holden, Newcombe, & Shipley, 2013). These memory biases have traditionally been presented in terms of "categorical" effects in spatial memory. "Categorical" effects have also been documented in other areas of visual perception, such as compositional biases in people's aesthetic preferences (the "inward bias" for instance, which describes the tendency for people to prefer objects with a clear facing direction to be placed off-center inside a bounding frame, facing the center), as well systematic biases favoring some 3D views of objects over others (often described as "canonical perspectives").

1.1 Uncovering visual memory priors

While previous work confirmed the presence of systematic biases in visual memory (Holden et al., 2013; Huttenlocher et al., 1991; Wedell et al., 2007), it typically employed experimental and modeling approaches with significant drawbacks. In all cases, deviations in spatial memory were measured for a small number of trials, and following only a single iteration of the memory task (a presentation, delay, and a response), resulting in a poor signal-to-noise ratio in the response data. In addition, the parametric approaches that were used to infer the location of "prototypes" biasing spatial memory usually made strong *a priori* commitments regarding their total number—a set number of locations determined a priori, or a single "canonical perspective" in preferences for 3D orientations of objects. However, what if the number of prototypes varies as a function of certain properties of an image, or a 3D object? In the case of natural images, do the number of "prototypes" vary as a function of some visual features? In prior work, only a single location was assumed to bias spatial memory inside the

segmented regions of natural images—the geometric centers of mass of segmented regions. But this work typically lacked the power necessary to reveal whether additional loci for these "prototypes" were present. As a result, while this work provided clear evidence for systematic biases in spatial memory (in 2D), and orientation preferences (in 3D), an exact picture of the priors themselves continues to elude traditional experimental designs, and parametric modeling approaches.

1.2 Serial reproduction

Serial reproduction chains are analogous to the so-called "telephone game." In an experimental setting, the procedure is as follows: a participant generates a response to some stimulus (such as indicating the position of an object in a spatial memory task). This response then becomes the stimulus for a second participant, who performs the same task—indicating the exact position of the object shown during the stimulus phase, which unbeknownst to him is actually the remembered position generated by the previous participant. The second participant's response may then be used as the stimulus for a third participant, who again produces a response, and so on and so forth.

Serial reproduction has a long history in experimental psychology, where it has been used to study how various biases distort information when it is transmitted from person to person (Bartlett, 1932). Figure 3.2 shows a schematic illustration of the experimental paradigm: a participant views a stimulus, such as a dot presented within a bounding shape, and is then asked to reproduce the stimulus as accurately as possible from memory. Critically, the reproduction created by the first participant is used as the stimulus for the second participant, who is then asked to do the same. At each iteration, the reconstruction produced by the previous participant becomes the stimulus for the next participant to reproduce. Famous early results include the transformation of an owl-like Egyptian hieroglyph into a small cat after ten iterations of a serial-reproduction drawing task (Bartlett, 1932). This result was interpreted in terms of inductive biases in memory: as veridical information from the input becomes degraded following successive iterations, the reconstruction of the ambiguous image is pulled towards a prototypical object with similar visual properties inside people's heads.

Serial-reproduction experiments have long been used to simulate phenomena in cultural transmission, evolutionary biology, anthropology, and cognitive science (Claidière, Smith, Kirby, & Fagot, 2014; Kirby, Cornish, & Smith, 2008), but it wasn't until recently that a rational analysis of serial reproduction considered how information should change as it is transmitted along a chain of rational agents (Xu & Griffiths, 2010). Under a rational analysis, reconstruction from memory is defined as the problem of inferring the most accurate state of the world from noisy data, such as an imperfect memory trace and perceptual noise during encoding of the image. This problem is

modeled using the framework of Bayesian statistics. Previous experience is captured by a prior distribution over possible states (a hypothesis space of world states). A posterior is computed, based on the likelihood, which indicates the probability of observing that information, given some hypothesis about the true state of the world. Xu and Griffiths (2010) examined the predictions of this Bayesian account of reconstruction from memory for serial reproduction. They found that serial reproduction by Bayesian agents defines a Markov chain with the following transition probabilities:

$$p(x_{n+1} \mid x_n) = \int p(x_{n+1} \mid \mu) p(\mu \mid x_n) d\mu,$$

where x is a noisy stimulus (such as an imperfect memory trace) and μ is the true state of the world that generated that stimulus (in this case, the veridical image that impinged on the visual system). This Markov chain captures the probability of a new stimulus x_{n+1} being created as a reconstruction of a previously seen stimulus x_n in each iteration in the serial reproduction chain, and has a stationary distribution, called the *prior predictive distribution*, which defines the probability of observing a stimulus x when μ is sampled from the prior:

$$p(x) = \int p(x \mid \mu) p(\mu) d\mu.$$

This process approximates a Gibbs sampler for the joint distribution on x and μ defined by multiplying $p(x \mid \mu)$ and $p(\mu)$. This finding is significant because it provides a mathematical formalism for describing the consequences of serial reproduction: assuming that participants share common inductive biases, the transmission chain will converge to a sample from their shared prior.

1.3 Crowdsourcing curated networks

While serial reproduction chains have been used in the past to study memory biases, to our knowledge, we are the first to use them to study perceptual priors in the context of visual perception, and in order to study visual memory representations in particular. This is due in part to the significant technical challenges posed by managing these chains at the scale required to quantify these priors. In order to quantify them with the precision needed to fully capture their intricacies, we often needed between two-hundred and five-hundred unique chains, each with between ten and over twenty-five nodes (links in the chain). These requirements would have made this work extremely prohibitive to undertake had it been confined to the experimental tools available in a traditional laboratory setting. While crowdsourcing platforms provided a partial remedy—by enabling the simultaneous participation of many workers in simple experimental tasks, our ability to arrange workers on Amazon

Mechanical Turk (AMT) into carefully curated serial reproduction chains was the final technical building block that made investigating these priors not only possible, but remarkably tractable and fruitful.

1.4 Objectives and layout of the present dissertation

The chapters that follow reveal new findings in multiple visual domains—visual aesthetics, including visual memory priors for the position of objects inside a frame, spatial memory priors in 2D images, and 3D orientation priors in memory. While the first chapter is primarily devoted to new work in spatial aesthetics using traditional experimental designs, with only one section revealing spatial memory priors with some similar characteristics to aesthetic preferences using serial reproduction chains (an "off-center" bias in compositional memory, although not the ubiquitous "inward bias"), the second and third chapters illustrate the true promise and sensitivity of curated serial reproduction chains for revealing visual memory priors. The primary objectives of this dissertation are to 1) demonstrate the extraordinary sensitivity of using serial reproduction networks and non-parametric modeling approaches to reveal human visual priors in a variety of visual memory domains, 2) discuss the implications for amending prior work with new (hitherto unknown) empirical findings for future methodological approaches, and 3) discuss the implications for revisiting current theoretical beliefs regarding visual memory biases based on earlier empirical work.

The first chapter presents empirical evidence for strong spatial biases in the preferred placement of objects in a scene that are conditioned on the implied motion speed, facing direction, and implied motion trajectories of objects. These biases reveal a reduction in the "inward bias" for objects that appear to be moving faster—while objects that are facing in one direction are preferred placed off-center, facing the center of its background frame, this bias is reduced if objects appear to be moving in the direction it is facing. In the first chapter, results from two experiments show that 1) the "inward bias" is reduced as a function of the implied motion speed of an object, and 2) the "inward bias" is maximized for objects with consistent facing and implied motion trajectories, but reversed when the implied motion trajectories are inconsistent with the facing directions. The chapter concludes with some results probing the spatial memory prior for facing objects in a natural scene, which reveals a clear off-center bias, although not an "inward bias".

The second chapter presents work on spatial memory priors, revealing hitherto unknown patterns in the prior for the memory of a dot presented over an image. The first section discusses results that were obtained for the remembered position of a dot initialized within the boundaries of a set of curved and angular geometric shapes (circle, horizontal and vertical ellipses, equilateral triangle,

square, regular pentagon). These results reveal for the first time a number of intricate patterns in the spatial memory prior—patterns that had previously eluded traditional experimental and parametric modeling approaches. Quantitative comparisons to estimates using results following only one iteration of the serial reproduction process (which is equivalent to using the response data following exposure to the original random placements of the dots, as in a traditional experimental design), and using the Category Adjustment Model (CAM) are also discussed. They demonstrate the superiority of our non-parametric approach, and the benefits of using the data following several iterations of the serial reproduction chain for obtaining a higher signal-to-noise ratio in estimating the prior. The section using simple geometric shapes ends with a discussion of additional results that were obtained when simple features of the shapes were manipulated—such as changes in the prior for regular polygons as additional vertices are added and the polygons begin to approximate a circle, as well as effects of varying the exposure-time during the retention phase of the task. Finally, we discuss the effects of rotating bounding triangular and square frames on the prior for the memory of a dot presented inside a circular region. This section is followed by results that were obtained for the same task, and using the same shape images, but after loosening the constraint that the dots be initialized randomly within the shape boundaries.

The second section discusses results that were obtained for the remembered position of a dot placed over a set of natural images, which also reveals that visual memory biases deviate from experimental and theoretical results put forward in prior work, including evidence for a bias towards remembering the location of random dots towards the centers of mass of segmented regions in images. Both series of experiments paint a much more nuanced picture of categorical effects in visual memory. The results for the task using natural images reveal visual memory priors that suggest serial reproduction may be a useful paradigm for revealing the structure of visual memory representations of images, and additional experimental evidence from a recognition task indicates that using the prior to reveal certain visual regions over an image boosts recognition accuracy when the correct prior is used. The chapter ends with results from a series of visual discrimination tasks that indicate that while visual discrimination is greater in some regions that have higher probability in the visual memory prior, it is also increased in other visual regions, suggesting that while the priors we measure are diagnostic of object identity in a recognition task, they do not correspond necessarily to the visual areas where visual discrimination is highest, suggesting that serial reproduction doesn't yield spatial memory priors in terms of location memory, but may instead be revealing the latent geometry of visual memory representations, and that distortions in the location memory of the dots are not due to spatial memory biases *per se*, but may be the incidental byproduct of deformation in visual memory representations.

The third chapter reveals the visual memory prior for the three-dimensional orientation of a set of 3-dimensional objects, using curated serial reproduction chains of participants over AMT. These results build on the results from the previous chapter, by showing how this paradigm is useful for deriving intricate (mostly) non-parametric estimates of human visual priors in a different visual task. Results reveal how intricate patterns that have eluded traditional experimental and modeling approaches are again made plain using our approach.

2

Effects of Implied Motion and Facing Direction on Positional Preferences in Single-Object Pictures

Among the many aspects of a picture that contribute to people's aesthetic appreciation, one of the most important is spatial composition (Amheim, 1974; Arnheim, 1954, 1983). Spatial composition refers to the positional arrangement of the depicted objects relative to each other and to the surrounding frame (Alexander, 2002; Arnheim, 1983). Perhaps the single aspect of spatial composition that has received the most scientific attention is horizontal balance: the perception of how equally and evenly the pictured objects are distributed to the left and right of the frame's center. (P. Locher, Overbeeke, & Stappers, 2005; P. J. Locher, 2003; P. J. Locher, Stappers, & Overbeeke, 1998; McManus, Edmondson, & Rodger, 1985) The usual supposition is that horizontally well-balanced pictures are aesthetically preferable to horizontally poorly-balanced ones. Palmer, Gardner, and Wickens (2008) studied one of the simplest of compositional choices: people's preferences for the position of a single object within a rectangular frame (Palmer, Gardner, & Wickens,

Much of the content of this chapter was published in Palmer and Langlois (2017).

2008). If balance were the primary factor governing people's aesthetic preferences for such single-object pictures, the object would always be preferred at or near the frame's center, where its spatial extent is about equally distributed left and right of center. Palmer and colleagues indeed found this to be true, but only for bilaterally symmetric objects that were facing forward, toward the viewer. Rightward- and leftward-facing objects, in contrast, were strongly preferred when they were positioned off-center so that they faced into the frame: pictures of right-facing objects were better liked when they were positioned left of center, and pictures of left-facing objects were better liked when they were positioned right of center. Palmer, Gardner, and Wickens (2008) called these two mirror-image effects the "inward bias" because in both cases viewers preferred the object to face into the frame more than out of the frame (Palmer et al., 2008). Bertamini, Bennett and Bode (2011) replicated this effect by showing an aesthetic bias for pictures of animals to have more space in front of them than behind them, although they called it the "anterior bias" (Bertamini, Bennett, & Bode, 2011).

These inward biases are very robust and have been replicated and extended in further studies (Sammartino & Palmer, 2012), including a recent analysis of the frame-based position of the faces of main characters in action films (Bode, Bertamini, & Helmy, 2016). Such findings provide clear evidence that balance – at least in its usual interpretation of physical balance – is not the dominant factor in spatial composition that it is usually portrayed as being. The present article considers another factor that may be as important, if not more so, in determining people's preference for unbalanced spatial compositions: namely, implied motion. Pictures are by definition static, so actual motion cannot be directly represented. There are nevertheless a number of ways in which motion can be implied in a picture of an object, including the depiction of an object's characteristic direction and speed of motion as well as the existence of motion blur in the object and/or the background. Any or all of these features might influence compositional preferences. For example, most objects that move – including people, cars, and horses – tend to move in a particular object-based direction: namely, with their front surfaces facing in the direction of motion. The facing direction of a movable object therefore strongly implies a likely direction of motion: forward. This correlation suggests that people may prefer moving objects to have more space in front than behind them within a frame to provide the object with more "room to move" in a forward direction. If so, implied motion of forward-moving objects would induce preference biases that coincide with the inward facing bias. Indeed, the inward facing bias might actually be determined, at least in part, by implied motion. Palmer et al. (2008) tested this possibility by measuring the inward bias for objects that characteristically move forward (e.g., a person, dog, and car) versus objects that are characteristically stationary but nevertheless have a well-defined front (e.g., a chair, daisy, and teapot). They found no differ-

ences between the two groups of objects and therefore concluded that the inward bias does not depend on motion (Palmer et al., 2008).

Why might an inward bias be present in static pictures depicting single objects that could move and/or faced rightward or leftward in the picture plane but not for the same objects when they move or faced forward? Sammartino and Palmer (2012b) proposed an affordance space hypothesis to account for this pattern of results (Sammartino & Palmer, 2012). The key idea is that every object is surrounded by a virtual spatial envelope, which they called its affordance space, whose distribution reflects its affordances: i.e., its opportunities for action and interaction with respect to the viewer (Gibson, 2014). The shape of the affordance space is hypothesized to represent the spatial extent and/or importance of functions that take place in corresponding regions around the object, due to the object's own behavior (for moving objects, such as a person, dog, or car), the behavior of the observer with respect to the object (for facing objects that do not move, such as a chair, flower, or teapot), or both. The affordance space of a chair, for example, would be more spatially extended around its front and top, because those are the regions around chairs of greatest importance when one uses them for their most important affordance of being "sittable upon." The crucial additional assumption of the affordance space hypothesis for predicting preferences in spatial composition is that people prefer pictures in which the affordance space, rather than the physical extent of the object, is centered within the frame. This account implies that objects whose affordance spaces are symmetrical in the picture plane will be preferred when positioned in the center of the frame (e.g., a front-facing person or chair), whereas the same objects will be preferred facing inward when their affordance spaces are asymmetrical in the picture plane, due to having a larger region in their affordance spaces around the surfaces most associated with relevant affordances, which usually means their front surface. This account squares well with Palmer et al.'s (2008) finding that objects that move forward produce an inward bias comparable to that of objects that merely face forward when shown in profile. However, the movable objects in that study were portrayed in stationary poses: e.g., a standing person rather than a walking or running one. Might pictures that actually depict objects in motion produce an inward-moving bias that is conceptually distinct and experimentally dissociable from an inward-facing bias? In this article we study the effects of implied motion of a single target object on people's aesthetic preferences for pictures of that object when depicted in motion. We consider the two fundamental aspects of such motion: its direction in Experiment 1 and its speed in Experiment 2. The primary question of Experiment 1 is whether motion direction and facing direction have dissociable effects on compositional preferences. Secondarily, we investigate whether motion or facing effects dominate when they conflict. The primary question of Experiment 2 is how speed might modulate the implied motion effects we demonstrate in Experiment 1. Would pictures

of faster moving objects produce larger inward biases, as would be expected from a straightforward "prospective" interpretation of the affordance space hypothesis (i.e., how the object is expected to move in the near future) or would they produce smaller inward biases, as would be expected from a "retrospective" interpretation of the affordance space hypothesis (i.e., how far the object has moved in the recent past)?

2.1 Directional Effects of Implied Motion versus Facing

Most objects characteristically move in a forward (i.e., front-facing) direction, meaning that their motion vector points in the same direction as their front surface faces. There are notable exceptions, of course, such as the species of crab that locomote sideways, but forward motion is by far the norm. Even so, there are circumstances in which objects that characteristically move forward actually move in other directions – whether intentionally or otherwise – and sometimes even in the opposite direction. In the present experiment, we dissociated facing direction from moving direction by studying forward versus backward motion events that depict people in the process of diving or falling. Because the images we used of the divers and fallers showed them far enough along their motion trajectories for viewers to discriminate easily between forward and backward implied motion, their faces were actually facing somewhat downward for the forward divers and fallers and upward for the backward divers and fallers. Nevertheless, their bodies were clearly facing right or left, coincident with the direction of motion for the forward divers and fallers but opposite the direction of motion for the backward divers and fallers. Based on previous results with pictures of stationary objects, we expected an inward facing bias in people's aesthetic preferences, but we also expected an inward motion bias, such that the combined inward bias would be greater for the forward-moving figures (for whom the inward-facing and inward-moving biases reinforce each other) than backward-moving figures (for whom the inward-facing and inward-moving biases conflict with each other). It is unclear whether the inward-facing or inward-moving bias will dominate in the backward motion conditions, however, where any net result is possible, depending on the relative strength of the two competing biases. We measured aesthetic biases using a "drag-and-drop" task, in which participants were asked to "drag" the target object into and around the background image using a computer mouse until they found the most aesthetically pleasing position for it, and to "drop" it by pressing a mouse button. (See Leyssen et al., 2012, for another study of spatial composition using this method (Leyssen, Linsen, Sammartino, & Palmer, 2012).) The position at which the target object's center was located when dropped was taken to be the most aesthetically preferred spatial composition. It differs from the two-alternative forced-choice paradigm used by Palmer et al. (2008) in that the

drag-and-drop task involves comparisons between different possible placements that are implicit and memory-based, but the drag-and-drop task is far more efficient in terms of the number of trials required (see Palmer, Schloss and Sammartino, 2013, for a methodological review of measuring aesthetic preferences (Palmer, Schloss, & Sammartino, 2013)). Dragging and dropping an object into a desired location also seems to be a more natural production task, mimicking the kind of decision process ones performs in editing images for composition in computer programs, such as Adobe Photoshop.

2.1.1 Methods

Participants

Twenty undergraduate students (6 men, 14 women) participated at the University of California, Berkeley, through the Psychology Department's Research Participation Pool (RPP) for course credit. All participants reported having normal or corrected-to-normal spatial vision. All gave informed consent, and the Committee for the Protection of Human Subjects (CPHS) at the University of California, Berkeley, approved the experimental protocol.

Design

The experimental design consisted of the orthogonal combination of two motion directions (left-moving and right-moving) and two facing directions (left-facing and right-facing), with 8 image exemplars (six divers and two fallers) in each of the 2 x 2 facing/moving direction conditions. Participants performed a drag-and-drop task, in which they were instructed to "drag" each object into a background image and to "drop" it in the most aesthetically pleasing location (see below for details).

Stimuli

The stimuli consisted of sixteen pictures that showed people either diving or falling in a direction that was either forward (toward their front side) or backward (toward their back side). Representative examples are shown in Figure 2.1. Each image was presented in both its original and its left/right reversed versions to manipulate facing direction. This stimulus design – 2 facing directions (left/right) x 2 motion directions (left/right) – allowed us to investigate inward versus outward biases in compositional preferences due to facing direction, motion direction, and their integration in both consistent and conflicting combinations. For example, forward divers/fallers have consistent facing and motion directions, whereas backward divers/fallers have conflicting facing and motion

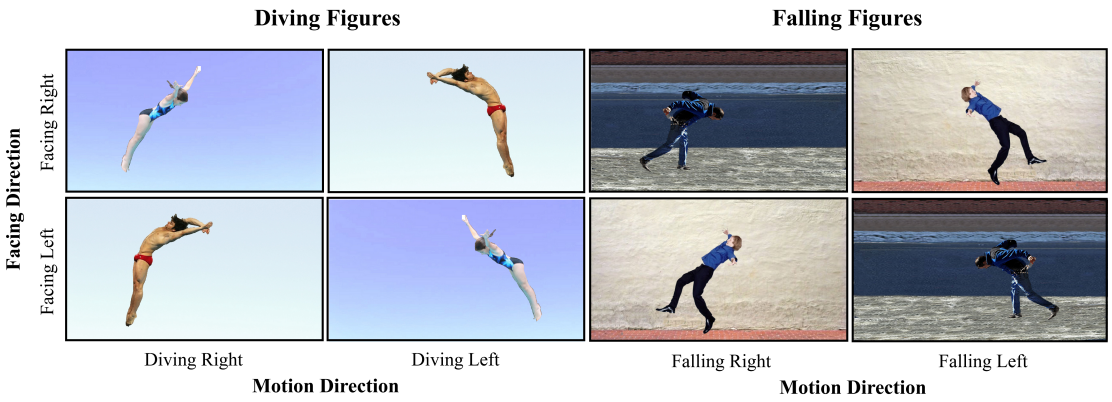

Figure 2.1: Representative examples of the stimuli used in Experiment 1. Divers and fallers had either the same facing and motion directions (forward divers and fallers), or opposite facing and motion directions (backward divers and fallers). These figures are shown in the average positions chosen by participants.

directions. All of the target figures (twelve divers and four fallers) were taken from naturalistic non-proprietary photographs found on the internet, and they were shown in their original background contexts whenever possible to maintain realism. The target objects were isolated and removed from the picture using Adobe Photoshop software, and the background was processed to fill in the regions in which the target objects had originally appeared with pixels from neighboring regions that were consistent with corresponding parts of the depicted background such that the original position of the object was undetectable. Each picture was presented in color. The background was positioned at the center of the screen, and the target object was initially placed at random on either the left, right, top, or bottom border of the background image such that half of it extended out of the boundary of the background image.

All backgrounds were 786 pixels wide x 436 pixels high. The diving and falling figures ranged from 195 to 329 pixels in width and 248 to 367 pixels in height. There was some unintended systematic variation in the widths of the forward-moving versus backward-moving figures, in that the average width of the forward-moving figures was 268.5 pixels and that of the backward-moving figures was 246.4 pixels. Nevertheless, there was a great deal of overlap in the image widths of the forward moving and backward moving figures.

Procedure

Participants viewed a computer screen from about 70 cm inside a darkened booth. They were shown images of each of the objects and were instructed to place it in the background image at the

location in which they found it to be the "most aesthetically pleasing." The target object was initially positioned with its center at the midpoint of one of the four sides. Participants could then move the object freely with the mouse to any location within the background image, but not beyond its borders. The x-axis of the mouse was therefore restricted to the range of -250 to +250 pixels horizontally in all cases. Once they placed the object in what they considered to be the most aesthetically pleasing location, they were instructed to left-click the mouse to "drop" the object at that position. If participants were not satisfied with their choice, they could repeat the procedure as many times as they wanted by clicking on a "back" button on the screen. Participants were shown all thirty-two images of the different divers and fallers with their corresponding backgrounds in random order.1 The horizontal position of the center of each target object was recorded.

2.1.2 Results and Discussion

The primary analyses were conducted on the average x-axis coordinates corresponding to the horizontal positions of the object centers computed across all eight exemplars within the 2 x 2 design: facing direction (left/right) and motion direction (left/right). The results (see Figure 2.2) show a very large effect of motion direction [$F(1, 19) = 76.22$, $p < .0001$, $\eta = .800$] due to a strong bias to prefer the target objects to be positioned moving inward, toward the center of the frame. There was also a reliable, though weaker, effect of facing direction [$F(1, 19) = 21.97$, $p < .0001$, $\eta = .536$], reflecting a bias to prefer objects facing inward, toward the center (i.e., with right-facing objects positioned farther to the left and left-facing objects facing farther to the right). The inward bias for motion direction is so much stronger than the inward bias for facing direction, however, that the net bias for backward divers/fallers was for them to face outward from the center, with less space in front than behind them! There was a small, but significant, interaction between motion direction and facing direction [$F(1, 19) = 5.76$, $p < .027$, $\eta = .233$] due to the fact that the difference between leftward and rightward facing directions was slightly larger when the divers and fallers were moving leftward than when they were moving rightward. The results show that whereas forward divers/fallers tend to be placed so that they both move and face inward (toward the center of the image with more space in front than behind them), backward divers/fallers tend to be placed so that they move inward, but face outward. This result indicates that, although both motion direction and facing direction have reliable effects, motion direction has the much larger influence. One reason why the effects of motion direction might be so much stronger than the effects of facing direction is that the present facing directions are not as directly lateral as the motion directions: as mentioned previously, the forward divers and fallers tend to be facing somewhat downward and the backward divers and fallers

tend to be facing somewhat upward rather than directly leftward or rightward. We doubt that this difference is sufficient to account for the much larger motion-facing bias, but there is no evidence in the present data to rule out the possibility. The images of the forward-moving divers and fallers were slightly wider horizontally (268.5 pixels) than the backward-moving divers and fallers (246.4 pixels). Although this stimulus difference of 22.1 pixels is less than 3 percent of the 786-pixel lateral extent of the background, we were nevertheless concerned that it might have influenced the magnitude of the motion and/or facing effects. If, for example, an influential factor were the distance between the back edge of the figural image and the closest edge of the frame, one would expect a negative correlation between image width and distance from the center of the frame to the center of the figure (i.e., the center of wider images would be closer to the center of the frame). Instead, we found a significant positive correlation between image width and absolute distance from the center ($r = +.52$, $p < .005$). This might be interpreted as indicating that participants tended to place wider figures closer to the closest frame edge than the narrower images, as would be expected if the more influential factor were the distance of the front of the figural image from the center. This is not the only explanation or even the most plausible one. A more consistent alternative is that the positive correlation between image width and strength of inward bias is actually due to the fact that there is a greater net inward bias for the slightly wider forward-moving figures, because their motion and facing biases are consistent, than there is for the slightly narrower backward-moving figures, because their motion and facing biases conflict. To control for this image-width confound statistically, we analyzed the effects of facing direction, motion direction, and image width on the preferred horizontal positions with linear mixed-effects models using the lme4 package (Bates and Sarkar, 2016) in the R environment (R Development Core Team, 2016) (Pinheiro, Bates, DebRoy, Sarkar, & Team, 2010). In the model, we specified subjects as the random factor in order to control for their intraclass correlation using random intercepts for each subject (Pinheiro and Bates, 2000). Facing direction, motion direction, and image width were specified as the fixed factors. We performed statistical significance tests using type II Wald chi-squared tests of the fixed effects in the model. The results indicate that there are still main effects of motion direction (Wald $\chi 2 = 504.23$, $p < .0001$) and of facing direction (Wald $\chi 2 = 24.62$, $p < .0001$), but no interaction between them (Wald $\chi 2 = 3.07$, $p = 0.08$). There is no main effect of width (Wald $\chi 2 = 1.46$, $p = 0.23$), no interaction between width and facing direction (Wald $\chi 2 = 0.95$, $p = .33$), and no three-way interaction among width, facing direction, motion direction (Wald $\chi 2 = 0.27$, $p = 0.60$). There was, however, a reliable motion direction x width interaction (Wald $\chi 2 = 11.37$, $p < .001$), indicating that width had a significantly different effect on the preferred position of all the rightward moving divers and fallers (i.e., both the forward and backward right-moving figures) when compared to the left-moving figures, regardless of facing direction. We are not

sure how to interpret this result. The present results thus confirm that both implied motion direction and facing direction influence the spatial composition of single-object pictures and that both factors produce biases opposite their own direction: rightward motion and/or rightward facing both favor leftward placement of the object and leftward motion and/or leftward facing both favor rightward placement of the object. These biases reinforce each other for forward-moving objects to produce a very large combined inward bias, but they conflict with and partly cancel each other for backward-moving objects. We have demonstrated these effects for human figures because it is relatively easy to depict implied forward versus backward motion through bodily poses. We presume that the results would be similar for inanimate moving objects (e.g., forward and backward moving cars), but this hypothesis is not easily tested because such objects, being rigid, look very much the same when moving forward versus backward in static pictures.

2.2 Speed Effects of Implied Motion

The results of Experiment 1 showed that the direction of implied motion strongly biases the preferred spatial composition of single object pictures and that its effects are substantially more powerful than the bias due to facing direction, at least with those images. In Experiment 2 we ask whether the speed of implied motion in a static picture might modulate these effects and, if so, how. Perhaps the most obvious prediction is that higher speeds might simply amplify the inward motion bias, as if more empty space in front of the moving object were required to accommodate the displacement of faster moving objects over a fixed future duration. This possibility implies that a picture of a left- or right-facing person would produce an increasingly larger inward motion bias when depicted standing, walking, jogging, and sprinting. There is another possibility, however. If observers were to conceive of the space behind the moving object as depicting the distance it had traveled in a fixed period of past time, the pattern of compositional biases would be reversed, with pictures of faster moving objects being preferred with less inward bias than more slowly moving objects. We investigated these possibilities by manipulating the speed and direction of implied motion for three types of objects – people, horses, and motor vehicles – using a variety of static cues to convey implied object speed, including the characteristic speed range of different types within the category (e.g., a tractor versus a race car), bodily poses of living objects (e.g., a galloping versus a walking horse), and the degree of motion blur in the background and/or certain moving parts of the objects.

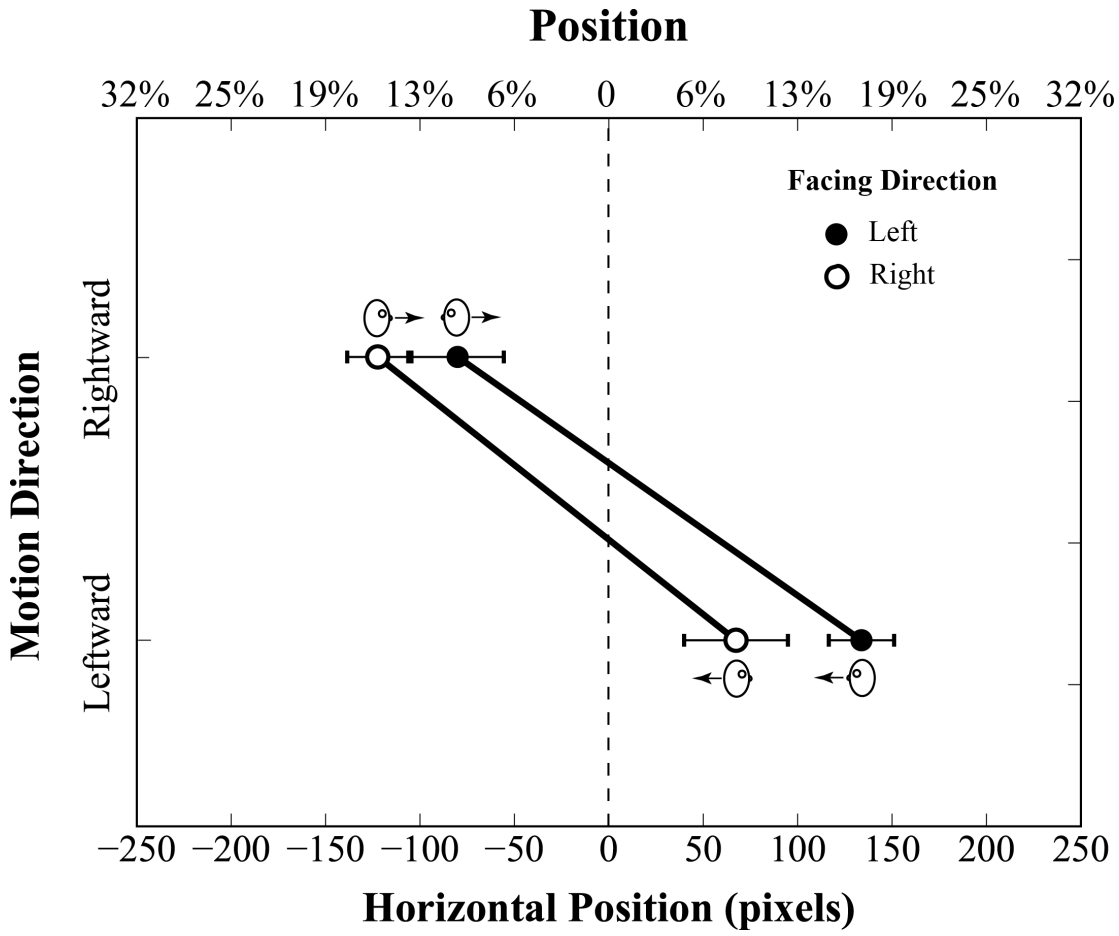

Figure 2.2: Results of Experiment 1. Average horizontal placement (x-axis) is plotted for rightward-moving divers/fallers (upper points) and leftward-moving divers/fallers (lower points), who are facing rightward (open circles) or leftward (filled circles). The top horizontal axis shows the percentage of the background image width from the center. The horizontally outermost points, indicating larger inward biases, represent the results for forward fallers/divers, whereas the inner points, indicating smaller inward biases, represent the results for backward fallers/divers. Arrows next to the face icons represent the direction of motion and the locations of facial features represent the direction of facing. Error bars indicate the standard errors of the mean. (Note that the canonical roles of the x- axis and y-axis are reversed in this graph to facilitate its interpretation.)

2.2.1 Methods

Design

The experiment consisted of a three-way factorial design: 3 object categories (person, horse, and motor vehicle) x 4 motion speeds (stationary, slow, medium, and fast relative to that object category) x 2 motion/facing directions2 (rightward and leftward) design. We used three different images in each cell of the 3 categories x 4 motion speeds design for a total of 36 images (see Figure 2.3). All participants performed two tasks. The first was a drag-and-drop task that was the same as described in Experiment 1: to place each object into a background image in the "most aesthetically pleasing" location (see below for details). Following completion of all trials in the drag-and-drop task, participants rated the relative speed of each object using a slider scale below the image as a manipulation check to see if our participants did, in fact, perceive the differences in relative speeds of implied motion.

Stimuli

Three object categories were used – human figures, horses, and motor vehicles – with three different objects sampled within each category at each speed. Four different speeds of implied motion were depicted relative to the characteristic range of speeds for the given object category: stationary, slow, medium, and fast. For the horse category, images of a motionless standing horse, a walking horse, a trotting horse, and a galloping horse were used to convey four different implied motion speeds. For the human figures, the implied speeds were conveyed with images of a stationary person standing, a person walking with a cane, a jogger, and a sprinting competitive runner. Finally, for the motor vehicle category, images of a broken-down car, a tractor, an ordinary passenger car, and a formula-one race-car were used. In each case, the specific examples of the object category at the different speeds were different. (See Figure 2.3 for examples.) To further convey the medium and fast implied motion speeds of images for all three object types, Adobe Photoshop was used to alter the backgrounds of the images by adding an appropriate degree of motion blur, augmented in the case of the faster moving cars by adding radial blur to the wheels. Rightward and leftward directed moving/facing versions of each image were also created by reflection about the central vertical axis in Photoshop, for a total of 72 images. Images of all objects were non-proprietary naturalistic photographs found on the internet. To maintain the realism of the original photographs, the original backgrounds of the objects (or the motion-blurred versions that were manipulated in Adobe Photoshop) were used in the drag-and-drop experiment. The regions where the objects originally appeared in the photographs were filled with pixels from neighboring regions in the image background using tools from

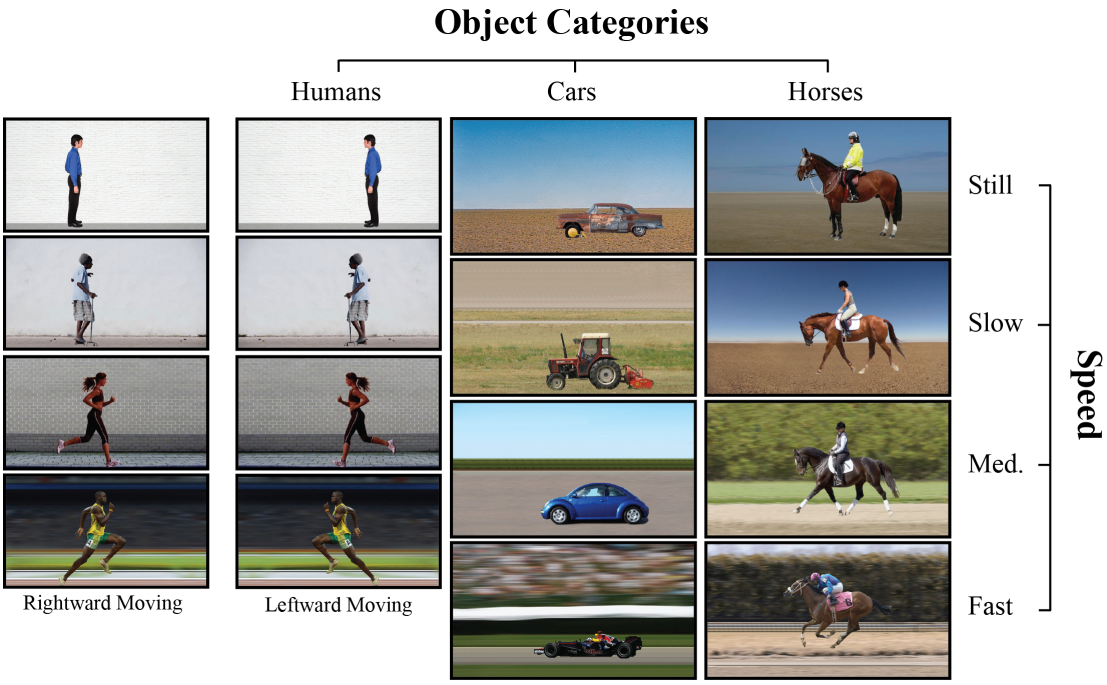

Figure 2.3: Examples of the images that were used in Experiment 2. Objects and background images were initially separate in the task, but integrated by participants when they dragged-and-dropped it. The objects are shown in the average horizontal positions that were selected by participants in the drag-and-drop task. Note the tendency to place faster moving objects closer to the center of the frame, which is most evident in comparing the still and the fast examples for each object category. (Rightward directed examples are shown only for the humans to save space.)

Adobe Photoshop in order to make them consistent with corresponding parts of the depicted background.

The dimensions of the background images were identical for the cars and horses (900 pixels in width, and 500 pixels in height). The cars and horses were all equal in width (400 pixels), although to preserve the correct aspect ratio, the heights of these objects varied (i.e., the slower car images are taller than the formula-one race cars). Being 400 pixels in width, the center of these objects could be placed no more than 250 pixels to the left and no more than 250 pixels to the right before one of their ends touched the bounding edge of the frame. The experimental program therefore did not permit participants to place any object beyond the -250 to +250 pixel range horizontally. The dimensions of the background images for the human object category were 749 pixels in width and 416 pixels in height and the human images themselves ranged from 67 pixels to 249 pixels in width. Procedure. Participants viewed the same computer screen from the same distance (about 70 cm) inside the same darkened booth as in Experiment 1.1 They performed two tasks: first "dragging and drop-

ping" the focal object of a picture to the position at which they found it most aesthetically pleasing (as in Experiment 1). Following a practice trial, participants were shown each of the 72 images of the different objects with their corresponding backgrounds in random order (intermixed with the objects and background images from Experiment 1). Only the horizontal (x-axis) position of the center of each object was recorded on each trial. Once participants completed the drag-and-drop task for all images, they were shown each of the objects in its rightward moving/facing version (in the center of the background used in the drag-and-drop task for that object) and were asked to make a slider scale rating to indicate how fast the object appeared to be moving. Participants rated the objects corresponding to each of the categories separately because we were primarily interested in people's perceptions of the relative speeds of the different objects within that category (rather than their absolute speeds) and wanted the range of speed judgments for each category to span the entire rating scale. The slider scale ranged from 0 to 500 pixels, and was located below the image. The scale was labeled "Slowest" on the left end and "Fastest" on the right end. We rescaled the data to range from 0 to 100.

2.2.2 Results and Discussion

We initially analyzed the ratings of speed (on the 0 – 100 scale) as a manipulation check to be sure that our participants perceived the speeds of the objects as intended. Indeed, the speed ratings, averaged over objects, increased monotonically from stationary (mean = 2.8), to slow (mean = 22.9), to medium (mean = 57.0), to fast (mean = 92.2). A two-way ANOVA showed that the speed ratings were indeed significantly different ($F(3,54) = 687.5$, $p < .0001$, $\eta = .974$). There was no significant effect of category ($F(2,36) = 2.97$, $p = .064$, $\eta = .142$) but there was an interaction between speed and category ($F(6,108) = 4.44$, $p = .0001$, $\eta = .198$). This interaction results from slight non-parallelism of the speed ratings for the slow and medium speeds for the different object categories. These results generally corroborate the fact that our manipulation of the four speed conditions was successful and that the differences between the category conditions were relatively minor. Analyses of the drag-and-drop placements were conducted on the average of the recorded x-axis coordinates corresponding to the horizontal positions of the centers of the three examples of each object in the drag-and-drop task. We analyzed the effects of object category, motion/facing direction, speed, and image width on the preferred horizontal positions with linear mixed-effects models using the lme4 package (Bates and Sarkar, 2016) in the R environment (R Development Core Team, 2016) (Pinheiro et al., 2010). In all of the models, we specified subjects as the random factor in order to control for their intraclass correlation using random intercepts for each subject. Motion/facing direction,

speed, and image width were specified as fixed factors for the main model we used to test the effect of implied speed. We included width as a fixed factor in order to test for effects of motion/facing direction and speed while controlling for variations in the horizontal extents (widths) of the target images, even though differences in target width were only present in the images of people. Horse and car images had a fixed width of 400 pixels. We performed statistical significance tests using type II Wald chi-squared tests of the fixed effects in the model. The results of the experiment are plotted in Figure 2.4. The pattern of the data is quite clear and systematic. Inward biases are evident for every condition, but somewhat surprisingly, they are larger for the objects perceived as moving at slower speeds. In fact, the largest inward bias is evident for the stationary conditions. Statistical analyses corroborated this pattern in the results. In keeping with previous findings, we observed a strong inward moving/facing bias (Wald $\chi 2 = 691.56$, $p < .0001$), showing that participants preferred leftward moving/facing objects to be placed right-of-center and rightward moving/facing objects to be placed left-of-center. This result replicates the robust inward bias previously reported for aesthetic preferences for the spatial composition of static, single-object pictures inside a frame (Bertamini et al., 2011), and Experiment 1 of this article). No main effect of speed was evident, because of the cancelling effect when the left and right moving/facing versions were averaged (Wald $\chi 2 = 5.94$, $p = 0.11$), but a significant motion/facing direction x speed interaction was observed (Wald $\chi 2 = 47.25$, $p < .0001$), resulting from the fact that the inward moving/facing bias becomes weaker as the object is seen as moving faster. There was no main effect of image width (Wald $\chi 2 = 0.0116$, $p = 0.91$), no interaction between motion/facing direction and width (Wald $\chi 2 = 0.089$, $p = 0.77$), and no interaction between speed and width (Wald $\chi 2 = 1.21$, $p = 0.75$). There was, however, a significant three-way interaction among motion/facing direction, velocity, and width (Wald $\chi 2 = 10.88$, $p = 0.012$). The data in Figure 2.4 are averaged over the three object categories because there was no main effect of category (Wald $\chi 2 = 2.84$, $p = 0.24$), no category x facing interaction (Wald $\chi 2 = 3.02$, $p = 0.22$), and no category x speed interaction (Wald $\chi 2 = 4.62$, $p = 0.59$). There was a significant category x motion/facing direction x speed interaction (Wald $\chi 2 = 18.19$, $p < 0.01$) due to the fact that the shapes of the speed curves are not parallel for the different object categories. The pattern of the data does not seem to be systematic in ways that suggests a coherent, meaningful interpretation, however, except for showing smaller speed effects for the horse images than for the people or car images (see Figure 2.5). This might result from people simply being less familiar with the speeds at which horses move at different gaits. Finally, we computed the correlation between the average magnitude of the speed ratings for each of the pictures and the average magnitude of the inward motion/facing bias in the drop-and-drag task. The results showed a reliably negative correlation ($r = -0.60$, $p < 0.0001$), indicating that pictures of objects with higher speed ratings were preferred in positions closer to the

center (i.e., with smaller inward biases), see Figure 2.7. This negative correlation is the opposite of what we originally expected, given our initial hypothesis that increased speed would amplify the inward moving bias obtained in Experiment 1.

These considerations lead us to examine the positional responses for the four speed conditions for each participant to find out whether some people are relying on the "room to move" principle, which would produce monotonically increasing curves (as objects move faster they get closer to the back edge of the frame), and others on the "distance traveled" principle, which would produce monotonically decreasing curves (as objects move faster they get farther from the back edge of the frame). The results are graphed in Figure 2.6. The graph shows the average positions of all the images in each of the speed conditions, for each subject. In other words, each line shows the data for one unique subject, and each of the four points on the line indicates the average positional data for each of the four speed conditions. The color of the line indicates the degree to which a subject's data is consistent with the "distance traveled" or "room to move" principle, although without the monotonicity constraint. The color is directly proportional to the overall average of all six pairwise differences between the positional data for each of the speed categories. The more a subject tended to place faster target objects ahead of slower target objects on average, the warmer (redder) the curve appears, the more a subject tended to place faster target objects behind slower target objects on average, the cooler (bluer) the curve appears. The results indicate that while few subjects have curves that change monotonically, most subjects adhered to a looser version of the "distance travelled" principle: They placed the faster target objects ahead of the slower target objects on average, although their curves do not necessarily show a monotonic decrease in inward bias across speed levels. Some subjects appeared to adhere to a very weak "room to move" principle, in the sense that they placed faster target objects behind slower target objects on average, although the average difference between the placements of the slower versus the faster target objects was smaller for the subjects that adhered to the "room to move" principle than it was for those who adhered to the "distance traveled" principle.

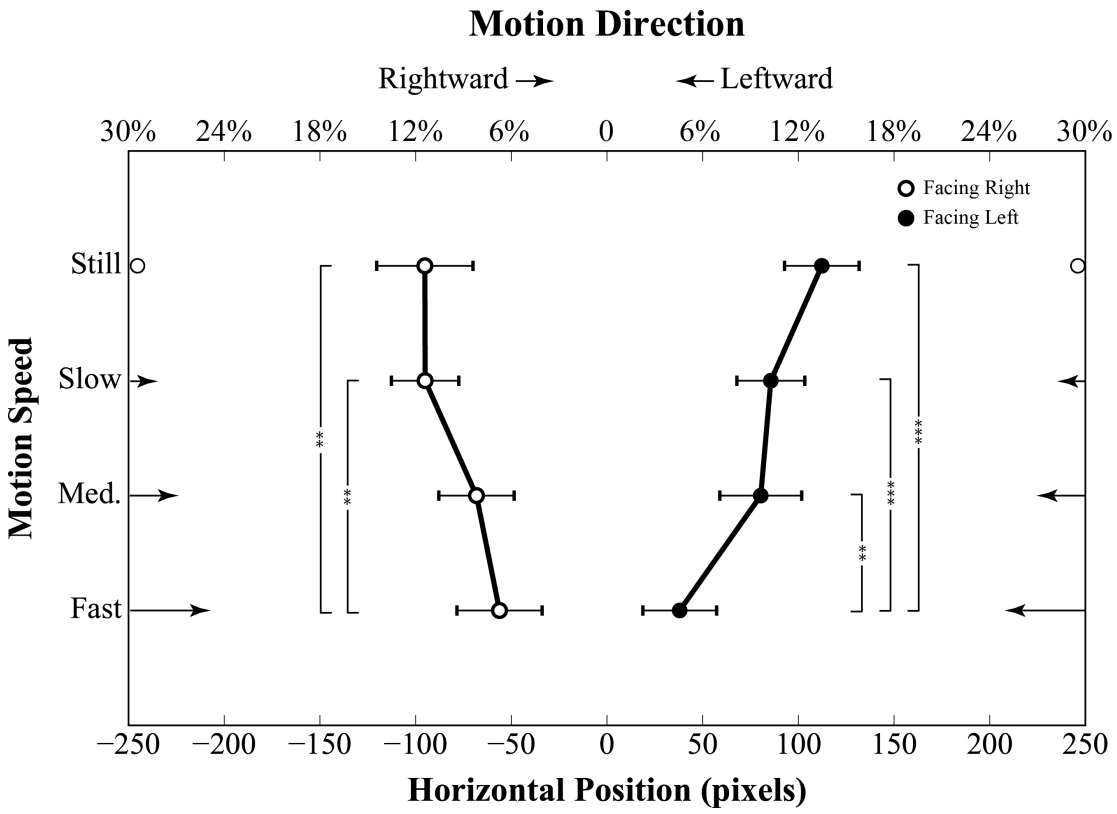

Figure 2.4: Results of Experiment 2. Average horizontal position (x-axis) of the center of target objects at their chosen positions, is plotted as a function of moving/facing rightward (open circles) versus leftward (filled circles) and implied object speed (y-axis). The top horizontal axis shows the percentage of the background image width from the center. (Because the width of the background images for the human figures was different from the width of the background images for the horses and the cars, this shows the percentage of the average background image width from the center). Asterisks represent statistical significance of specific comparisons as indicated, using Bonferroni corrections for multiple comparisons (** $p < .05/12 = .004$). Error bars show the standard error of the mean for each speed condition, for each facing direction.

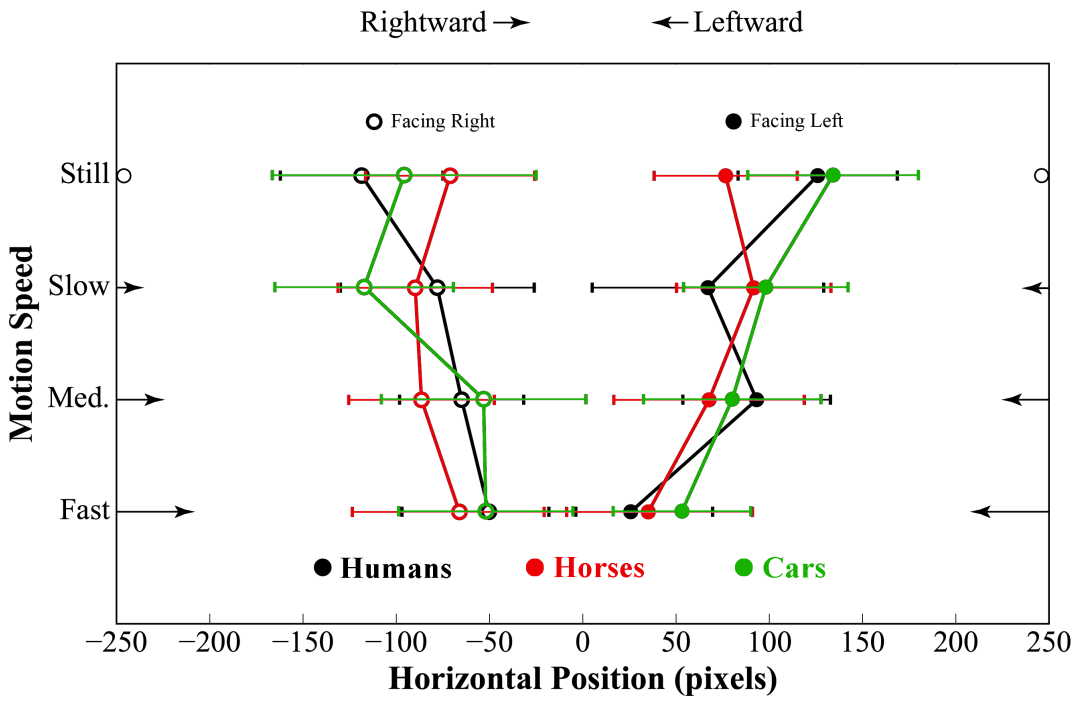

Figure 2.5: Further results of Experiment 2. Average horizontal position (x-axis) of the center of target objects at their chosen positions, is plotted as a function of moving/facing rightward (open circles) versus leftward (filled circles) and implied object speed (y-axis) for the three object categories (humans, horses, and cars). Error bars show the standard error of the mean for each speed condition, for each facing direction.

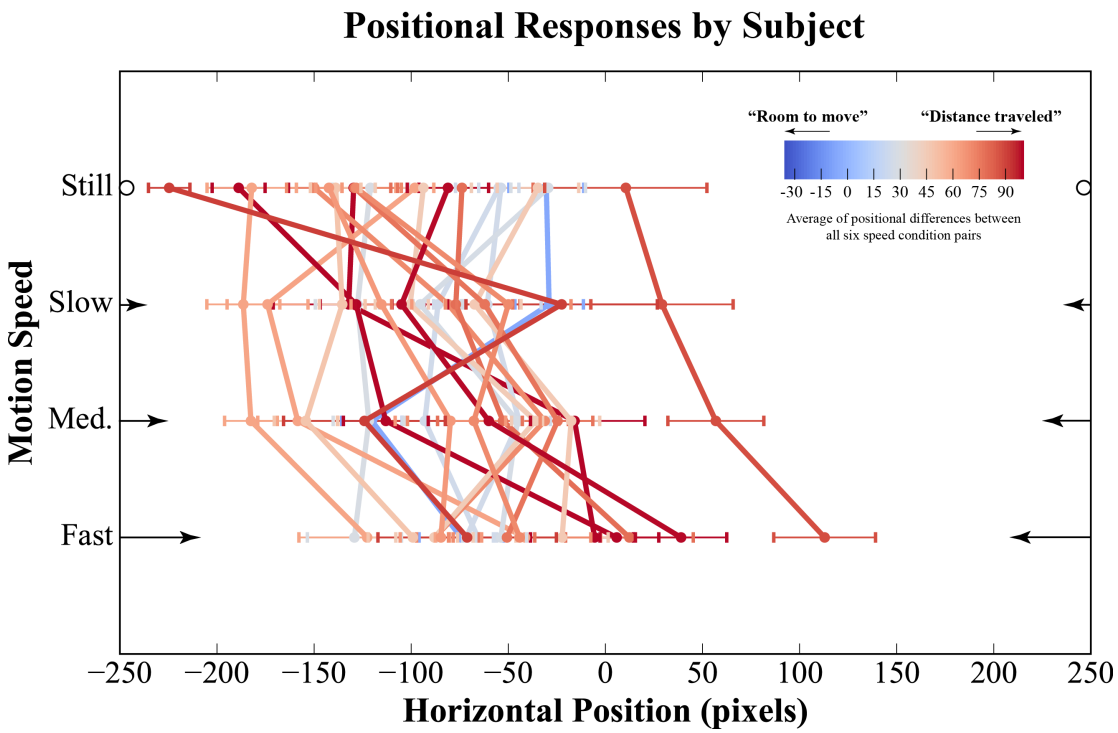

Figure 2.6: Average horizontal position (x-axis) of the centers of target objects in each speed condition, plotted as a function of motion speed (y-axis), for each subject. The left facing and right facing examples are merged (as if all were right-facing), and each line corresponds to a unique subject. The redder lines correspond to subjects who placed faster target objects ahead of slower target objects on average ("distance traveled" principle). The bluer lines correspond to subjects who placed faster target objects behind slower target objects on average ("room to move" principle). For each subject, the saturation of the line color is directly proportional to the average of the six pairwise differences in the positional data corresponding to each of the four speed categories (The average of the six following differences: Fast – Med, Fast – Slow, Fast – Still, Med – Slow, Med – Still, Slow - Still). Error bars show the standard error of the mean for each subject, for each speed condition.

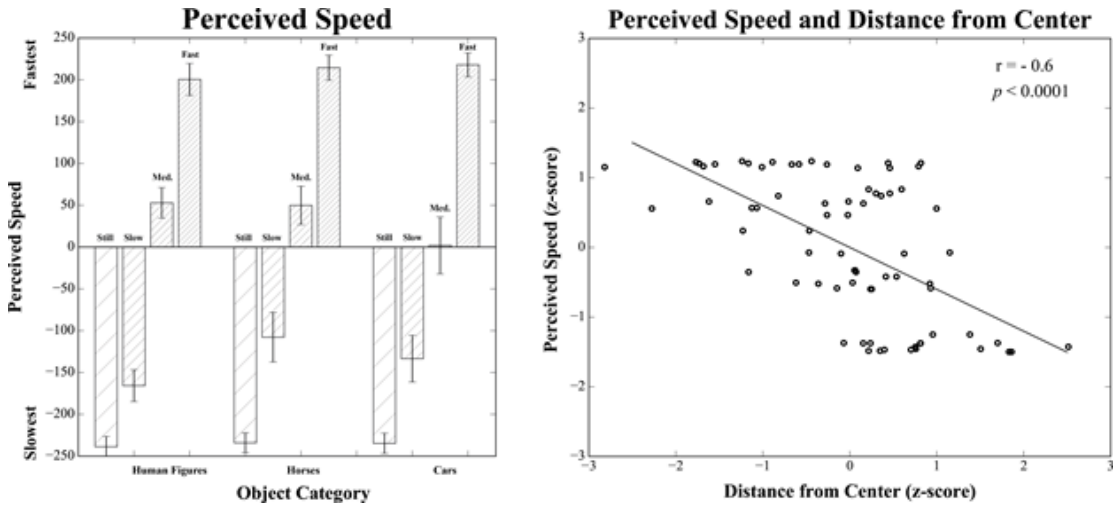

Figure 2.7: Subjective speed ratings of the foreground objects used in the experiment, and correlation between subjective speed ratings and horizontal placement (z-scores). The correlation between the average magnitude of the speed ratings for each of the pictures and the average magnitude of the inward motion/facing bias in the drop-and-drag task are shown, confirming a reliably negative correlation ($r = -0.60$, $p < 0.0001$), indicating that pictures of objects with higher speed ratings were preferred in positions closer to the center (i.e., with smaller inward biases). This negative correlation is the opposite of what we originally expected, given our initial hypothesis that increased speed would amplify the inward moving bias obtained in Experiment 1, but it confirms that our visual manipulations of the foreground objects were consistent with our participants' subjective judgments of their implied speed.

2.2.3 General Discussion

The results of the experiments described here replicate and extend the inward bias in aesthetic preference for spatial compositions of pictures (Palmer et al., 2008; Sammartino and Palmer, 2012a, 2012b; also called the "anterior bias" by Bertamini et al., 2011). It occurs for pictures of a single object in a rectangular frame when the object faces leftward or rightward in the picture plane and produces a preference for the object to be positioned on the side of the frame opposite to the direction it faces. Here we explored the inward bias further by investigating whether an analogous effect is also present for implied motion of the depicted object. In Experiment 1, we showed that although the inward bias due to the facing direction of objects (the inward-facing bias) is still evident in people's aesthetic preferences for composition of moving objects, it can be overpowered by an inward-moving bias when the object is moving backward. That is, when the facing and motion directions are the same, as in forward motion, people favor compositions exhibiting a strong inward bias in which the object both moves and faces inward. When the facing and motion directions are opposite, however, as in backward motion, they favor compositions in which the object is moving into

the frame but facing (somewhat) out of the frame. These effects were found using images of forward and backward human divers and fallers, because it was relatively easy to convey a clear direction of motion through diving and falling poses that could easily be perceived as forward or backward motion. Finding a case in which the inward facing bias is overwhelmed by another factor (motion direction) was somewhat unexpected. Both the inward-facing bias and the inward-moving bias are potentially compatible with the affordance space hypothesis, given that asymmetries are likely to arise in the shape of the affordance spaces of divers and fallers due to both the direction that they are looking (the facing bias) and the direction they are moving (the inward-moving bias). The results of Experiment 1 show that these two compositional biases are largely additive. One theory about the inward-facing bias is that the preferred displacement of a side-facing object is to be off-center, facing inward, because people are actually centering an object's affordance space, which is suggested to extend farther in the front-facing direction of an object than in other directions for most objects (Sammartino and Palmer, 2012b). The results of Experiment 1 are compatible with an affordance space hypothesis, provided that the affordance space on an object can differ dynamically, depending on the perceived movement and action of the object in the depicted event. Experiment 2 investigated how the inward-moving bias might be modulated by the implied speed of the depicted object. Although the inward moving/facing bias held in this experiment as well, varying the speed of implied motion showed that faster-moving objects exhibited a smaller inward bias than slower-moving objects. In other words, faster objects were preferred closer to the center than objects that appeared still or moving slowly. We further validated these results by collecting ratings from the same participants on how fast the objects appeared to be moving and found that these ratings were negatively correlated with the magnitude of the inward moving/facing bias (i.e., faster moving objects produced smaller inward biases). The results of Experiment 2 thus replicate the robust inward motion/facing bias in aesthetic preferences demonstrated in the inward motion bias of Experiment 1, but they also reveal a somewhat surprising reduction in that bias for objects depicted as moving faster. The reduction is surprising because the data from Experiment 1 clearly imply that people systematically prefer more space to be present in the direction of implied motion. We take a "room-to-move" hypothesis to be the most straightforward explanation of the directional bias effects reported in Experiment 1. By extension, however, this account predicts that more (i.e., faster) implied motion should produce a larger inward bias to provide more space in front of a faster moving object than a slower or stationary one. Unfortunately, we observed the opposite in the speed effects obtained in Experiment 2: faster objects were preferred with less space in front of them and more space behind them. There is a coherent rationale for these speed effects in Experiment 2, however. Perhaps people prefer the spatial position of an implicitly moving object to represent how far it has traveled in some fixed amount

of time in the recent past from an off-center reference position. If the inward bias for a stationary object is taken as a baseline, this "distance-traveled" hypothesis thus implies that the inward bias will progressively diminish as the speed of the object increases. Perhaps the simplest way to conceptualize the relation between these two hypotheses is that room-to-move is a future-oriented principle of motion foresight whereas distance-traveled is a past-oriented principle of motion hindsight. The results of Experiment 1 thus seem to conflict with those of Experiment 2 in the sense that the directional bias so evident in Experiment 1 suggests a future-oriented (room-to-move) account of implied motion effects, whereas the speed effects apparent in Experiment 2 support a past-oriented (distance traveled) account. How might this conflict be reconciled? It seems unlikely to be due to differential context effects, because the data for both studies were collected in a single session with the same participants and with the trials randomly intermixed, precisely to avoid such context-specific strategic effects. One possible reconciliation is that people's compositional preferences are determined by two different processes, one that depends on the perceived direction of motion and the other on the perceived speed of motion. Initially, the viewer may identify a reference position for a picture of a stationary object that might move based on the directionality of its characteristic motion as depicted by its facing direction. This initial reference position exhibits a strong inward bias with an expanse of empty space ahead of it where it might be located in the near future (i.e., representing room-to-move). This initial reference position is then modified according to the object's perceived speed, including an expanse of space behind it to represent where it had been located during a fixed period of time in the recent past (i.e., representing distance traveled). This two-process account seems to be a plausible explanation for the pattern of results obtained in these two experiments. The present data thus appear to provide evidence of biases in aesthetic preferences for the spatial position of a single object that are based on assumptions about both the prospective and retrospective motion of the depicted object.

Although these results were obtained in the context of aesthetic preferences, the results from the second experiment are reminiscent in some ways of representational momentum effects observed in the context of spatial memory (Freyd & Finke, 1984; Hubbard, 2005). Representational momentum refers to a bias in an observer's memory of a previously viewed moving target to be displaced forward in the direction of the target's motion. Freyd (1983a and 1983b) showed pairs of "frozen action" photographs taken from the exact same action (such as a sequence of two photographs of a girl jumping in the air, or a person jumping off a wall). The pictures were shown either in the correct temporal order, or in the reverse temporal order. The first picture was presented for a short 250 ms, and was followed by a 250 ms retention interval. Next, the second picture (either in the correct temporal order, or the reverse temporal order) appeared and remained on the screen until the viewer indicated

if it was the "same" as the first picture, or "different," via a key press. Response times measuring the interval between the visual onset of the second picture and the key press revealed that subjects took on average 59 ms longer to reject a distractor (when the second picture was different from the first), if the distractor was in the correct temporal order, as opposed to the reverse temporal order (Freyd & Finke, 1984). This effect illustrates a representational bias in people's spatial memory for simulating implied motion forward in time in ways that are consistent with spatial-temporal correlations in the natural world, and even physical properties of objects and motion (Hubbard, 2005). Our results appear consistent with these representational momentum effects, but are different in important ways. First, the reduction in the inward bias for objects with faster implied motion does not necessarily reflect a bias in spatial memory, since subjects were never asked to remember the placements of these objects and to reproduce them in the drag-and-drop task (they were never even shown the images before the task). Our results are more generalized than the representational momentum effects mentioned here, and reveal a general bias for placing faster objects of completely different kinds closer to the center, as if people were simulating their displacement forward in time. While these results may not be due to particular biases in spatial memory, they may reflect representational biases due to perceptual priors shaped by the accumulated experience of spatial-temporal correlations and physical dynamics of moving objects in the world. An alternative to the ecological explanation, which explains the reduction in inward bias for faster objects in terms of perceptual priors shaped by the accumulated experience of spatial-temporal correlations and physical dynamics of objects in the world, is that our results owe more to semantic and narrative factors than to visual experience. Aesthetic preferences for spatial compositions have been shown to be extremely sensitive to semantic context (Sammartino and Palmer, 2012). Sammartino and Palmer (2012) found that while people tend to prefer standard compositions with a neutral title (a title that simply describes the content of the picture), they preferred nonstandard compositions if the caption "fit" the spatial implications in the composition (for example, a rear-view of a plane with the title "Departing") (Sammartino & Palmer, 2012). While semantic labels were not used in the experiment presented here, the strong spatial implications conveyed by the visual characteristics of the objects and background images (the varying degrees of implied motion speed and directionality of motion that we manipulated) may have influenced our participants' compositional choices nonetheless.

2.3 Off-Center Compositional Biases Revealed Through Serial Reproduction

Where do aesthetic preferences for the spatial composition/arrangement of objects come from? Could biases in spatial memory be predictive of aesthetic preferences? In two follow-up experi-

ments, we asked whether the inward bias in people's aesthetic preferences could be due to spatial memory biases. We conducted two serial reproduction experiments in which subjects viewed a stationary object with a clear facing direction (we used images very similar to those used in the previous experiments). The task was to remember the exact position of the object after it was presented briefly, following a delay. We created a serial reproduction chain with 10 iterations. Figure 2.8 shows the images used, as well as kernel density estimates of the initial uniform distribution of object positions, and of the positions of the objects at the 10th and final iteration. Although an inward bias isn't present, there are clear lateral (off-center) biases. It is unclear whether these priors are responsible for aesthetic biases, and additional experiments that include objects with both facing directions within a single experiment may be needed. However, the clear tendency for people to remember the position of facing objects towards the side remains.

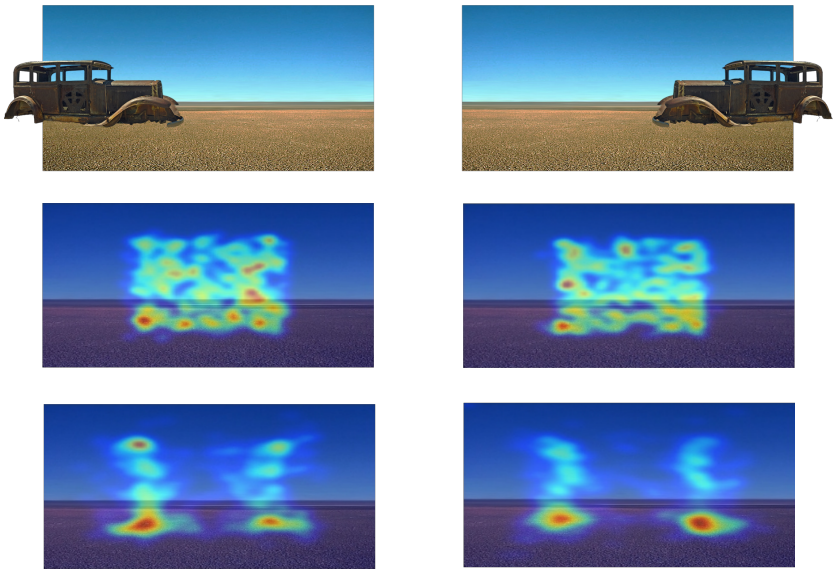

Figure 2.8: Results obtained for the remembered position of stationary right-facing and left-facing objects similar to ones used in the previous experiments following ten iterations of a serial reproduction chain. The object was initialized in a position within the background frame sampled from a uniform distribution. After ten iterations, a clear off-center bias emerges in the remembered position of the object, regardless of its facing direction, both for the experiment in which the right-facing object was shown, as well as the experiment in which the left-facing object was presented. An important difference between these experiments and those presented in the previous section is that for the serial reproduction memory experiments, only one facing direction was ever shown.

3

Uncovering Visual Priors in Spatial Memory Using Serial Reproduction

3.1 Spatial Memory Priors for Simple Geometric Shapes

Retrieving detailed visual information from memory requires efficient representations of often complex and noisy visual scenes. In Bayesian accounts of reconstruction from visual memory, the memory system integrates sensory information with knowledge acquired from previous experience ("priors"). Effective use of this information may reduce variability in visual memory and improve overall reconstruction accuracy (Weiss, Simoncelli, & Adelson, 2002). Using priors is usually advantageous because they capture regularities in the structure of the world that are innate or observed over our lifetimes. However, this can lead to substantial biases during reconstruction. This is because prior information may deviate significantly from our observations, especially when a visual scene is unexpected given previous experience.

The first sections on priors for geometric shapes were taken from Langlois, Jacoby, Suchow, and Griffiths (2017).

In many cases, visual priors are categorical (or prototypical), represented in memory as schematic or simplified objects (Huttenlocher et al., 1991). In one experimental paradigm that reveals categorical effects in the context of spatial memory, participants are asked to remember the location of a dot presented within a circle or other bounding shape. After a brief presentation and a delay, participants reproduce the dot's location by placing it in the recalled position (see Figure 3.1). Results indicated that participants tend to misplace dots toward a central (prototypical) location in each of the quadrants of a circle (Huttenlocher et al., 1991). Following these results, others tried to predict prototypical positions in spatial memory for dots presented inside a variety of geometric shapes (circle, square, triangle, vertical oval, horizontal oval, and pentagon) (Wedell et al., 2007). In the study, participants were shown thirty-two dots aligned along two concentric circles within each shape. A parametric model with four components (prototypes) was fitted to the remembered positions of the dots, confirming that visual memory of these shapes shows substantial categorical effects.

Past parametric approached to characterizing categorical biases have a number of limitations—specifically, a relatively small number of to-be-remembered locations, and noisy measurements (responses following only a single iteration of the task), resulting in limited resolution for capturing the locations of the categories. In addition, earlier investigators used a parametric model with a fixed number of categories (the Category Adjustment Model, or CAM). The choice of the model, and the number of categories that were used were often not fully justified, requiring certain *a priori* assumptions. Here, we propose to use a paradigm based on serial reproduction to characterize visual memory biases without needing to rely on parametric modeling and with substantially better resolution and accuracy.

In this chapter, we explore spatial memory priors in a task where participants were asked to remember the position of a small black dot inside a variety of geometric shapes. Operating under the assumption that people share the same inductive biases, or spatial memory priors, we show that serial reproduction appears to converge on these priors remarkably quickly, revealing patterns that are consistent with some established findings, although in many cases revealing new and intricate patterns that were previously unknown. Finally, we demonstrate the advantages of using a non-parametric kernel density estimation procedure to characterize the prior.

3.1.1 Methods

Participants

Participants were recruited online using Amazon Mechanical Turk. All gave informed consent. The experimental protocol was approved by The Committee for the Protection of Human Subjects

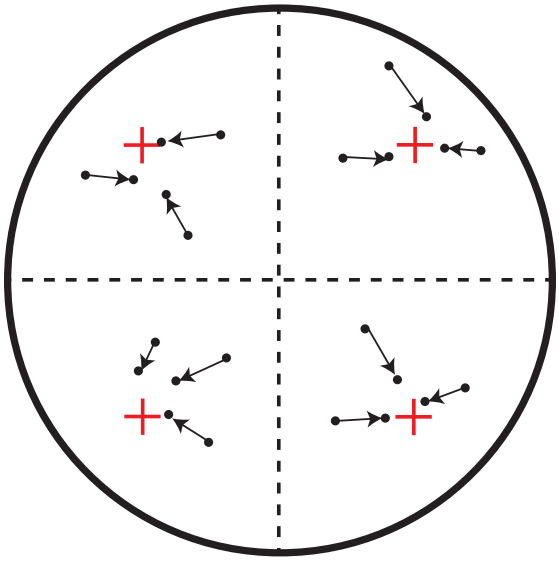

Figure 3.1: Illustration of prototype effects in memory for points in a circle. The red crosses represent prototypes, and the small points are typically misremembered as being closer to those prototypes.

(CPHS) at the University of California, Berkeley. Each experiment required approximately 70-100 participants. A total of 570 participants took part in Experiment 1 and an additional 590 took part in Experiment 2.

Stimuli

All images were approximately 400 × 400 pixels in size. Each shape was a 6-pixel-wide black outline over a white background. The sizes and colors of the backgrounds for the images were intended to ensure that the images would be clearly visible in any standard browser window (unlikely to become occluded), and such that the boundaries of the images would be invisible.

Procedure

We carried out a series of serial reproduction experiments. Participants were presented with timed displays (a shape outline with a dot initialized somewhere within the boundaries of the shape), and were instructed to reproduce the exact location of the dot inside of the shape. Once complete, their response was sent to another worker (again, as a timed display), who was instructed to reconstruct this display from memory, and so on. A total of ten iterations were completed for each chain. See

Figure 3.2 for a schematic diagram of the serial reproduction procedure.

Practice trials. Participants completed ten practice trials in order to become familiar with the user interface. During these practice trials, they were presented with a circle (a black outline of a circle over a white screen), with a dot initialized somewhere within its boundaries. This display was presented for 4000 ms, followed by a blank screen lasting 1000 ms. Next, the circle was presented without the dot and remained on the screen until the participant positioned the dot in the location that they remembered. As soon as the participant clicked, the dot appeared under the mouse cursor. Participants could reposition the dot as many times as they needed. Once done, they pressed a button to proceed to the next trial.

Experimental trials. Following the ten practice trials with the circle, there were ninety-five experimental trials with exactly one of the shapes. In Experiment 1, the shape could be a circle, triangle, square, vertical oval, horizontal oval, or a pentagon. In Experiment 2, the shape could be a regular polygon with more than five vertices. For each of the 95 experimental trials, the presentation time was reduced to 1000 ms. As with the practice trials, the position of the shape on the screen was randomized somewhere inside a larger canvas in order to control for participants resorting to tracking the position of the dot by trivially marking its absolute position on their computer screens. In addition, participants were given trial-by-trial feedback regarding their accuracy. If their responses were within eight percent of the width and height subtended by the shape on the screen, they were told that their response was accurate (a message in green font: "This was accurate"), and received a small monetary bonus. If not, they received no bonus beyond the basic payment for the HIT, in addition to any bonuses accrued from the previous trials, and were presented with a red message ("this was not accurate"). These trials were discarded from the experiment. Participants could not provide multiple responses within a chain.

Experiment 1: We used the same six shapes as Wedell et al. (2007): A circle, equilateral triangle, square, vertical and horizontal ovals, and a regular pentagon (Wedell et al., 2007). For each shape, we initialized the position of five hundred dots within its boundaries (for the circle, we initialized four hundred dots).

Experiment 2: Because our method revealed a variable number of peaks (prototypes) in the prior for the angular shapes in Experiment 1, and that these appeared to be due in large part to the number of vertices in the polygons (all were regular polygons—an equilateral triangle, square, and pentagon), we wanted to determine the point at which the prototypes begin to merge into the four prototypes in the prior for the circle. We did this by conducting the same experiment with polygons containing increasingly more vertices (approximating a circle more closely as vertices were added).

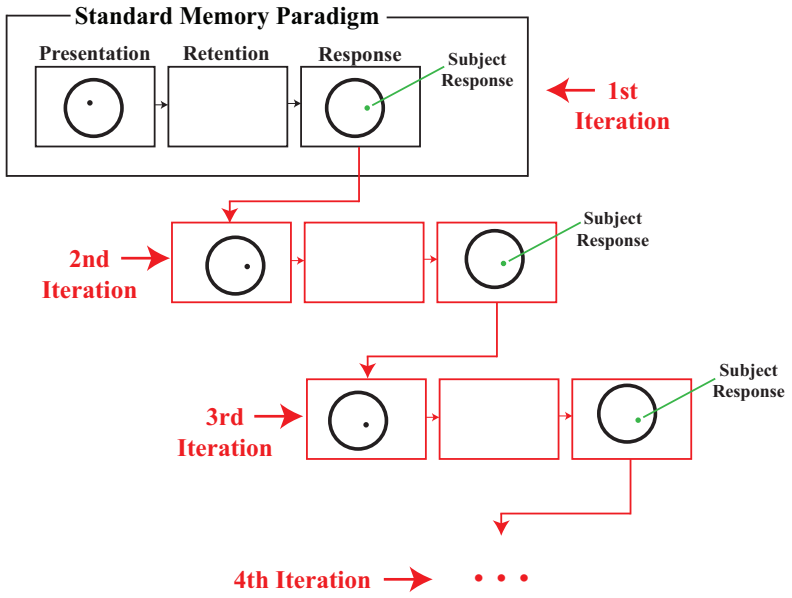

Figure 3.2: Serial reproduction chain for one trial in the memory task, illustrating the iterative process for a single dot being remembered. The trial in black represents a standard memory paradigm. In red are additional iterations of the task using the result from the previous iteration as the new stimulus, which form the nodes in the serial reproduction process.

3.1.2 Results

Our results are presented in two parts. First, we present all our results for Experiment 1, in which we used the same shapes as Wedell et al. (2007). We demonstrate that using a serial reproduction paradigm, as well as non-parametric kernel density estimation, replicates some (but not all) of their key findings. In particular, we find four prototypes arranged in the centers of the four quadrants of the circle, as well as the horizontal and vertical oval shapes, as they did. However, we also show new and intricate patterns in the priors for the angular shapes (triangle, square, and pentagon). We evaluate the predictions of the model by Wedell et al. (2007) on the data we obtained following one iteration, for all the shapes, and compare them to predictions that we obtain from our estimates of the prior following all ten iterations. In addition, we show quantitative evaluations of the change in copying accuracy for the equilateral triangle. Second, we show the results for Experiment 2, where we illustrate the effect of adding vertices to regular polygons on the prior, revealing hitherto unknown grouping effects of the prototypes in spatial memory that occur as regular polygons begin to approximate a circle.

Measuring spatial memory priors

Serial reproduction results. Figure 3.3 shows visualizations of the estimates that we obtained following ten iterations of the serial reproduction experiment using four hundred initial seeds for the circle. Each panel shows the results for each of the ten iterations, including the initial seeds. Notice that the prototypes begin to emerge in as early as the fourth iteration. For the panel showing the results of the tenth iteration, we show an estimate of the prior using our non-parametric kernel density estimate in lieu of plotting the points. Figure 3.4 shows the results that we obtained for the equilateral triangle. Notice the emergence of bimodal peaks near each of the vertices. This finding suggests that for this shape, there are a total of six prototypes in the prior, grouped in pairs at each corner.

Simple shapes. In Figure 3.5 (panels B and D), we show the kernel density estimates that we obtained for all the shapes. In the case of the circle, vertical oval, and horizontal oval, our results are consistent with past findings (shown in panels A and C). However, we discover bimodal peaks in the vertices of the angular shapes (prototype pairs clustered at each of the corners). This result is particularly striking for the triangle and the square shapes. The same result is present for the pentagon shape, although unlike the peaks in the prior for the triangle and square, those in the pentagon are not quite rotationally invariant, although all three geometric shapes are, suggesting that the shapes and orientations of the modes in the priors are not a simple function of the presence of edges, or the angles at these edges.

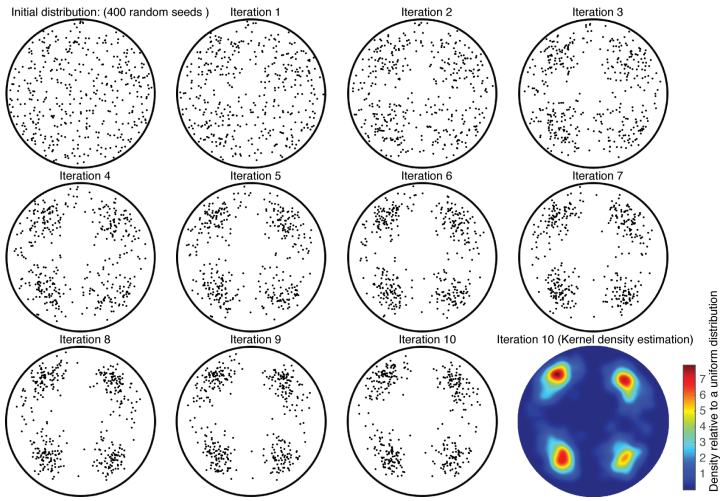

Figure 3.3: Serial reproduction of 400 dots presented in a circle, for ten generations (iterations) of the process. The top left scatterplot shows the positions of the original seeds (sampled from a uniform distribution) inside the circle shape. The remaining subplots show scatterplots of the results of the serial reproduction chain for iterations 1-10. The subplot of the tenth iteration, in the bottom right, also shows the kernel density estimate. Note that from one iteration to the next, points that were originally scattered uniformally within the circle boundary begin to converge on each of the nearest prototypes at the center of each of the four quadrants in the circle. By the tenth iteration of the process, four clusters are clearly discernable.

Convergence analysis. For the triangle results, we completed a convergence analysis (See Figure 3.4, panels B and C), using the Jensen-Shannon divergence (JSD). To estimate the variability of these JSDs, we generated 100 bootstrapped data sets sampled from the original data (with replacement). For each one, we computed the JSDs of consecutive iterations (see panel B). The JSD between the initial distribution and iteration 1 was significantly larger than that between the two final iterations ($p = 0.02$) and there were no significant differences between the distance between iterations 9 and 10 compared with iterations 8 and 9 ($p = 0.43$).

As another measure of convergence, we also computed the JSD between all iterations and the last iteration (Jacoby & McDermott, 2017) (see panel C). The distance between the last two iterations was significantly smaller than the distances between iteration 10 and each of the remaining iterations (0 through 8). The distance between iteration 10 and 9 was marginally larger than the distance between iteration 10 and 8 ($p = 0.041$). These analyses suggest that convergence occurs at or near the tenth iteration. To test if the responses of participants became more "prototypical" over the course of the experiment (as they progressed through their trials), we used the estimate of the prior from the final iteration to measure the average log-likelihood of their responses. We used data from

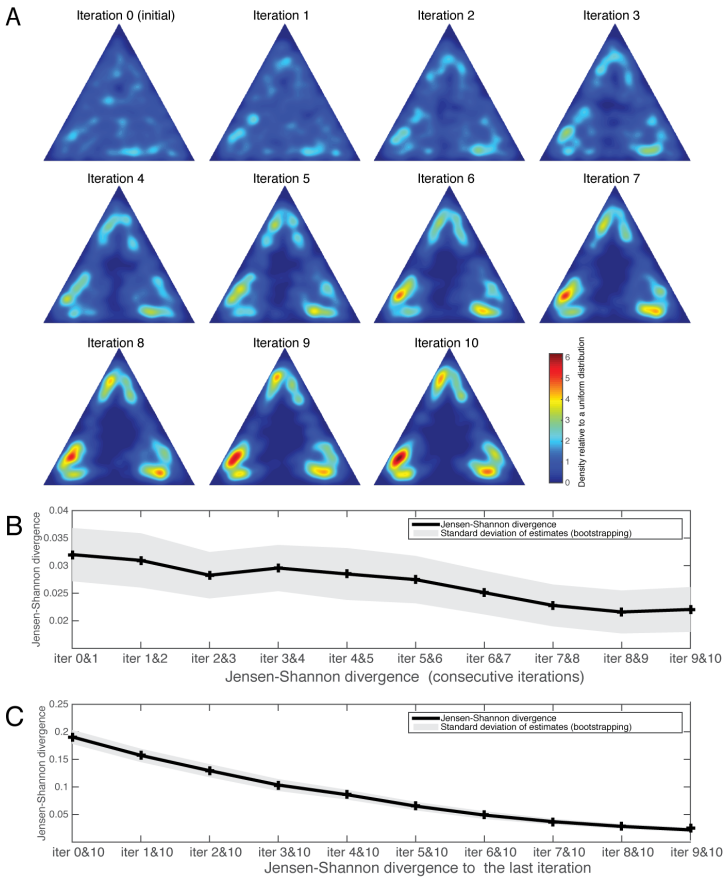

Figure 3.4: Results we obtained for each iteration in the chain for the triangle shape. A. Kernel Density Estimate (KDE) for the initial distribution and the 10 iterations. B. Convergence analysis using the Jensen-Shannon divergence (JSD) between consecutive iterations. C. JSD between all iterations and the last iteration. Note that both measures decrease with the number of iterations, and suggest that convergence occurs at or near the tenth iteration in the serial reproduction chain.

the 83% of the participants who performed more than 80% of the trials within the accepted criteria (responses within 8% of the height and width of the shape on the screen). We found that the log-likelihood significantly improved when comparing the first and second half of their responses ($t(49) = -2.47$, $p = 0.008$), and when comparing the first 10 trials to the last 10 trials of each of the subjects ($t(49) = -2.04$, $p = 0.046$).

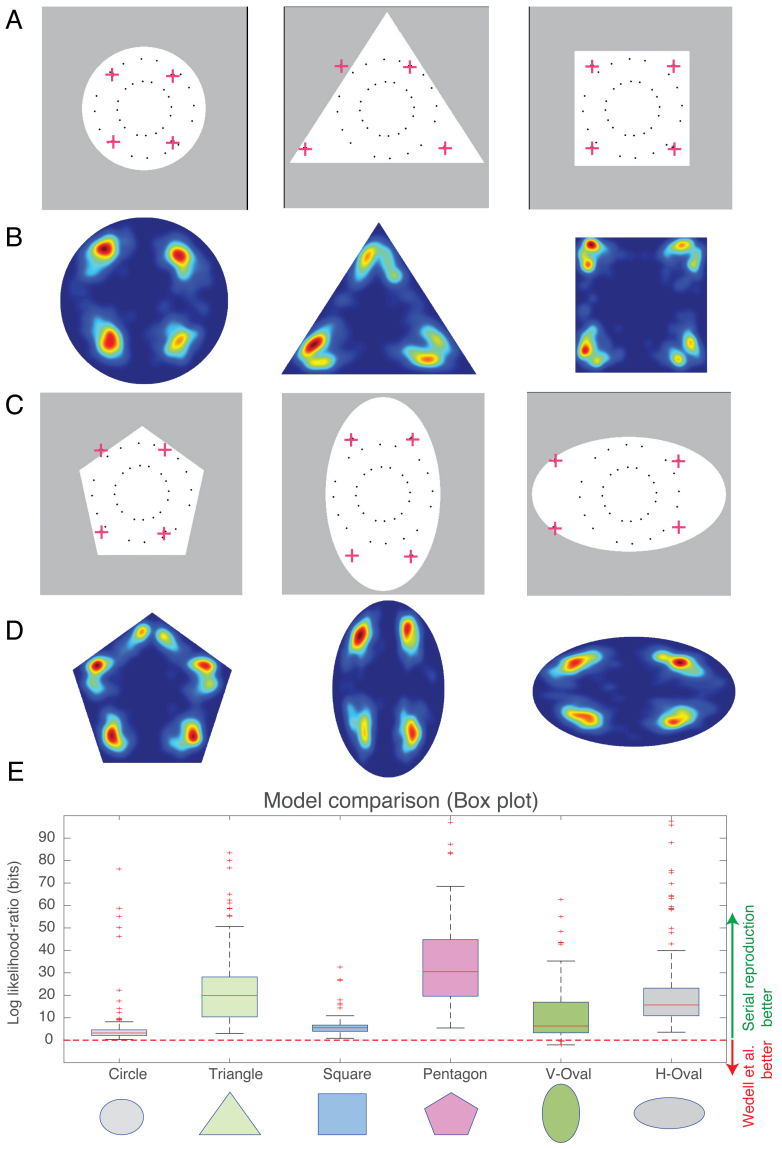

Figure 3.5: Kernel density estimates for the priors were estimated for all six shapes using the tenth iteration of the serial reproduction chain. A. and C. Original result by Wedell. B and D. Kernel density estimates with serial reproduction. E. Boxplots showing model comparisons. We computed the log likelihood difference for the two models as explained in the main text. In all cases the serial reproduction model was significantly better ($p < 0.01$ for all shapes except vertical oval ($p = 0.03$) resulting in positive log-likelihood ratios.

Model comparisons.

Using a combination of non-parametric kernel density estimation and serial reproduction lets us uncover intricacies in the prior for angular shapes (including bimodal peaks at the vertices) that paint a nuanced picture of human spatial memory priors. In addition, our approach enables us to obtain more than just point estimates of the locations of prototypes in spatial memory. Nevertheless, we provide a comparison between point estimates obtained from our method to those obtained from the model by Wedell et al. (2007), for each shape, using the same number of parameters. This model, called the Category Adjustment Model (CAM), describes the remembered position for a dot i (a response vector $\vec{R}_i$) as a weighted average of the actual location at which the dot was presented, which they refer to as the "fine-grain memory representation", and the weighted sum of the prototype locations, using the following equations:

$$\vec{R}_i = w\vec{S}_i + (1 - w)\sum_{j=1}^{4} v_{ij}\vec{P}_j \tag{3.1}$$

$$v_{ij} = \frac{e^{-c\|\vec{S}_i - \vec{P}_j\|}}{\sum_{k=1}^{4} e^{-c\|\vec{S}_i - \vec{P}_k\|}} \tag{3.2}$$

where $\vec{S}_i$ and $\vec{R}_i$ are vectors in $\mathbb{R}^2$ containing the x and y coordinates for each point i in the stimulus phase (iteration 0), and in the first response phase (iteration 1), respectively. The $\vec{P}_j$ terms correspond to the four prototype vector coordinates being estimated by the model, in addition to weights w that correspond to the relative strength of the veridical memory (as opposed to the strength of a prototype in the prior).The v_{ij} capture the relevance weight of each of the four j prototypes for each point i. In other words, the strength of the influence of prototype j for each point i. The parameter c corresponds to a "sensitivity" parameter that models the sharpness of the prototype boundaries.

We generated 100 split-half samples of the points for iteration 0 (inital seeds), iteration 1, and iteration 10. Next, for each sample, we obtained estimates of the prototype locations for four prototypes (the same number used by Wedell et al. (2007)) by running their model using the training half of iteration 0 and the same points in iteration 1. In order to ensure a fair comparison, we sampled four points under local maxima from the Kernel Density Estimate (KDE) fit to the same points in iteration 10. This gave us four prototype estimates from the Category Adjustment Model, and four points corresponding to local maxima in the KDE we fit to the points in the training half of iteration 10 (which can only be obtained from our paradigm), for each training split half. We evaluated the accuracy of these two sets of four prototype estimates by computing the sum of the negative-

log-likelihood values from a KDE that we fit to the remaining points in the testing half of iteration 10. Next, we computed the log likelihood difference for the two models, for each of the shapes. In all cases, the serial reproduction model performed significantly better ($p < 0.01$ for all shapes except the vertical oval ($p = 0.03$) resulting in positive log-likelihood differences. Boxplots showing all the results are displayed in Figure 3.5E.

Grouping of prototypes.

The apparent increase in peaks in the prior for more complex regular shapes afforded the opportunity to consider changes to the prior in the limit, as the shapes begin to approximate a circle. We computed the entropy of the obtained KDEs to quantify their complexity. Complexity increased with the number of vertices (going from a triangle to a heptadecagon, or seventeen-sided regular polygon). However, the prior for a icosihenagon (twenty-one sided regular polygon) begins to reveal the transformation of the corner peaks into one of the quadrant peaks. Entropy further decreases for the icosipentagon ($p < 0.001$), revealing a prior that appears nearly identical to the prior for a circle, and with similar entropy ($p = 0.68$) (see Figure 3.6).

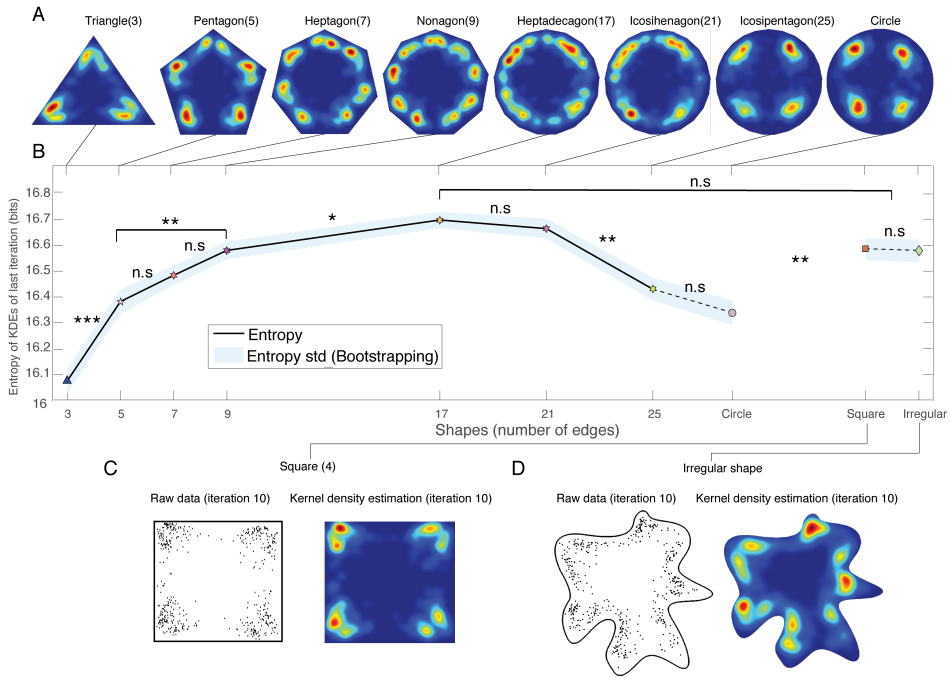

Figure 3.6: Grouping, and complexity of prior estimates. A. KDEs for regular polygons of increasing complexity. B. Entropy of the last iteration computed for all shapes. Entropy increases steadily with shape complexity (3 to 17 vertices). After the number of vertices exceeds 21, entropy stabilizes, and peaks start grouping toward the nearest quadrant center (as with the circle). We used the Bonferroni correction for multiple comparisons.

Discussion

In this first section, we present results from a preliminary foray into exploring spatial memory priors using serial reproduction: a process in which information being transmitted through successive participants leaves behind only a signature of the transformation process itself: the perceptual and reconstructive biases of those participants. This iterative process provides an effective tool for greatly amplifying biases in perception and memory.

We used a serial reproduction paradigm in the context of a spatial memory task. KDEs of the dots' final positions revealed detailed structure in priors over location. We found that the priors for circles and ovals show peaks at the center of each of their four quadrants, but also discovered that angular shapes show bi-modal peaks at the vertices in the prior. The modes appear on either side of each vertex, and do not seem to be a simple function of the angle at each vertex, since they are not rotationally invariant in all cases. We provided quantitative comparisons between the perfor-

mance of a parametric model, and point estimates derived from the KDEs we obtained following ten iterations of the chain. These comparisons demonstrated that our estimates were significantly better than those obtained from the parametric model (we used the same number of parameters—four prototype estimates, even though our method yields kernel density estimates that clearly reveal more than four in some cases). In future work, we intend to determine if priors differ across individuals, by repeating the experiments so that each participant completes a subset of chains in their entirety (within-subject design). While some studies show differences between within and between-subject designs (Claidière et al., 2014), most studies showed high agreement between these versions (Jacoby & McDermott, 2017; Xu & Griffiths, 2010). We provided quantitative comparisons between the performance of a parametric model, and point estimates derived from the KDEs we obtained following ten iterations of the chain. These comparisons demonstrated that our estimates were significantly better than those obtained from the parametric model. For comparison, we used the same number of parameters—four prototype estimates, even though our method yields kernel density estimates that clearly reveal more than four in some cases. However, while determining whether individual differences exist in people's spatial memory priors using a within-subject design is a worthwhile and important direction for future experiments, we are inclined to believe that our results would not change using a within-subject-design.

Finally, we demonstrated a novel application of our method to more complex shapes. We show that while priors tend to have peaks near the vertices in the angular shapes, as more vertices are added, there appears to be a limit on the complexity that determines how the prior is represented, resulting in a grouping of the peaks into quadrants for regular polygons approximating a circle.

Our results suggest that our approach may provide an opportunity to uncover complex priors for a wide range of perceptual phenomena that would otherwise elude traditional experimental approaches, and parametric models. We plan to use it to measure memory biases when there is more than one point to be remembered (Lew & Vul, 2015), and to probe for structured priors in memory for local orientation (chapter 3 presents some preliminary findings on memory priors for the three-dimensional orientations of objects) (Wei & Stocker, 2016). This section ends with some additional results from experiments in which we manipulated some simple features of the shapes (bounding frames of the circle, and rotations) in order to understand the enigmatic patterns we observed (and which had been confirmed in prior work (Huttenlocher et al., 1991)) for the circular and elliptical shapes. We also show some results for an exposure-time experiment: we demonstrate that varying the exposure time seems to shift the moment at which modes in the prior seem to multiply as vertices are added. This result opens the door for future work capitalizing on the increased resolution we can obtain in order to probe what visual features are used, and which are discarded during a vi-

sual memory task when the sensory information is further degraded by shorter exposure times. The next section presents results using the same shapes, but when the spatial domain in which the dot was initialized wasn't restricted to the boundaries of the shapes. The following section presents spatial memory priors that were revealed using the same task (dot memory task) but with natural complex images, which gave us an opportunity to explore the effects of higher-order visual features and semantic content on spatial memory biases. We also show results from an analysis in which we extracted several visual features in order to determine the extent to which they were predictive of the visual memory priors.

Effects of manipulating a bounding frame

We conducted additional experiments to ascertain the effects of manipulating some simple features of the geometric shapes on the prior. In particular we investigated the effect of adding a bounding frame on the circle shape (a square and triangular bounding frame), as well as the effect of rotating the bounding frame (see Figure 3.7). The original results for the circle and elliptical shapes present an enigma, and do not lend themselves to an easy interpretation as the results with the angular shapes do. While the vertices of the angular shapes seem like clear basins of attraction (although why they would yield the bi-modal patterns we observed is still open to interpretation), the presence of modes inside the quadrant centers of the circular shapes remains a mystery even though they were known far before we measured them here.

One explanation for why we see the presence of four modes inside the quadrants of the circle has to do with people's frame of reference when they perform the task. Perhaps that implicitly they are using canonical Euclidean axes centered in the circle (horizontal and vertical axes), or they are using the corners of the screens of the devices they completed the task with. While we took measures to avoid people from deliberately tracking the point locations relative to some fixed frame of reference like their screens, by jittering the trial displays by random offsets, the results seem to suggest that people use some kind of horizontal and vertical frame of reference. We introduced a bounding frame to the circular region (either an equilateral triangle, or a square) in order to test the effects of having different bounding shapes (or frames of reference) for the circle. In addition, we rotated the bounding frame to measure whether we would measure corresponding rotations in the prior. Our prediction for the circle bound by a square was: if people do implicitly use some kind of vertical/horizontal frame of reference when doing the task (such as a bounding square, or canonical Euclidean basis centered in the circle's origin), that the bounding square, if rotated would either serve as a new frame of reference, or it would cause Stroop-like interference. In both cases, we pre-

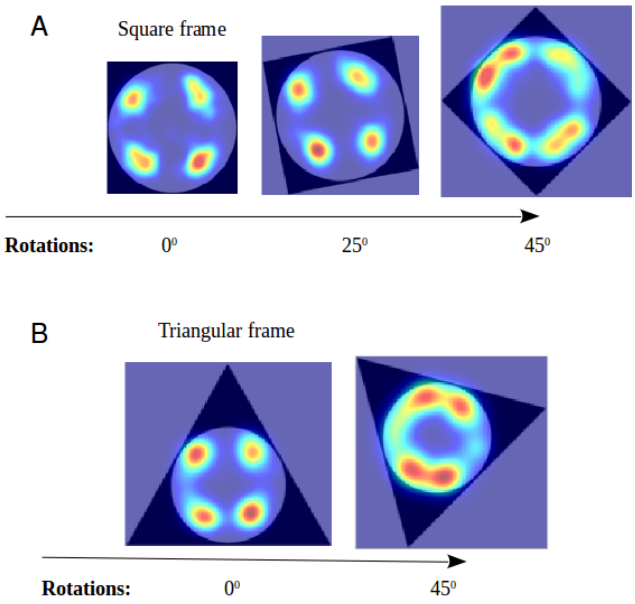

Figure 3.7: Effect of adding an angular bounding shape around the circle, and of rotating the shape on the spatial memory prior, for A: a square bounding frame and B: a triangular bounding frame. Note that the introduction of the bounding frames has no effect on the prior for the circle in both cases. A small rotation of the square bounding frame results in an identical rotation of the prior. However, a 45-degree rotation yields eight modes in the circle prior, clustered in four pairs. A 45-degree rotation of the triangular frame yields a different pattern altogether.

dicted that we would observe a corresponding rotation in the prior (that the four modes would rotate by the same angle as the bounding frame). For the triangular frame, we predicted that it would result in three modes instead of four, and that rotating it would also result in a corresponding rotation of the prior itself. The results are shown in Figure 3.7. They reveal an unexpected interaction between the effects of the bounding frame shape (square or triangle) when combined with the rotation. Surprisingly, the presence of a square and triangular bounding frame with no rotation does not appear to alter the prior—as before, we see four distinct modes that emerge following 20 iterations of the task. This result for the square is consistent with our prediction (although we expected that the presence of vertices might cause some distortions despite the fact that the points were restricted to being within the circular boundary). The result for the triangle is surprising: the modes appear unchanged, suggesting that altering the number of vertices in the bounding frame does not have an effect on this pattern. Introducing as small rotation of 25 degrees yields a corresponding rotation of the modes for the square (we did not do the same experiment with the triangle). This is consistent with our prediction, and suggests that introducing the bounding shape changes the frame of reference used by our participants, as if they were adopting it as a new basis for estimating the dot locations. However, a full 45 degree rotation yields unexpected changes: For the circle with the bounding square, the four modes appear to multiply into eight, and for the circle with the bounding triangle, they appear to multiple into six. These results suggest that manipulating a frame of reference does disrupt the circle pattern, but raises more questions than it answers —a complicated interaction between rotational and geometric effects of the bounding frame. Future work will attempt to tease out theses effects with additional manipulations, such as larger bounding frames, where the edges are not tangent to the circle, as well as effects of varying the exposure time during the retention phase of the task.

Effects of varying the exposure time

We determined that given specific parameter settings, there is a limit on the complexity that determines how the prior is represented (adding vertices to regular polygons as they approach a circle in the limit results in a grouping of the modes at the vertices into the four quadrant centers of the circle). We conducted an additional experiment in order to test whether varying the exposure time during the retention phase of the experiment has a similar effect, see Figure 3.8.

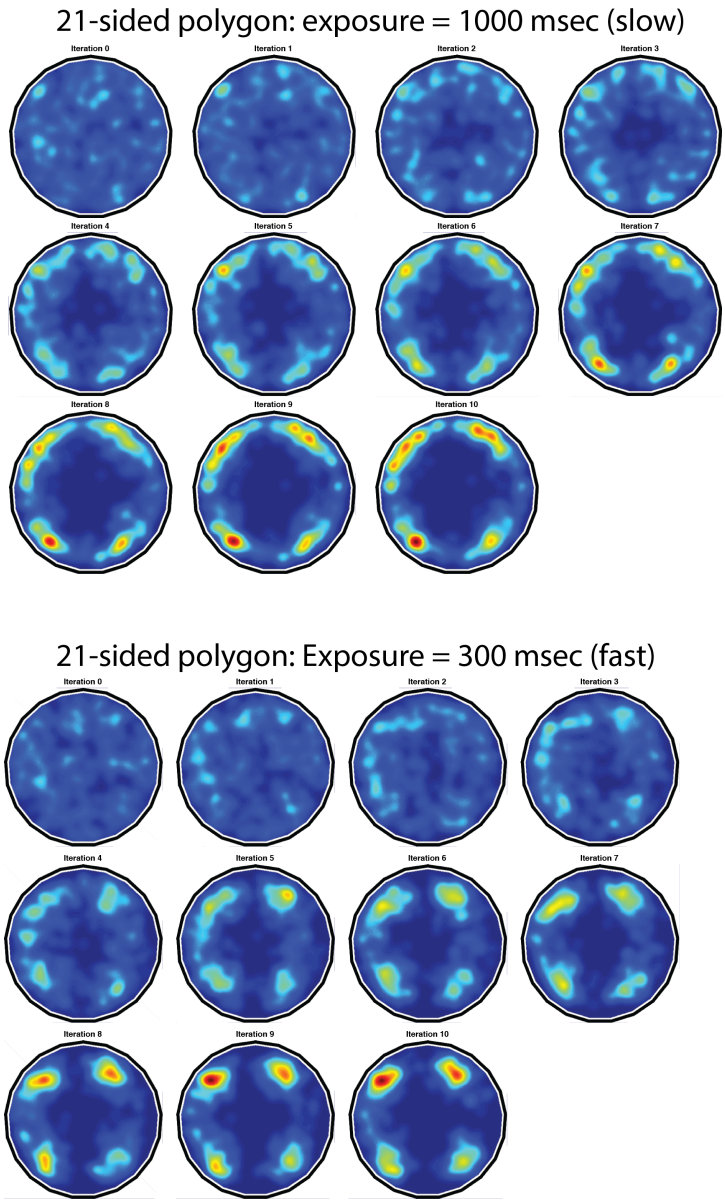

Figure 3.8: Effect of manipulating the exposure time during the retention phase of the spatial memory task, for a 21-sided regular polygon. Longer exposure times yield higher complexity in the prior, and shorter exposure times yield a prior that is similar to the one for the circle. This result suggests that the limit on the complexity with which the prior can be represented depends not only on the physical characteristics of the shape, but also on the duration of the retention phase.

Scale invariance of the prior for the circle

As a means to better understand the pattern in the prior we confirmed with the circle, we wondered whether the prior would change as a function of the scale of the image. Perhaps that reducing the visual angle within which participants had to estimate the remembered location of the dot would improve their performance, if performance is contingent on discrimination accuracy. Regions farthest from the boundaries of the circle (close to the center) would be relatively closer to the boundary for smaller images. Figure 3.9 shows the results of conducting the same experiment that yielded the original circle results, but when the circle was scaled to a fraction of its original size (the sizes of the scaled down images were roughly the sizes that appear on this page). Remarkably, this manipulation revealed a total scale invariance in the structure of the spatial memory prior.

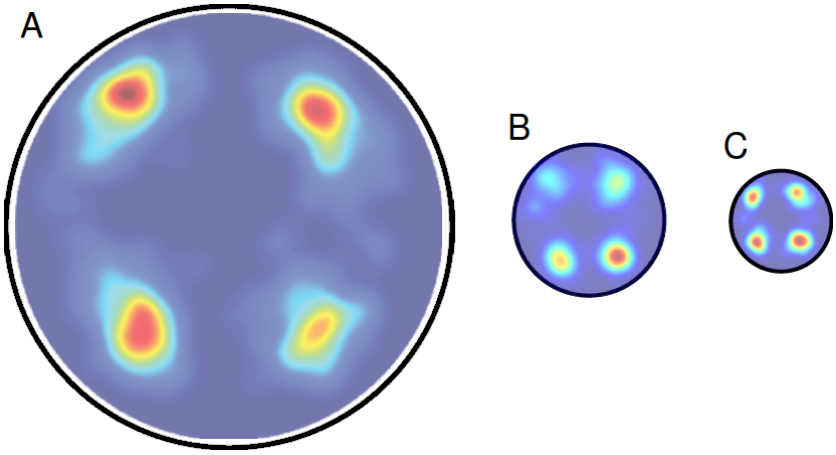

Figure 3.9: Scale invariance in the spatial memory prior for the remembered position of a dot shown inside a circle. A. Our original result for the circle: overlay of the kernel density estimate on top of the circle image. B and C show the results for the smaller circles. The prior appears to be invariant to the scale of the circle image.

3.2 Priors for Shapes Without Spatial Domain Restrictions

We repeated the spatial memory experiments with the same geometric shapes (circle, triangle, square, pentagon, vertical and horizontal ellipses), but without restricting the spatial domain of the initial seed locations. Since dot locations were only presented within the shape boundaries in the original series of experiments, we could not determine what the prior would look like on either side of the shape boundaries. In addition, we wanted to loosen this restriction in order for the experiments to be consistent to those we conducted with natural images (which are discussed in the next section). The results are largely similar to the original shape results, although there are notable differences:

modes appear near the vertices on the outer edges of the shape boundaries in addition to internal edges.

3.2.1 Methods

Participants

As before, participants were recruited online using Amazon Mechanical Turk. All gave informed consent. The experimental protocol was approved by The Committee for the Protection of Human Subjects (CPHS) at the University of California, Berkeley. Each experiment required approximately 70-100 participants.

Stimuli

All images were approximately 400×400 pixels in size. Each shape subtended a shaded gray region over a white background. The sizes and colors of the backgrounds for the images were intended to ensure that the images would be clearly visible in any standard browser window (unlikely to become occluded), and such that the white boundaries of the images would be invisible.

Procedure

We carried out a series of serial reproduction experiments. Participants were presented with timed displays (a shape outline with a dot initialized somewhere within the boundaries of the shape), and were instructed to reproduce the exact location of the dot initialized over the image (and not necessarily within the shape boundaries). Once complete, their response was sent to another worker (again, as a timed display), who was instructed to reconstruct this display from memory, and so on. A total of twenty iterations were completed for each chain. See Figure 3.2 for a schematic diagram of the serial reproduction procedure.

Practice trials. Participants completed ten practice trials in order to become familiar with the user interface. During these practice trials, they were presented with an image containing a circular gray region, with a dot initialized somewhere over it. This display was presented for 4000 ms, followed by a blank screen lasting 1000 ms. Next, the gray circular disc was presented without the dot and remained on the screen until the participant positioned the dot in the location that they remembered. As soon as the participant clicked, the dot appeared under the mouse cursor. Participants could change their responses as many times as they needed before clicking the "next" button. Once done, they pressed the "next" button to proceed to the next trial.

Experimental trials. Following the ten practice trials with the circle, there were ninety-five experimental trials with exactly one of the shapes. In Experiment 1, the shape could be a uniform gray circular disc, gray shaded triangle, square, vertical oval, horizontal oval, or pentagon. For each of the 95 experimental trials, the presentation time was reduced to 1000 ms. As with the practice trials, the position of the shape on the screen was randomized somewhere inside a larger canvas in order to control for participants resorting to tracking the position of the dot by trivially marking its absolute position on their computer screens. In addition, participants were given trial-by-trial feedback regarding their accuracy. If their responses were within eight percent of the width and height subtended by the shape on the screen, they were told that their response was accurate (a message in green font: "This was accurate"), and received a small monetary bonus. If not, they received no bonus beyond the basic payment for the HIT, in addition to any bonuses accrued from the previous trials, and were presented with a red message ("this was not accurate"). These trials were discarded from the experiment. Participants could not provide multiple responses within a chain.

3.2.2 Results

Results including scatter-plots are presented for the circle and the triangle in Figure 3.10 and Figure 3.11. They reveal that loosening the spatial domain restrictions on the original random positions of the points results in peaks in the prior with a hierarchical structure: for the circle, four locations straddling the border of the gray disc boundaries, near the centers of the quadrants emerge, but they are divided around the boundary, into bi-modal groups.

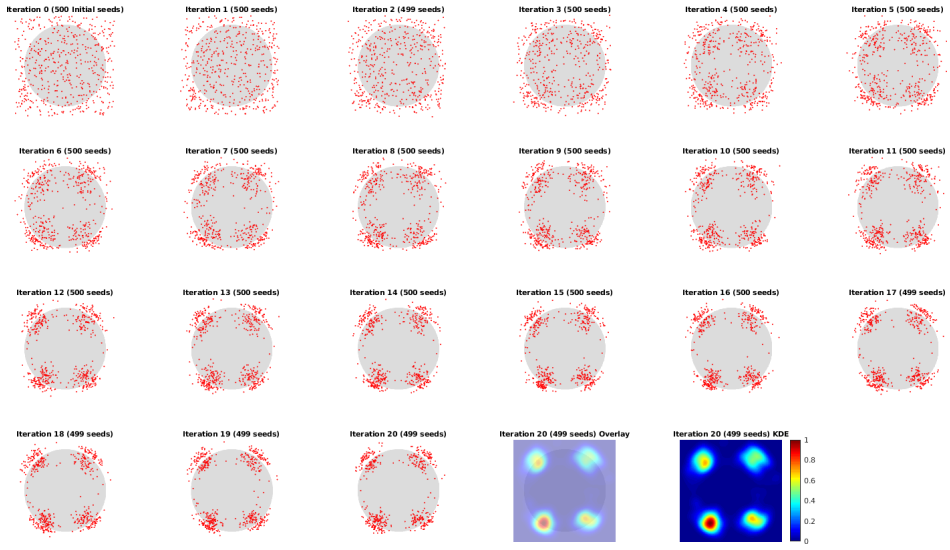

Figure 3.10: Serial reproduction of 500 dots, for twenty iterations. The top left scatter plot shows the positions of the original seeds (sampled from a uniform distribution). The remaining subplots show scatter plots of the results of the serial reproduction chain for iterations 1-20. The subplot of the tenth iteration, in the bottom right, also shows an overlay of the circle image with the kernel density estimate, followed by just the KDE.

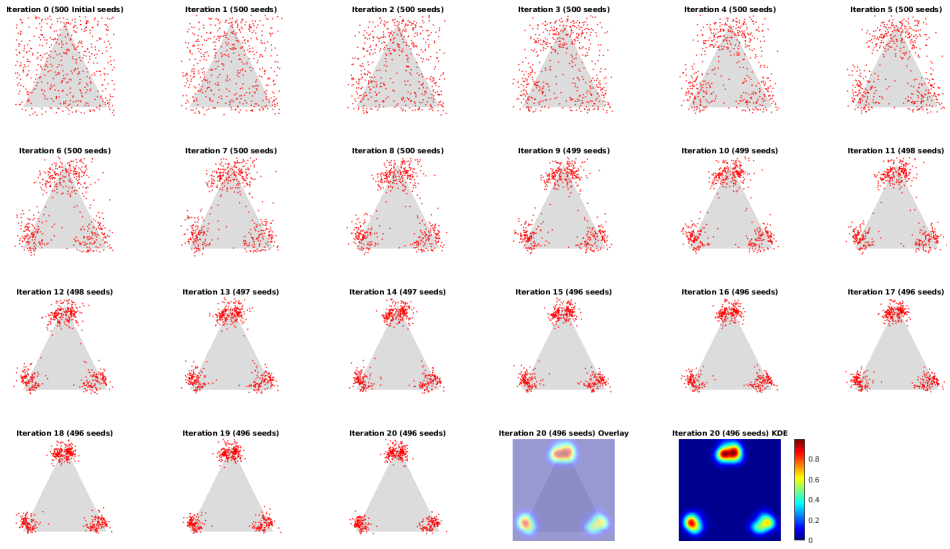

Figure 3.11: Serial reproduction of 500 dots, for twenty iterations. The top left scatter plot shows the positions of the original seeds (sampled from a uniform distribution). The remaining subplots show scatter plots of the results of the serial reproduction chain for iterations 1-20. The subplot of the tenth iteration, in the bottom right, also shows an overlay of the triangle image with the kernel density estimate, followed by just the KDE.

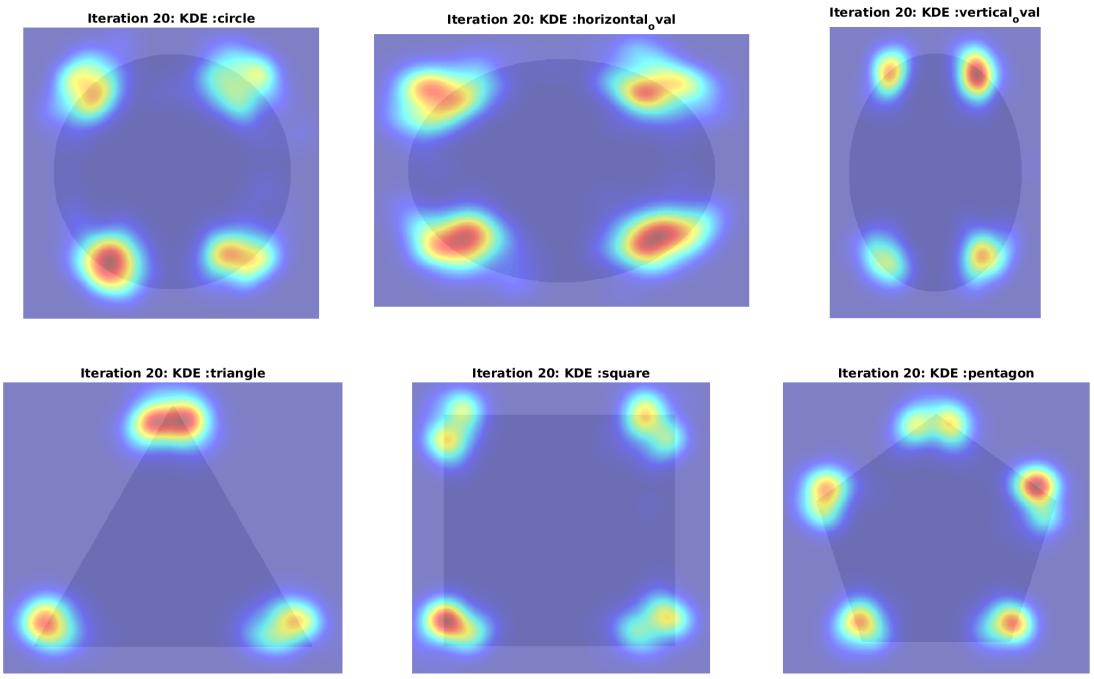

Figure 3.12: All the shaded shapes, with an overlay of the kernel density estimates obtained for each from the distributions of points at the 20th and final iteration of the serial reproduction chains. Note: in the actual experiments, the vertical and horizontal ellipses were the same dimensions. The vertical ellipse is only shrunk here in order to fit into the figure.

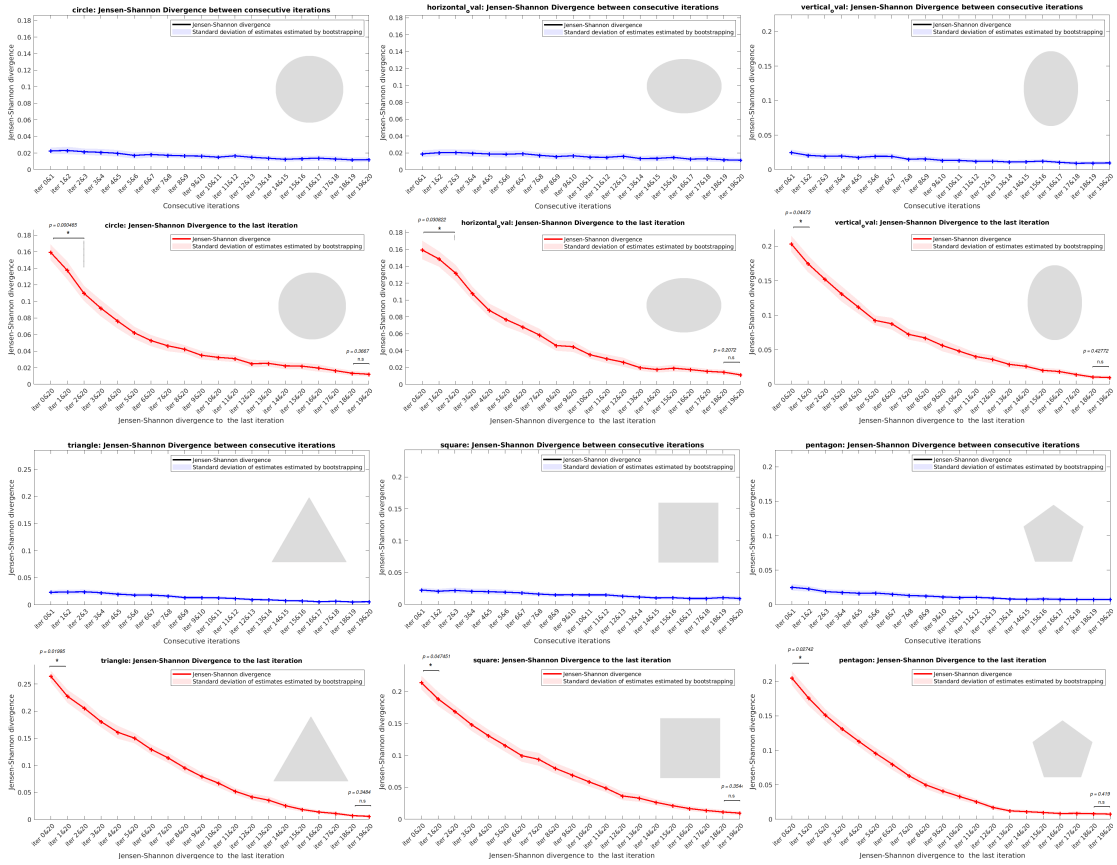

Figure 3.13: Results of convergence analyses using the Jensen-Shannon Divergence (JSD) metric, for all the shapes. For each shape, we show the JSD between consecutive iterations, as well as the JSD between each iteration and the last iteration. In all cases, we show that there are no significant differences in JSD between consecutive iterations, as expected. Most importantly, we show that when comparing JSD between each of the twenty iterations and the final iteration, there is a clear and significant decrease in JSD between iteration 1 and iteration 3 for all shapes, but convergence occurs at the end of the chains, since there are no significant differences in JSD between the 19th and 20th iterations. In some cases, convergence appears to occur even earlier. These results show that we used a sufficient number of iterations, and that the priors we obtained appear to be good estimates of the stationary prior distributions for all the shapes.)

Estimating the priors with serial reproduction chains requires that we identify the correct number of iterations that are needed to achieve convergence. In other words, at what iteration does the estimate of the prior stop changing significantly? This is the point at which we can say that we have finally converged on a stationary estimate of the prior distribution. A method for testing the convergence of the chains is using the Jensen-Shannon Divergence (JSD) to compare the distribution at the *ith*

iteration to the original (uniform) distribution. Figure 3.13 shows the results of the JSD analysis for all the shapes. For all the estimates we used 50 bootstrap samples with replacement of the data. The subplots in the first two rows show the results for the curved shapes (circle, horizontal and vertical ellipses). The blue curves show the JSD between consecutive iterations, and the red curves show the JSD between each iteration and the final (20th) iteration.

3.3 Spatial Memory Priors for Natural images

Prior work investigated spatial memory biases using the same task (remembering the exact position of a dot) but using natural images instead of simple shapes (Holden et al., 2013). These reports presented evidence that people tend to remember the location of the dots towards the centers of mass of segmented regions in the images (Holden et al., 2013). We repeated our experiments using natural images in order to probe for effects of natural image features on spatial memory biases, and to test whether spatial memory is indeed biased towards centers of mass in segmented regions. Given the sensitivity of our approach, and the fact that we could reveal hitherto unobserved subtleties in the context of geometric shapes, we were hopeful that doing the task with natural images could provide a strong test of whether memory is actually biased towards the centers of mass of segmented regions. We also expected that it might reveal systematic and nuanced effects that other approaches failed to measure in past work. We selected images from the PASCAL dataset that corresponded loosely to three image categories: faces, scenes, and objects. We used the PASCAL dataset because it includes human segmentations (from which we could compute centers of mass), and eye-tracking fixation data. Using 250 chains with 20 iterations, our results reveal strikingly complex patterns in the prior.

Figure 3.14 shows the results using a natural image of a face. The 21 scatter plots show the positions of the red dot at iteration 0 (the original random seed positions), followed by their positions at iterations 1 through 20. The bottom right subplots show an overlay of the image with the kernel density estimate obtained from the scatter plot at the final 20th iteration, and the kernel density estimate without the image. 500 unique chains with 20 iterations of the process were used, and they reveal clear memory biases that appear to be influenced by higher order semantic image features —the dots appear pulled towards they eyes, nose and mouth of the face.

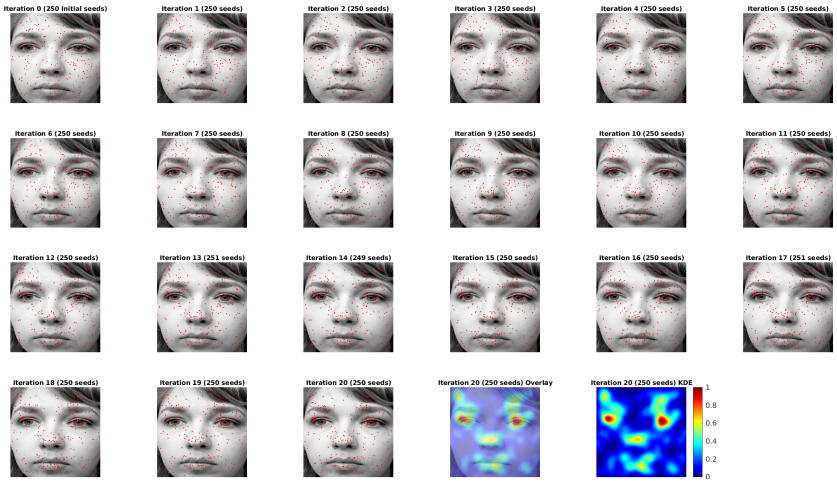

Figure 3.14: Serial reproduction of 500 dots presented over a face, for twenty generations (iterations) of the process. The top left scatter plot shows the positions of the original seeds (sampled from a uniform distribution). The remaining subplots show scatter plots of the results of the serial reproduction chain for iterations 1-20. The subplots in the bottom right also show the kernel density estimate (KDE), and an overlay of the image with the KDE. Note that from one iteration to the next, points that were originally scattered uniformally converge towards salient features of the face: the eyes, nose, and mouth

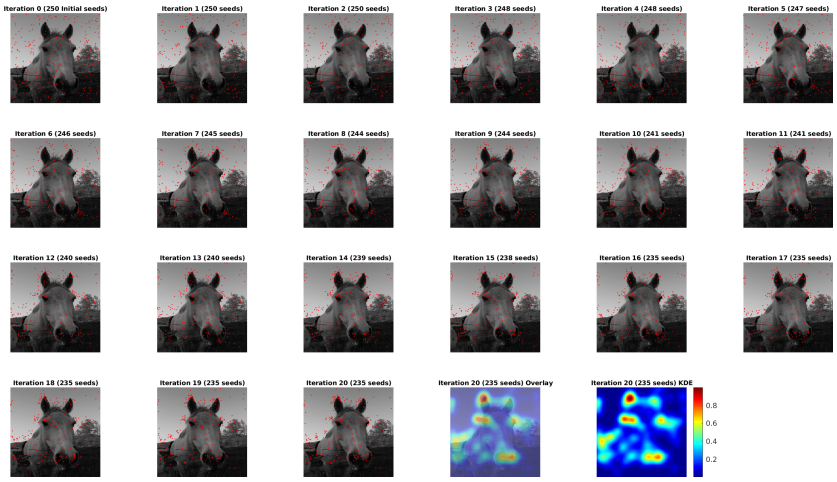

Figure 3.15: Serial reproduction of 500 dots presented over a horse's face, for twenty generations (iterations) of the process. The top left scatter plot shows the positions of the original seeds (sampled from a uniform distribution). The remaining subplots show scatter plots of the results of the serial reproduction chain for iterations 1-20. The subplots in the bottom right also show the kernel density estimate (KDE), and an overlay of the image with the KDE. Note that from one iteration to the next, points that were originally scattered uniformally also converge towards salient features: the eyes, ears and snout

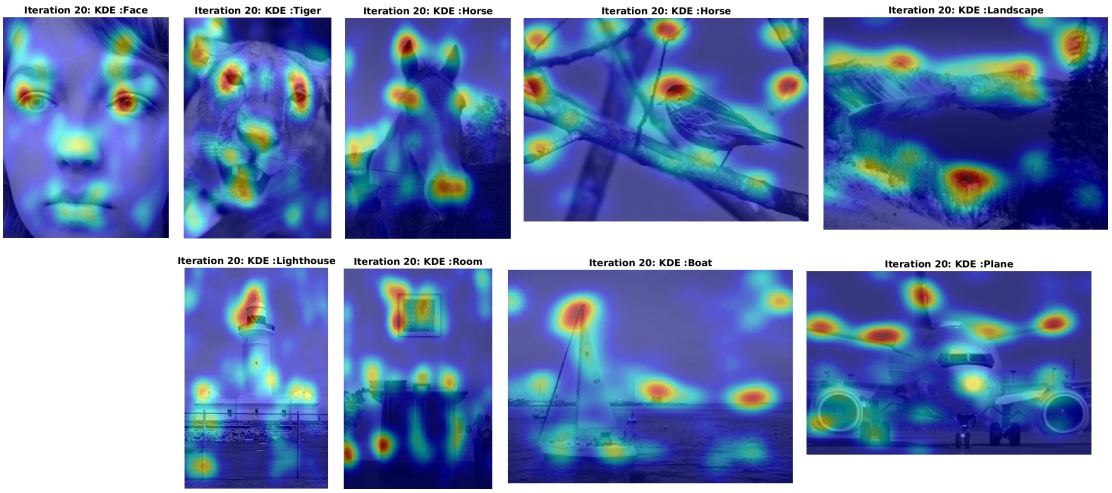

Figure 3.16: Natural images overlayed with kernel density estimates of the spatial memory prior estimated from the final iteration of the serial reproduction chains. Images of faces show similar patterns: memory for the dot location is biased towards the eyes, nose, and mouth. Biases in other images, while harder to interpret, show clear regularities and structure.

Model Comparisons: Predicting the Spatial Memory KDE using Image Features

The results obtained for the small set of natural images revealed clear, replicable, and qualitatively interpretable results. In the case of face images, we observed clear spatial memory biases towards similar areas (such as the eyes). A natural question is: What image features are responsible for these biases? and can we predict the kernel density estimates we obtained on the basis of certain image features? Looking at the results for the face images naturally brings to mind eye-gaze fixation patterns, but comparisons between these results and kernel density estimates of eye-gaze fixation patterns obtained for the same images in a free viewing task (available from the PASCAL dataset), shown in Figure 3.17, indicate that they re not very predictive of spatial memory biases.

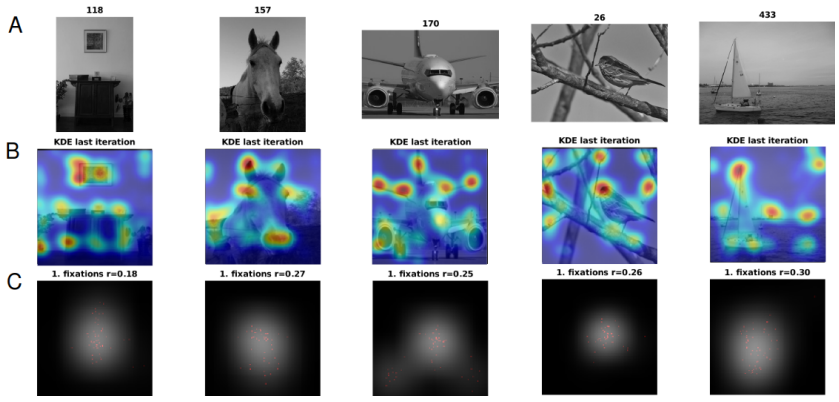

Figure 3.17: Correlation kernel density estimates of eye-gaze fixations obtained from the PASCAL dataset. The relatively low correlations indicate that eye-gaze fixation patterns are not very predictive of spatial memory biases. A shows the original images used in our task, and for which eye-tracking data on a free viewing task were available. B shows an overlay of our spatial memory kernel density estimates over the images. C shows a scatter plot of fixations, and Gaussian kernel density estimate.

We generated a number of rudimentary image features, including Canny edges, Harris corners, Shi-Tomasi corners, and entropy maps. The entropy maps were generated by obtaining distributions of normalized cross-correlations between small image patches of an image, and all other patches in the image using a stride of 1 pixel in both dimensions. For each patch, we computed the entropy of the distribution of correlation values. This gave us a measure of a patch's "confusability," in other words, a measure of how unique it is relative to other image patches. Figure 3.19 illustrates how the entropy maps were computed, and shows an example (entropy map for the face image), and explains the layout of the prediction evaluations that are shown in subsequent figures. Each column corresponds to an image, and each row corresponds to an image feature map. In all cases, correlations between the feature map and the final spatial memory KDE are shown.

For each feature detector, we performed a detailed grid search of all the parameter settings within the ranges that are usually specified for these algorithms given the dimensions of the images we used. We also introduced an additional smoothing parameter (the standard deviation of a Gaussian kernel that was convolved with the final feature map). The goal was to find the maximally predictive parameters for a given feature detector to see how well it could predict our spatial memory kernel density estimates. For a given image, we started with the Canny edge detector. For each image, we extracted 2,400 edge maps, one for a unique set of parameter settings. The Canny edge detector has four parameters: the first and second threshold for the hysteresis procedure, the aperture size of the Sobel filter (which computes the gradient in the image), and the norm of the gradient magnitude (either an L2 norm or an L1 norm). Increments of 10 within a range of 0 to 190 were used for the

first and second parameters, three aperture sizes (3, 5 and 7 pixels) were used for the third parameter, and a Boolean setting indicating which norm to use for computing the magnitude was used for the fourth parameter. This yielded 2,400 total Canny edge feature maps for each image. For the evaluation, an additional smoothing parameter was used on the final maps. A total of 280 feature maps for the Harris corner detector were used, which corresponded to all the unique settings along a grid search for its three parameters (the size of the neighborhood considered for corner detection, called the "block size", the aperture size for the Sobel derivative operator, or "k-size", and a free parameter used in the Harris detector equations). For our evaluation, we varied the the block size between 2 and 9 pixels, the Sobel filter aperture size between 1 and 7 pixels, and the final free parameter between 0.01 and 0.13 at increments of 0.02 (The appropriate range for this parameter setting). As with the Canny edge detector feature maps, we introduced an additional smoothing parameter before the final evaluations of the Harris corner maps. For the Shi-Tomasi corner detector, which is similar to the Harris corner detector, a total of 1,944 maps were generated, which were all the unique parameter settings for three parameters: the number of corners to be found (which varied between 10 and 240 at increments of 10, and two additional parameters that varied between 0.1 and 0.9, with increments of 0.1). Finally, as a baseline feature map, we used the inverted grayscale of the original image (We inverted it in order to have black pixels yield values of 1, and white areas yield values of 0).

Figure 3.20 shows both the maximally predictive feature maps, for all the features (these are denoted by "Method" numbers—"Method 1" corresponds to the Canny edge detector, and so on). Above each of the feature maps is the maximum Pearson correlation r value that could be achieved following the full grid search of all parameter settings for that feature. These r values are also displayed in the barplots in the bottom row, for each image. The corner and edge detectors out-performed the entropy maps in nearly all cases, although the correlations were not particularly high in general. As expected, simply using inverted grayscale images yielded the lowest correlation values. Figure 3.21 and Figure 3.22 show the same results, for the PASCAL images, and the results are the same overall. Because the PASCAL images came with segmentation maps, we were able to obtain spatial memory priors from experiments using these maps instead of the original images, and we show the results of using these to predict the oringinal image KDEs. Using the KDEs from the segmentation map experiments, predictions outperformed all the maps generated from the automated feature detectors we used. While this comes with important caveats—such as the fact that these feature maps came from human experiments, and were not generated directly from the images, they suggest that spatial memory priors are biased towards "semantic corners," or higher-level visual features that correspond to the semantic content of the images rather than the low-level pixel intensities. Figure 3.22 includes, in the sixth row, prediction from using the centers of mass of the

segmented blocks, followed by smoothing.

The fact that the spatial memory priors we estimated from the experiments that employed the segmentation maps are so similar to those that we estimated from the original images is remarkable, and indicates that lower-level visual information in the images, such as corners, gradients, and edges, are not the real determinants of the biases in spatial memory, at least not until they are interpreted as semantically meaningful features (an eye, an ear, the edge of a wing, or the tip of a building, etc). For ease of comparison, Figure 3.18 shows the results for the natural images and the segmentation maps side-by-side. In the discussion that follows, we sketch out future approaches for predicting our priors that would involve using contemporary computer vision algorithms, such as pre-trained semantic segmentation neural networks, as well as variants on convolutional neural networks to see if we can extract the "semantic corners" that are common to the images and the segmentation maps. One idea involves designing a convolutional neural network trained with constraints on how much of an image may be used in training (a Lagrangian constraint on the budget of pixel-wise activations, which the network would need to learn how to distribute optimally for different image categories).

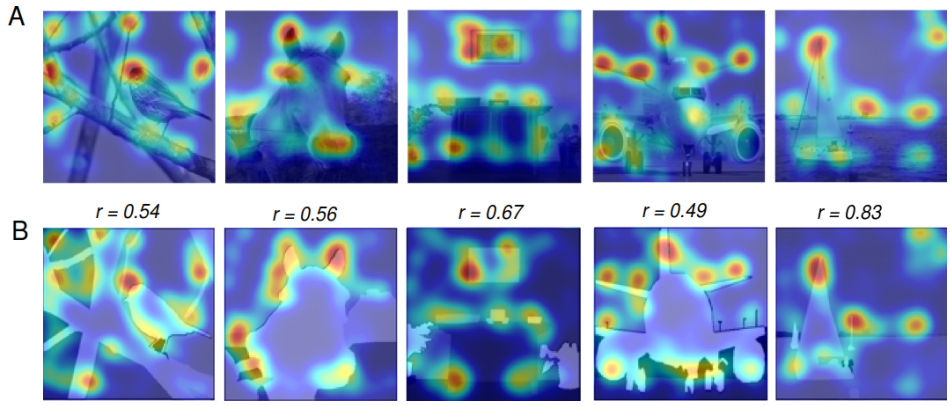

Figure 3.18: Spatial memory priors and corresponding images. Panel A shows the natural images used in the original experiments, and Panel B shows the results using the segmentation maps in the spatial memory experiments in lieu of the natural images. The Pearson correlations between the corresponding spatial memory prior KDEs are also shown.

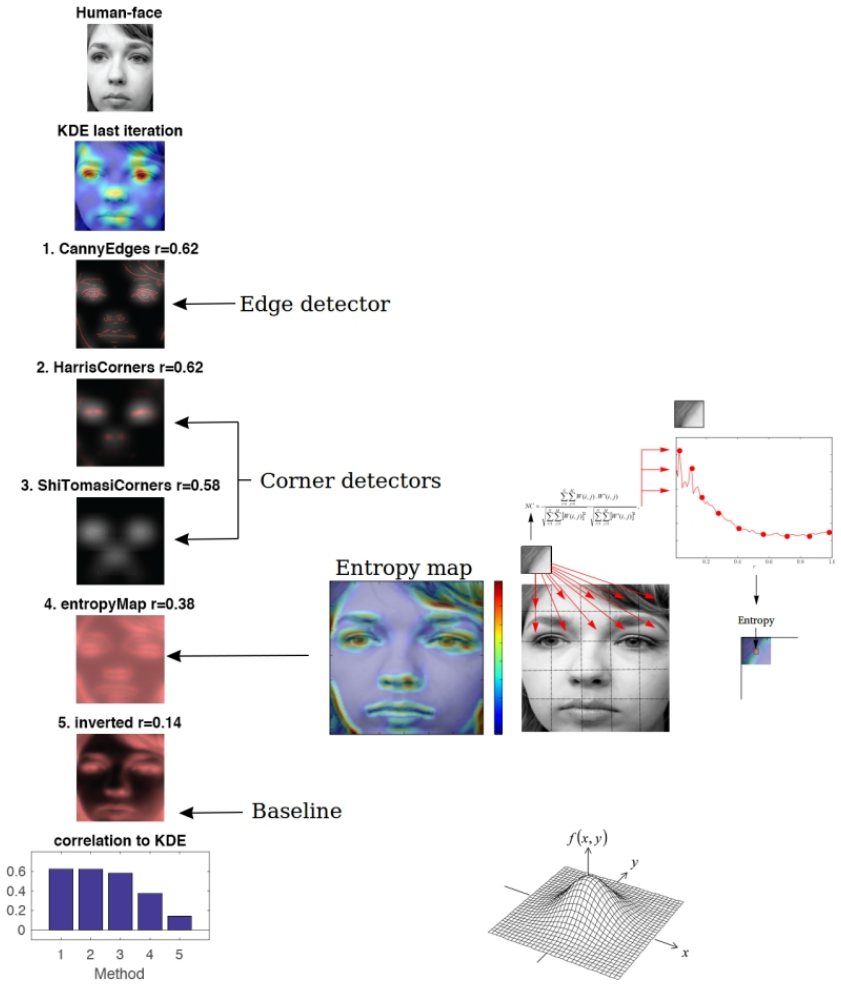

Figure 3.19: Illustration of the various features that were used, including an illustration of how the "entropy maps" were computed. We performed a grid search over all the parameters of the edge and corner feature detectors (with an additional smoothing parameter) and selected the parameter settings that yielded the highest correlation between the resulting feature maps and the final iteration kernel density estimate of the spatial memory prior. The "entropy map" was computed by comparing a small $8x8$ image patch to all other patches in the image (using a stride of 1 pixel in both dimensions), using normalized cross-correlation (NCC) (akin to template matching). For each image patch, this yielded a distribution of values. We computed the entropy of these distributions, which gave us a measure of the patches "confusability" to other image patches. This illustration shows the face image with the entropy map displayed over it.

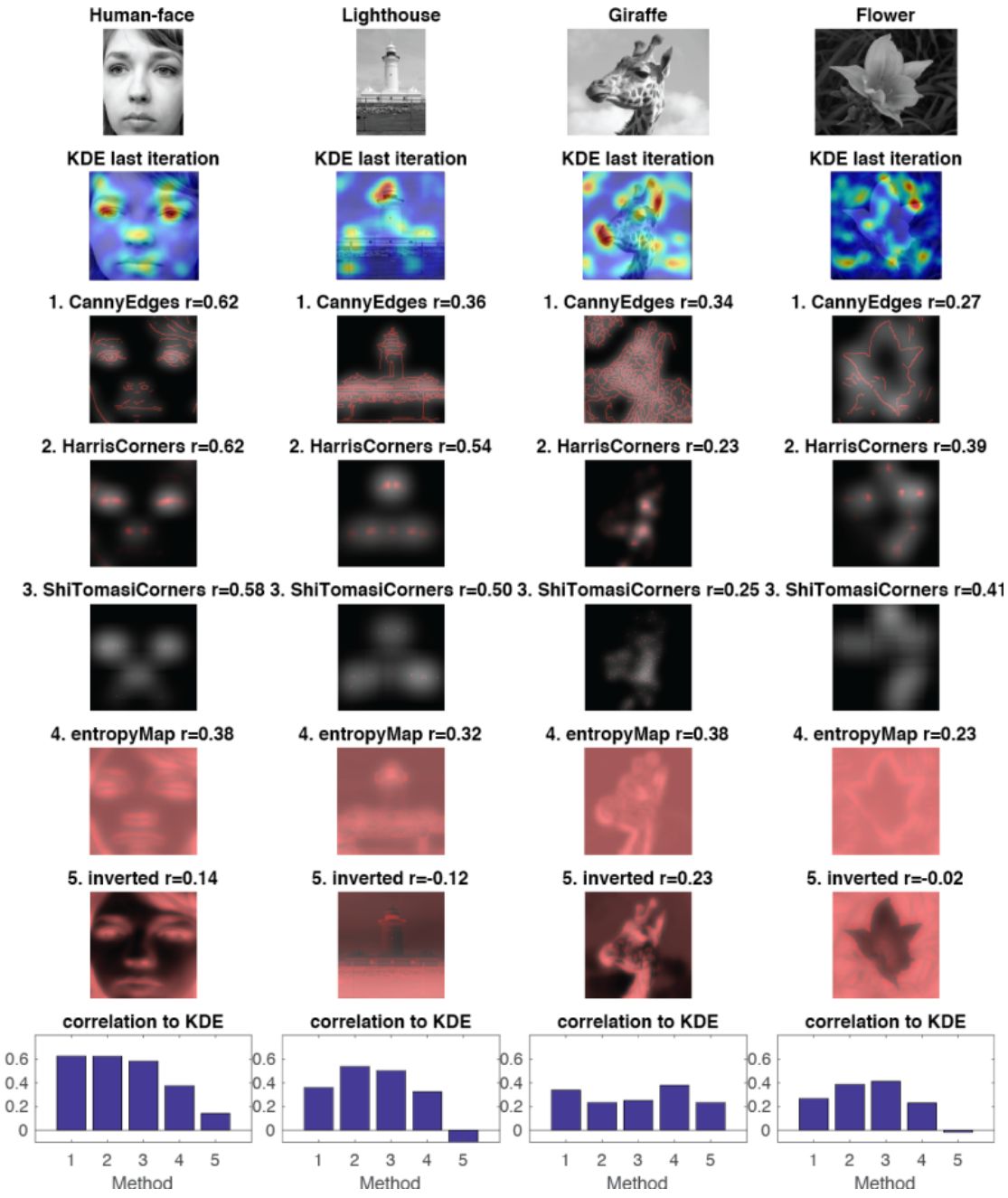

Figure 3.20: Model Comparisons for face, lighthouse, giraffe, and flower images. The first row shows the original images, the second row shows images with kernel density estimates of the prior obtained form the final iteration of the serial reproduction chains. The third, fourth, and fifth rows show the maximally predictive results (trying to predict the final iteration KDEs) following a grid search of all the parameters of the Canny edge detector, Harris and Shi-Tomasi corner detectors, followed by an additional smoothing (convolution with a Gaussian kernel). The sixth row shows the optimally predictive results of using an entropy map (described in the text), followed by an additional smoothing. The seventh row is intended as a baseline (lower-bound) prediction—and shows the maximum prediction that can be achieved using the inverse of a gray-scale of the original image to predict the spatial memory prior. The bottom row shows barplots of the correlations between each of the feature maps shown in each row (denoted by the "method" number, and the KDEs). These correlations are also displayed above each of the subplots in each row, over each image.

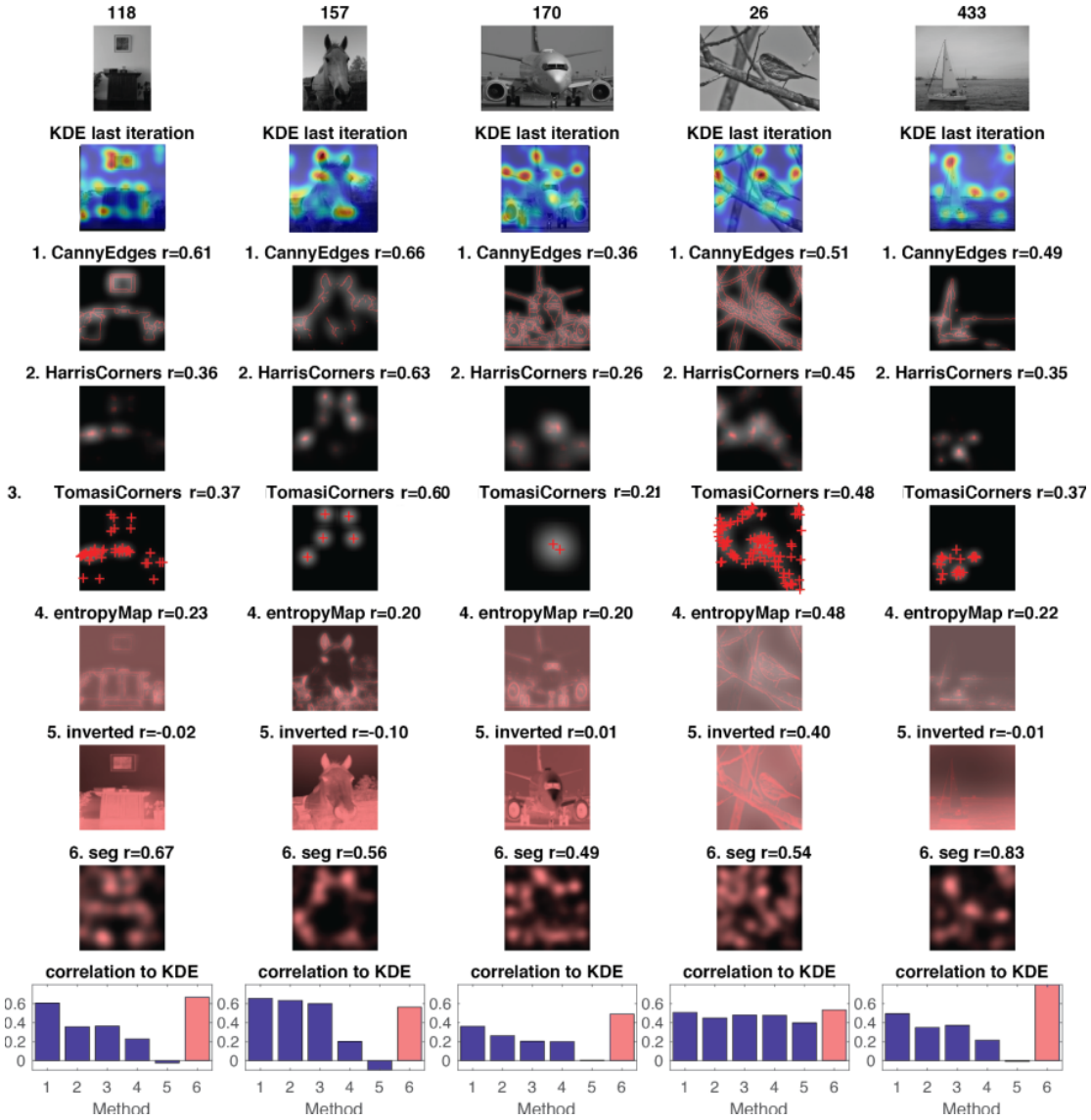

Figure 3.21: Model Comparisons for the PASCAL images. The first row shows the original images, the second row shows images with kernel density estimates of the prior obtained form the final iteration of the serial reproduction chains. The third, fourth, and fifth rows show the maximally predictive results (trying to predict the final iteration KDEs) following a grid search of all the parameters of the Canny edge detector, Harris and Shi-Tomasi corner detectors, followed by an additional smoothing (convolution with a Gaussian kernel). The sixth row shows the optimally predictive results of using an entropy map (described in the text), followed by an additional smoothing. The seventh row is intended as a baseline (lower-bound) prediction—and shows the maximum prediction that can be achieved using the inverse of a gray-scale of the original image to predict the spatial memory prior. The next-to-last row shows the final iteration KDEs obtained from the same memory task, using segmentation maps of the original images. The bottom row shows barplots of the correlations between each of the feature maps shown in each row (denoted by the "method" number, and the KDEs). These correlations are also displayed above each of the subplots in each row, over each image.

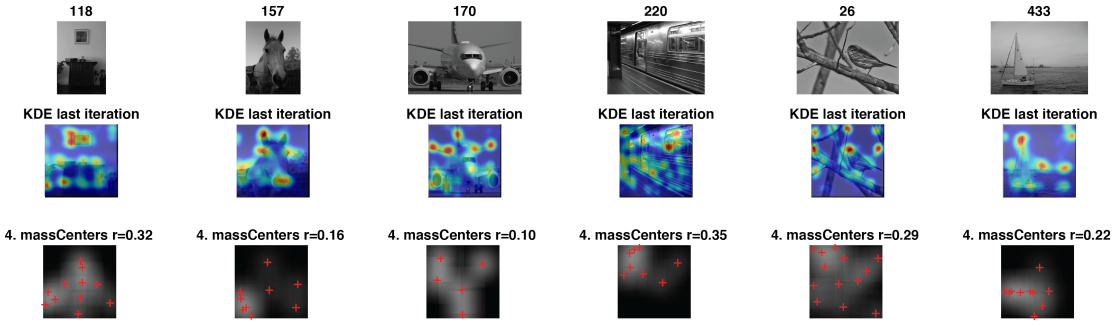

Figure 3.22: Model Comparisons PASCAL images (with an additional Train image from the dataset). This figure illustrates the performance of the "center of mass" hypothesis: The centers of mass inside the segmented regions do not appear to be highly correlated to the modes in the prior we measured.

Convergence of Serial Reproduction Chains

As with the shape priors, we show the Jensen-Shannon Divergence (JSD) measures for two of the natural image results. For all the estimates, we used 50 bootstrap samples with replacement of the data. As before, blue curves show the JSD between consecutive iterations, and red curves show the JSD between each iteration and the final iteration. The first and second rows show the results for the face and tiger images. While the results are not as clear for the natural images as they were for the shapes, they indicate that our estimates likely are converging at or near 20 iterations of the process.

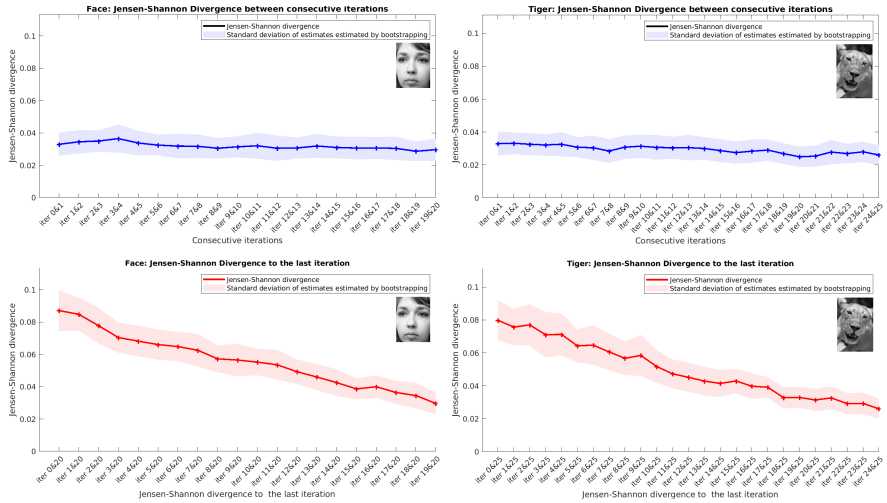

Figure 3.23: Results of convergence analyses using the Jensen-Shannon Divergence (JSD) metric, for natural images. For each image, we show the JSD between consecutive iterations (in blue), as well as the JSD between each iteration and the last iteration (in red).

3.4 Priors and Visual Recognition

If the serial reproduction process reveals the latent geometry of visual memory representations, or compressed representations of the images that we see, one experimental prediction that we can make is that recognition accuracy for images that are masked by their corresponding visual memory prior will be improved relative to recognition accuracy for the same images when masked by different priors. We tested this prediction for both the geometric shapes, as well as the natural images. We found that recognition accuracy was indeed significantly greater for every image that we tested, with d' scores that were significantly greater for images that were masked using Gaussian windows sampled from their corresponding prior than for the same images when masked using Gaussian windows sampled from a different prior chosen at random. Our approach of masking the images with Gaussian windows is similar to the work of Gosselin and Schyns (Gosselin & Schyns, 2001).

3.4.1 Methods

Participants

We recruited participants on Amazon Mechanical Turk. We generated seventy HITs (see the design below) with five assignments per HIT (five unique turkers per HIT). Due to the nature of the HIT (a recognition task, in which multiple exposures to a masked image would likely influence accuracy), we blocked workers from completing more than one of the HITs.

Design

We used eight of the PASCAL images for which we had obtained kernel density estimates of the visual memory prior in the dot game. For each HIT, a total of eight masked images were presented in a forced choice task (where workers were instructed to pick the correct label, from a set of possible image labels). In each HIT the particular set of eight image trials corresponded to a unique permutation of four correctly masked images and four incorrectly masked images (a permutation of a multiset comprised of four instances of incorrect masking and four instances of correct masking). Using the multinomial coefficient:

$$\binom{n}{k_1, k_2, \ldots, k_m} = \frac{n!}{k_1!\, k_2! \cdots k_m!} \tag{3.3}$$

This yielded 70 unique sets, from which we generated the 70 unique HITs. During the experiment, each of the eight images in a given set were presented in random order, and were flashed on

the screen for 250 ms. The forced-choice labels were then presented on the screen, and instructions were to choose the correct label as quickly and accurately as possible. We obtained responses from five unique workers for each of the sets (these workers were also barred from accepting other HITs, since the same images were used across HITs). During the experiment, workers completed some practice trials with images that were not used during the experimental trials (see example).

Of the eight images in each of the HITs, four were masked by Gaussian windows sampled from their corresponding priors (250 centroids for 2D Gaussian kernels with standard deviations of 10 pixels were sampled uniformally from each prior). The remaining four images were masked by Gaussian windows sampled from a randomly chosen prior from the set of seven remaining prior distributions. We normalized each of the masks, and generated the stimuli by an element-wise product of the mask values with corresponding grayscale image intensities.

Stimuli

A representative sample of 5 sets from the full 70 sets of images are shown in Figure 3.24. Again, these images were the same grayscale images used in the dot game, and the masking was accomplished by forming the element-wise product of the image pixel intensities with the corresponding values in the mask image, which was obtained by normalizing a 2D matrix made up of the stacked sum of 250 Gaussian windows with centroids sampled from the prior. An additional parameter we used was a scaling factor that controlled the overall brightness of the final image. For the results presented below, we used images which were darkened by a factor of 2.

Results

We calculated the d scores for each of the images, as well as the overall d score by averaging over all the images. d' scores were computed for each image, and for each condition (correct kernel, or incorrect kernel), by calculating the False Alarm (FA) rate (the number of times a given label was selected when the image shown was not an image of that label, over the number of times that the presented images were not images of that label), and the HIT rate (the number of times that a given label was selected when the image shown was an image of that label, over the number of times that all the presented images were images of that label). d' was then calculated using the following formula:

$$d' = Z(HIT) - Z(FA) \tag{3.4}$$

Where the function $Z(p), p \in [0, 1]$, is the inverse of the cumulative distribution function of the Gaussian distribution.

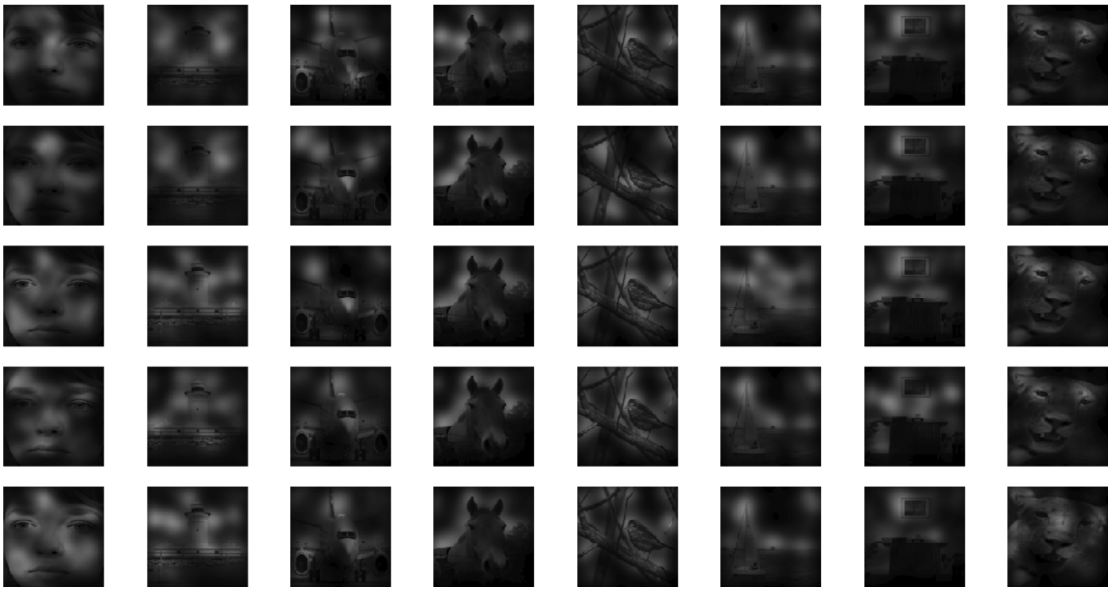

Figure 3.24: A couple of representative image sets for the recognition experiment. The caption indicates whether the image is masked with Gaussian windows sampled from its own (correct) visual memory prior, or with Gaussian windows sampled from one of the remaining image visual memory priors, chosen at random.

In Figure 3.25, we show a barplot of the d scores that were obtained for each of the images, for each of the two possible conditions (Image was masked by the 'correct' kernel, or by the 'incorrect' kernel), when images were presented for 250 ms. Error bars correspond to the standard deviation of the distribution of scores that we obtained by repeating the calculation one hundred times for each of the the image response datasets by generating bootstrapped samples (with replacement) that were the same size as the original dataset. Results show that recognition accuracy d' scores were significantly greater for all the images when they were masked by Gaussian windows sampled from their corresponding visual memory prior distributions. The barplots also show the overall (average d' scores for both conditions, averaged over image category).

In Figure 3.26, we show a barplot of the same scores, for the same set of images, when images were presented for 100 ms instead of 250 ms. The findings are unchanged: higher recognition accuracy for images that were flashed with the correct mask when compared to the same images when they were flashed with the wrong mask.

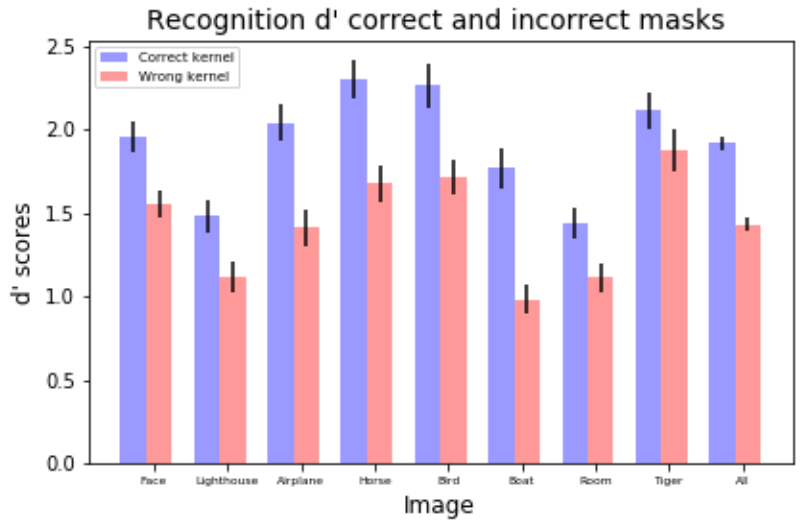

Figure 3.25: d' scores for each of the eight images used in the image recognition task when images were presented for 250 milliseconds. In all cases, d' scores were significantly higher for images that were masked using Gaussian kernels sampled from their corresponding visual memory priors obtained at the end of a serial reproduction process

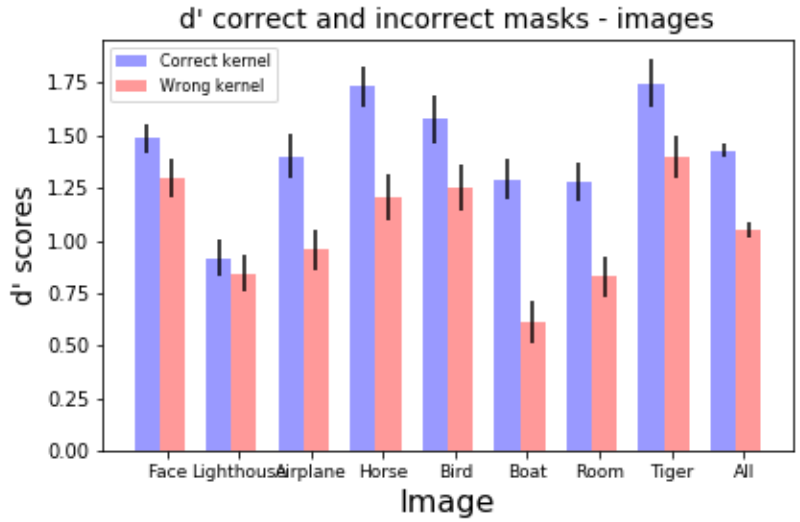

Figure 3.26: d' scores for each of the eight images used in the image recognition task when images were presented for 100 milliseconds. In nearly all cases, d' scores were significantly higher for images that were masked using Gaussian kernels sampled from their corresponding visual memory priors obtained at the end of a serial reproduction process. On average, d' scores were higher for the correctly masked images.

In figure 3.27, we show a barplot of the d' scores that were obtained for a set of the geometric shapes (a square, pentagon, triangle, and circle), using the same experimental design. We repeated the same experiment with a shorter exposure time, and replicated the results (see figure 3.26).

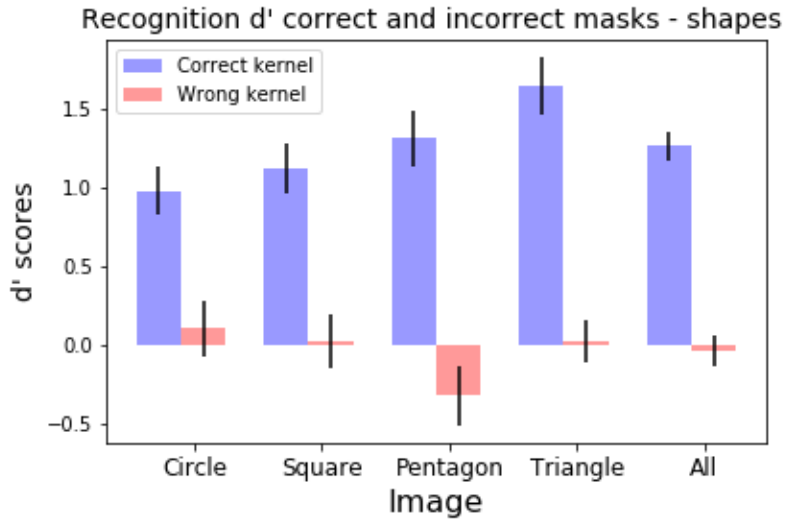

Figure 3.27: d' scores for each of the four shapes used in a similar recognition task. In all cases, d' scores were significantly higher for shapes that were masked using Gaussian kernels sampled from their corresponding visual memory priors obtained at the end of a serial reproduction process

3.5 Priors and Visual Discrimination

While the recognition results we obtained reveal that visual areas biasing spatial memory also happen to be more diagnostic of an image's identity than other regions, a result that is consistent with our model evaluations suggesting that they correspond to "semantic corners" rather than low-level visual features—corners, edges, confusability of image patches), could these areas also simply be those where visual discrimination is most sensitive? The experiments presented here set out to answer this question using the circle as a case-study. Having the priors allows us to make directional hypotheses, and we designed the following experiments to test if discrimination is highest on the borders of the circle where the attractor regions (modes) in the prior are located, and whether it changes in other places along the border where the prior does not contain modes. we sampled regularly spaced points along a line that passed over two diagonal modes in the circle prior (the line contained the locations of the two maxima in the prior), and we also sampled points placed at regular intervals along a horizontal line, containing points at similar distances from the edge of the circle

as the locations of the modes in the first line (but these points had low probability in the spatial memory prior).

3.5.1 Methods

Participants

We recruited participants on Amazon Mechanical Turk (AMT). We generated two HITs (see the design below) with forty assignments per HIT (forty unique participants per HIT), although some were excluded due to missing data for more than 50 percent of the trials. For the first experiment (containing points along a line connecting modes in the spatial memory prior), only 29 workers completed the full experiment. For the second experiment, 37 participants took part and completed the experiment in full.

Design

The basic task in the experiment was for participants to indicate if they noticed a shift in the location of a dot presented over the circle. They viewed the original display with a dot presented over it, and following a delay, they saw a second display, also with a dot presented over it. The displays were shifted by random offsets that were different for the first and second presentations. Other parameters, such as the duration of the first presentation and the fixation (retention period) matched those that were used in the spatial memory experiment (1.5 seconds, and 1 second, respectively). Half of the trials did not show the slightest change between dot locations for the first and second presentations (the "Same" condition), and the other half did contain a shift (the "different" condition). Figure 3.28 illustrates the task, and the two possible conditions.

The first experiment contained 36 points spaced at regular intervals on a diagonal line connecting two mode locations in the prior. There was a total of 148 trials in the experiment—37 trials corresponded to the "different condition" where a given point from the line was shown in the first presentation, and a point adjacent to it on the line was presented in the second presentation. These trials were counterbalanced, yielding 74 "different condition" trials (counterbalanced pairs A-B and B-A positions). There was an identical number of "same condition" trials, where the dot was in the same exact point location in the first and second presentations (A-A, and B-B repeated twice). Participants viewed all 148 trials presented in random order, and simply indicated whether they thought that the dot had shifted position, or if it was the same in both presentations. The second experiment contained fewer points—just 25 points in each condition, yielding a total of 100 trials.

Results

Results are presented in Figure 3.29 and Figure 3.30. Figure 3.29 shows the discrimination results for the first experiment. The barplot in panel A shows the d' discrimination results for points that were at the same distance from the closest mode location (point P1 or P2, where P1 and P2 corresponded to the exact locations of two of the peaks in the spatial memory prior). Each bar corresponds to a collection of points at the same distance from the closest mode location, and the error bars are standard deviations obtained from bootstrapping 100 samples with replacement and computing d' using:

$$d' = Z(HIT) - Z(FA) \tag{3.5}$$

Where the function $Z(p), p \in [0, 1]$, is the inverse of the cumulative distribution function of the Gaussian distribution. We used the exact same procedure as described in the previous section (Recognition experiments) to compute the False Alarm (FA) and HIT rates. The results in panel A clearly show increased discrimination accuracy for points near or under the modes in the spatial memory prior. Panel B shows the percent discrimination accuracy as a function of distance from the closest mode, for the "same" and "shifted" conditions. Panel C illustrates the locations of the points along the line, color-coded according to proximity to the closest mode: blue points were close to a mode, and red points were far from a mode (the threshold was chosen so that the points that were close to a mode roughly equalled the number of points that were far from a mode). This distinction ("close" versus "far") is illustrated in order to facilitate the interpretation of the d' results shown in panel E, which shows d' discrimination results for the points that were far from a mode (in red), and the points that were close (in blue). Panel D shows the visual memory prior on top of the circular region, along with the points. Finally, panel F shows distributions of d' results obtained by bootstrapping 100 samples with replacement, along with a quadratic curve fit. d' results overall appear to show that regions that have higher probability under the prior are also regions where visual discrimination is highest, although the second experiment, presented below, tests whether discrimination is high elsewhere near the edge of the circular region, where there are no modes in the prior. Panel B reveals a main effect of condition: points in the "same condition" tended to be more accurately classified as being "same", whereas more errors were made in correctly classifying the points in the "shifted" condition as being "shifted." This may be due to a product of the variances in the overall random shift of the displays, as well as the unique shifts of the dots themselves within the displays.

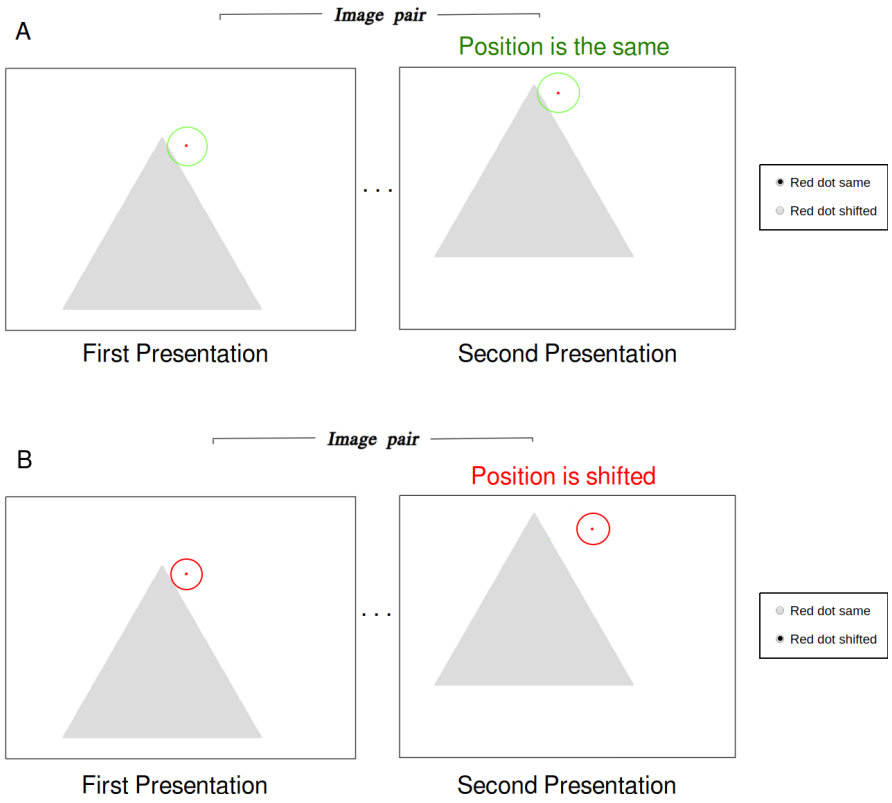

Figure 3.28: The discrimination task: Participants were instructed to indicate if the position of the red dot had shifted, or if it was the same in both the first and second presentations. A illustrates the case when the position was unchanged ("same condition"), and B illustrates the case when the position was shifted "different condition." In both cases, the overall display was shifted at random (for both presentations). The example here shows the dot placed over an image containing a shaded triangular surface, but in the actual task, the dot was presented over an image containing a gray circular disc. The red and green circles are only included here for illustrative purposes, and were not shown in the experimental trials.

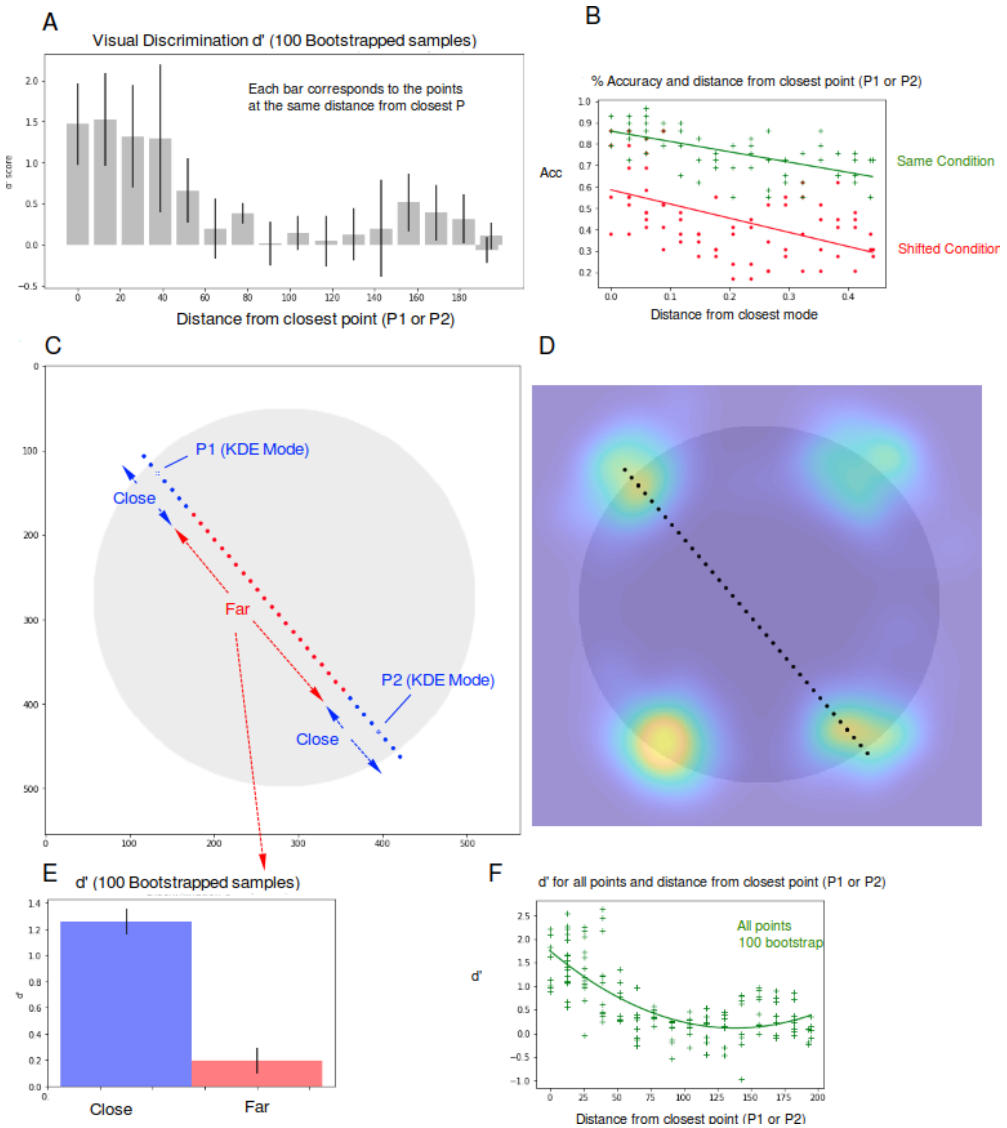

Figure 3.29: Visual discrimination results for points placed along a diagonal line connecting spatial memory KDE modes. The barplot in A shows d' discrimination results for points that were at the same distance from the closest mode location (point P1 or P2). Each bar corresponds to a collection of points at the same distance from the closest mode location. The error bars are standard deviations obtained from bootstrapping 100 samples with replacement and computing d'. B shows percent discrimination accuracy as a function of distance from the closest mode, for the "same" and "shifted" conditions. C illustrates the locations of the points along the line, color-coded according to proximity to the closest mode: blue points were close to a mode, and red points were far from a mode. This distinction is illustrated in order to facilitate the interpretation of the d' results shown in panel E. D overlays the visual memory prior on top of the circular region, along with the points. F shows distributions of d' results obtained by bootstrapping 100 samples with replacement, along with a quadratic curve fit.

Figure 3.30 shows the results for the experiment using a horizontal line (that did not include locations under the modes of the spatial memory prior). The results are very similar: Panel A shows the d' discrimination results for points that were at the same distance from the closest point location (point P1 or P2, where P1 and P2 corresponded to points at nearly the same distance from the edge of the disk as the modes in the prior). Each bar corresponds to a collection of points at the same distance from the closest point P1 or P2, and the error bars are standard deviations obtained from bootstrapping 100 samples with replacement and computing d'. The pattern of results is very similar to what we obtained before, indicating that discrimination might be high near the edge of the disc in general, and not necessarily in an area that has higher probability in the prior.

Panel B also shows the percent discrimination accuracy as a function of distance from the closest mode, for the "same" and "shifted" conditions, and Panel C illustrates the locations of the points along the line, color-coded according to proximity to the closest point P1 or P2, as before (again, the threshold was chosen so that the points that were close to a mode roughly equalled the number of points that were far from a mode). Panel E which shows d' discrimination results for the points that were far from a mode (in red), and the points that were close (in blue). Panel D shows the visual memory prior on top of the circular region, along with the points in order to illustrate the location of the points relative to the prior. Finally, panel F shows distributions of d' results obtained by bootstrapping 100 samples with replacement, along with a quadratic curve fit. d' results overall appear to show that regions that are near the border of the circular disc, although not under modes in the prior are associated with higher discrimination sensitivity. As in the first experiment, Panel B reveals a main effect of condition: points in the "same condition" tended to be more accurately classified as being "same", whereas more errors were made in correctly classifying the points in the "shifted" condition as being "shifted."

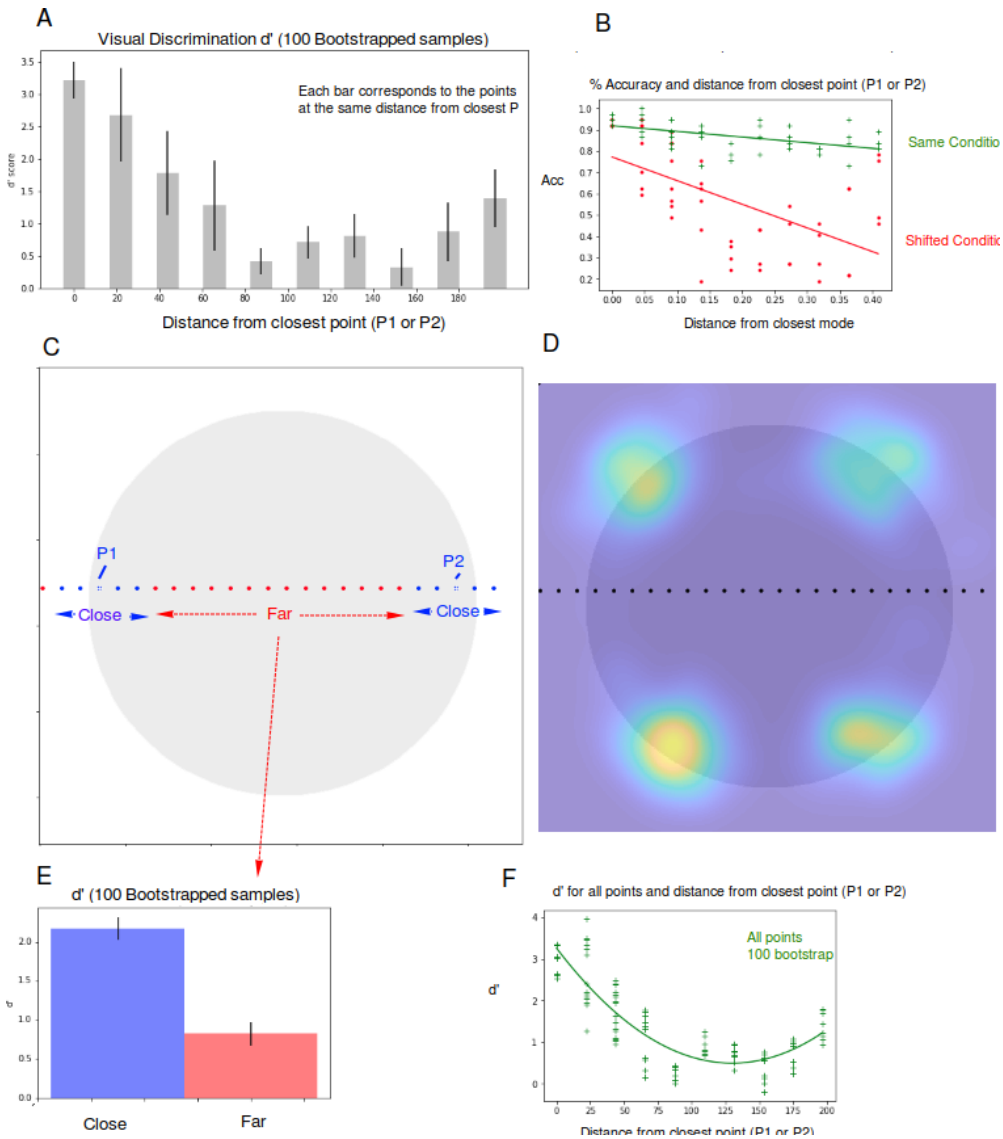

Figure 3.30: Visual discrimination results for points placed along a horizontal line connecting two points near the circle edges at nearly the same distances from the edges as the modes in the prior. The barplot in A shows d' discrimination results for points that were at the same distance from the closest edge point location (point P1 or P2). Each bar corresponds to a collection of points at the same distance from the closest edge point location. The error bars are standard deviations obtained from bootstrapping 100 samples with replacement and computing d'. B shows percent discrimination accuracy as a function of distance from the closest edge point, for the "same" and "shifted" conditions. C illustrates the locations of the points along the line, color-coded according to proximity to the closest edge point: blue points were close to either P1 or P2, and red points were far from either point P1 or P2. This distinction is illustrated in order to facilitate the interpretation of the d' results shown in panel E. D overlays the visual memory prior on top of the circular region, along with the points. F shows distributions of d' results obtained by bootstrapping 100 samples with replacement, along with a quadratic curve fit.

3.5.2 Discussion

While the KDEs we estimated in the spatial memory experiments appear to have some clear structure (structure that was particularly evident in the case of the geometric shapes), why are the peaks in these distributions located over some visual regions and not others? What is it about these regions that seem to form basins of attraction that bias people's spatial memory? Are they regions that correspond to increased visual discrimination acuity? Or regions that correspond to some salient features of the images? Or are they regions that are diagnostic of the image in some way, and support visual recognition? These were the questions that motivated the series of experiments described in this section. What emerges is that the visual regions that bias spatial memory also appear to be regions that are more diagnostic than others in a recognition task: sampling Gaussian windows form the spatial memory prior of an image, and then masking the image itself with these Gaussian windows results in improved recognition accuracy (relative to the recognition accuracy obtained with the same images when they are masked with Gaussian windows sampled from the incorrect spatial memory prior). However, these regions are not necessarily associated with improved visual discrimination sensitivity, although more experiments are needed to establish whether this is true generally. A natural question then is: Why would spatial memory for what amounts to a small perturbation over an image (a small colored dot appearing over the image) be misremembered towards regions that happen to be visual regions that support recognition better than other regions in the same image? Is there some relationship between spatial "location" memory and visual recognition? Or instead, could it be the case that the spatial memory biases we are measuring are by-products of the geometry of visual memory representations? If different regions in memory are encoded with variable precision, with some regions being remembered with greater precision and accuracy than others, perhaps that spatial memory for the dot's location would be biased accordingly in a manner that reveals that geometry.

The recognition finding is also consistent with our findings from some preliminary feature-based prediction evaluations. The lesson we learned from the series of feature-based model predictions we completed is that spatial memory priors are not easily predicted by low-level visual features (such as edges, corners, and gradient-based features), and that the priors do not change substantially when a natural image is replaced by a simplified semantic segmentation map in our memory task: With these model evaluations, we attempted to predict the priors with corner and edge detector features extracted from the images, or other engineered features (entropy maps), which yielded results that were almost always poorer than the predictions we obtained from the priors we estimated from identical spatial memory experiments, but using segmentation maps of the original images. This

finding supports the idea that people's memory is biased towards regions that correspond to "semantic corners"—which may also be visual regions that support visual recognition. A natural extension of this finding would be to use more sophisticated feature learning models (such as semantic segmentation convolutional neural networks) in order to determine whether the features that are learned by these models when carving out semantically meaningful objects and regions in the visual world happen to be similarly predictive of the structure we see in the spatial memory priors. A current drawback of our evaluations is the fact that the improved prediction we have from using the spatial memory priors estimated from the segmentation maps is confounded by the fact that they were also obtained by humans, unlike the other features, which were extracted directly from the images using simple computer vision algorithms. In a similar vein, a worthwhile avenue for future work would consist in training state-of-the-art convolutional neural networks to classify images with a Lagrangian constraint imposed on the number of pixels in the image that may be used (or a fixed budget of pixel-wise weights). Such a constraint would force the network to maximize its classification accuracy while deriving optimal distributions of pixel weights for different images simultaneously.

While our recognition results are significant, suggesting that our task reveals semantically meaningful regions that may correspond to the geometry of visual memory representations (variable precision in encoding for different visual regions), we can't rule out the possibility that spatial memory biases are due simply to variable discrimination sensitivity for different visual regions. Although there is no reason to believe that "semantic corners" and regions that support improved visual discrimination are necessarily different, an explanation for spatial memory biases in terms of variations in visual discrimination sensitivity seems more parsimonious. And although more experiments are needed, we presented data using the prior we obtained for a circle, and determined that discrimination appears to be increased near regions along the circumference of the shaded disc—in areas that are both most likely, *and* least likely under the spatial memory prior. This finding indicates that the modes in the spatial memory prior are not simply due to the presence of areas that facilitate visual discrimination. If that were the case, we would expect it to resemble a ring around the circle's circumference, and not four distinct peaks straddling the disc's edge near the quadrant centers. Other facts, such as the total scale invariance of the circle prior, which presumably would be associated with changes in discrimination (sensitivity in areas that subtend smaller visual angles should be increased), suggest that the priors are not easily explained by simple variations in discrimination sensitivity. Nevertheless, explaining the patterns in terms of "semantic corners" does not explain the very first findings with the shapes when the spatial domains of the initial point locations were restricted to the regions inside the shapes themselves—how are the four windows inside the quadrant centers

of the circle revealing visual areas that are in any way diagnostic of the presence of a circle? Using them as masks would only reveal uniform regions within the boundaries. The recognition results we obtained, which confirmed improved d' results corresponded to the priors and images that were obtained/used when the domain restrictions were removed. These observations warn against broad generalizations based on our findings so far, and suggest that we should err on the side of caution in our attempts at coming up with a unified theory until these inconsistencies are resolved.

4

Multi-view three-dimensional object representations in memory revealed by serial reproduction

4.1 Priors for orthogonal multi-view 3D object representations

Seminal theories of human similarity judgments and categorization introduced the notion of "prototypical" exemplars, which are perceptual or abstract entities of some class or category that contain the most information about that category (Rosch, 1999). In this work, Eleanor Rosch theorized that categorization is done on the basis of similarity to a "prototype." In further work examining predictions of the Prototype Theory for object recognition, Palmer, Rosch and Chase (1981) explored whether people have privileged ("canonical") 3D perspectives of objects. They described a "canonical perspective" as the perspective view that is the "best" or most "typical" view of the object, and they made a number of predictions about what the characteristics of those perspectives should be: They should 1) be the perspectives that first come to mind when people are asked to imagine an object, 2) they should be the perspectives that are most similar to other views of the same object, and most different from views of other objects (in other words, the perspectives that both minimize

within-group categorization error, and at the same time, maximize between-group categorization errors), and 3) they should contain the most information of greatest salience about the object. Finally, they should be most easily recognized as perceptual instances of the object itself—in other words, they should be the exemplars that yield that highest recognition accuracy (categorization). Remarkably, they revealed for the first time that "canonical perspectives" in fact do exist—that people tend to favor one perspective over others for a given object, and that the privileged view differs depending on the object. Furthermore, they found that these views also happened to yield higher recognition accuracy (in terms of naming latency), and that they were correlated with the views that subjects adopted when taking pictures of the same objects. They were also the views that subjects tended to agree on intuitively as far as their "canonicalness."

Palmer et al. proposed a number of different theoretical explanations for why these "canonical perspectives" exist inside people's heads, in spite of the fact that visual perception must be robust to perspective transformations. They proposed two theories: The first one, called the *"frequency hypothesis"* is consistent with the notion of a prior on visual perspectives that we typically take when interacting with a given object: it proposes that the canonical perspective simply corresponds to the perspective from which we are most likely to view an object. The second theory, called the *"maximal information hypothesis"* argues that the canonical perspective is simply the one that is the most informative about the overall 3D structure and identity of the object. In other words, the *"maximal information hypothesis"* corresponds to the idea that the canonical view is the one that looks relatively unchanged under minor perspective transformations.

Given that many common objects exhibit some kind of structural symmetry, we wondered whether canonical perspectives might be correspondingly symmetric as well. Palmer, Rosch and Chase (1981) selected objects that were bilaterally symmetrical so as to minimize the number of views that would have been needed to present to participants in their experiments in order to have a full discrete coverage of possible views of that object. While they confirmed the clear presence of unimodal canonical views for each, is it possible that these views come in pairs for these objects due to their bilateral symmetry? Further work using images returned by online search engines estimated the modes of the distribution of 3D perspectives for a variety of objects, and largely confirmed the findings of Palmer, Rosch and Chase (1981). However, in some cases, the canonical views they identified in the distributions of 3D views for some objects were the mirror image of those identified by Palmer, Rosch and Chase (1981)—suggesting that canonical views for bilaterally symmetric objects may be at least strongly bi-modal, with one privileged view for each side of the object (Mezuman & Weiss, 2012). In this work, we set out to capture the presence of more than a single canonical perspective for some of the same bilaterally symmetric objects used by Palmer, Rosch and Chase (1981):

A teapot, shoe, and clock. We took inspiration from the results of our work in the 2D visual memory domain, and the possibility that we might uncover intricacies that might have escaped detection in earlier work. Although "canonical views" may not necessarily be the same as the *remembered* 3D views of an object, we expected that "canonical views" might still influence people's memory.

4.1.1 Methods

Participants

All participants were recruited online using Amazon Mechanical Turk. All gave informed consent. The experimental protocol was approved by The Committee for the Protection of Human Subjects (CPHS) at the University of California, Berkeley. Each experiment required approximately 100 participants.

Stimuli

The stimuli used in these experiments were 3D objects that could be viewed from any angle by rotating a camera oriented towards the origin of the object, and at a fixed distance (travelling on the surface of a sphere around the object). We started with a detailed mesh model of a typical teapot (which is one of the objects used by Rosch and Palmer), as well as a detailed model of a shoe, and a clock. These everyday objects were chosen because they happen to be among those that yielded the original findings by Rosch and Palmer.

Procedure

We carried out a series of serial reproduction experiments. Participants were presented with timed displays of the 3D object, viewed from a fixed vantage point, sampled uniformally from the surface of a sphere (the domain of possible camera angles, given a fixed distance from the object). Following the retention phase (which lasted 4 seconds), a blank screen (lasting 1 second) was presented, followed by a blank canvas with the object viewed from a different random perspective. Instructions were for the subject to orient the object (equivalent to rotating the camera view) so that it matched the orientation seen during the retention phase. Participants could take as much time as they needed to complete the trial, and change the orientation of the object as many times as they wished. Once done, they could press a "next" button to complete the next trial. There were 10 practice trials followed by 95 experimental trials. In total, there were 700 chains (for the teapot and shoe, and only

350 chains for the clock), with 10 generations in the chain (for each object, two separate experiments were done, each with 350 chains and 10 generations).

4.1.2 Results

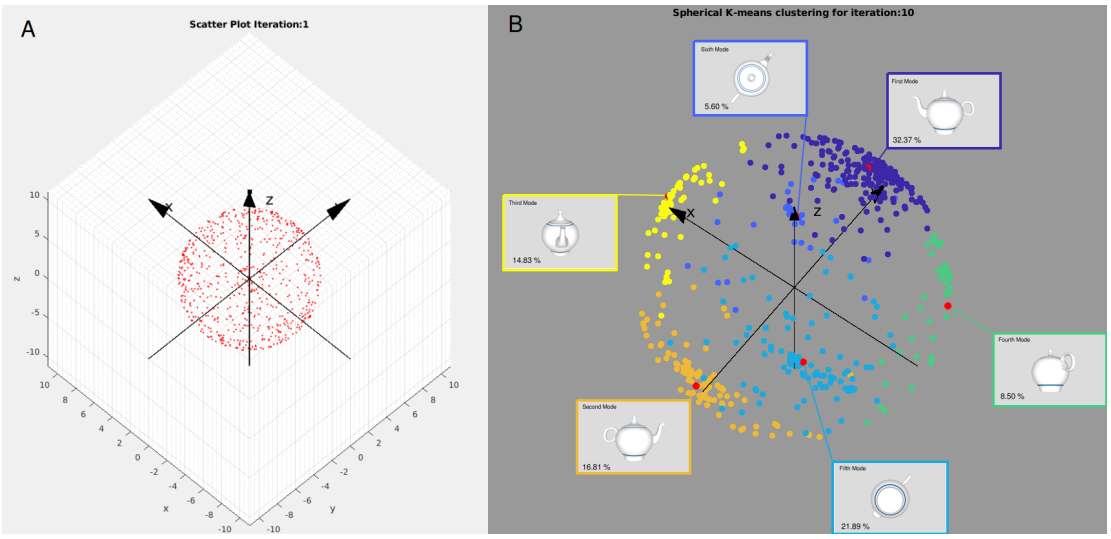

Figure 4.1: 3D Orientation Memory Priors: Teapot. Panel A shows scatter-plots of the nearly 700 views that were originally presented—distributed uniformally over a sphere in the original iteration. Panel B shows the distribution of camera views at the 10th iteration of the serial reproduction process, and is color-coded according to the results of a spherical k-means clustering (k = 6). Each color denotes one of 6 clusters, and the corresponding views are provided. They are the views that correspond to the k-means centroids, shown as red circles. While the distribution of views in the first iteration is completely uniform, by the 5th iteration clear clusters emerge around the ends of the axes of the standard basis (where the ends of the axes intersect the sphere), and are clear by the 10th iteration.

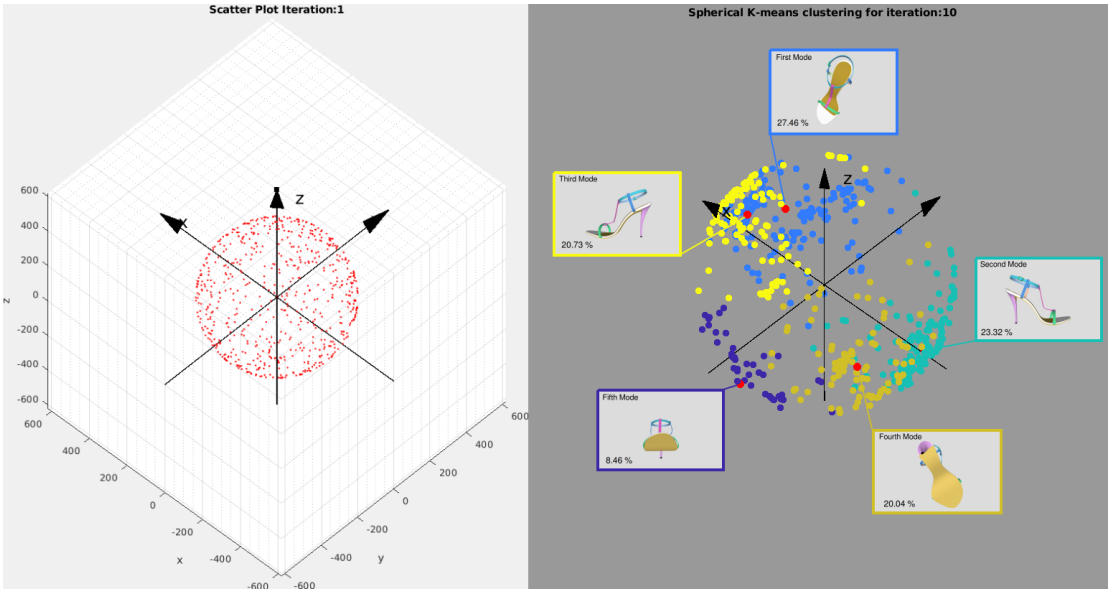

Figure 4.2: 3D Orientation Memory Priors: Shoe. Panel A shows scatter-plots of the nearly 700 views that were originally presented—distributed uniformally over a sphere in the original iteration. Panel B shows the distribution of camera views at the 10th iteration of the serial reproduction process, and is color-coded according to the results of a spherical k-means clustering (k = 5). Each color denotes one of 6 clusters, and the corresponding views are provided. They are the views that correspond to the k-means centroids, shown as red circles. While the distribution of views in the first iteration is completely uniform, by the 5th iteration clear clusters emerge around the ends of the axes of the standard basis (where the ends of the axes intersect the sphere), and are clear by the 10th iteration.

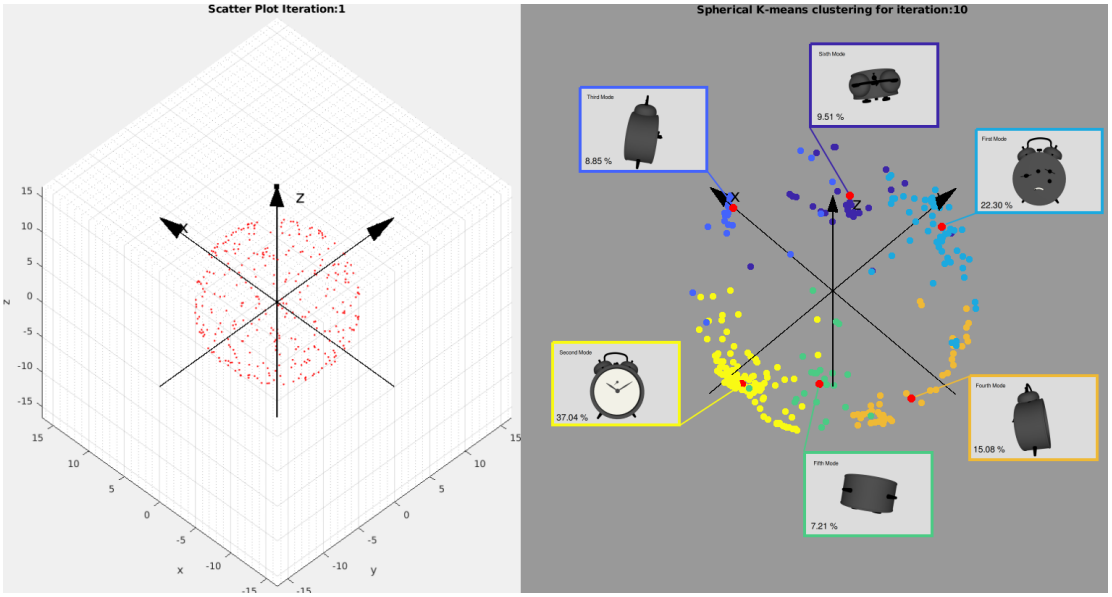

Figure 4.3: 3D Orientation Memory Priors: Clock. Panel A shows scatter-plots of the nearly 350 views that were originally presented—distributed uniformally over a sphere in the original iteration. Panel B shows the distribution of camera views at the 10th iteration of the serial reproduction process, and is color-coded according to the results of a spherical k-means clustering (k = 6). Each color denotes one of 6 clusters, and the corresponding views are provided. They are the views that correspond to the k-means centroids, shown as red circles. While the distribution of views in the first iteration is completely uniform, by the 5th iteration clear clusters emerge around the ends of the axes of the standard basis (where the ends of the axes intersect the sphere), and are clear by the 10th iteration.

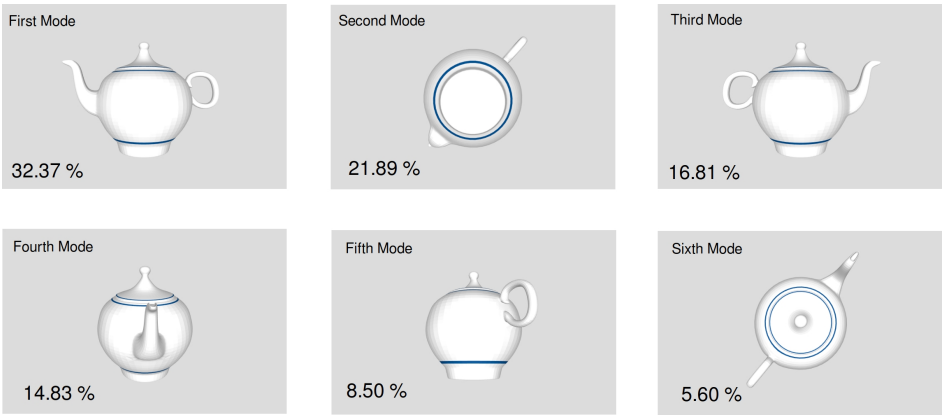

Figure 4.4: Top 3D modal views obtained from spherical k-means clustering of the distribution of views at iteration 10 for the Teapot. Clockwise from left to right: The ranked perspectives that correspond to the centroids of the spherical k-means clustering analysis (k = 6). These views clearly correspond to diagrammatic views of each of the sides of the object—the left and right sides, front and back, followed by the bottom and top of the teapot . The modes were ranked according to the proportion of samples in each cluster, also shown for each view.

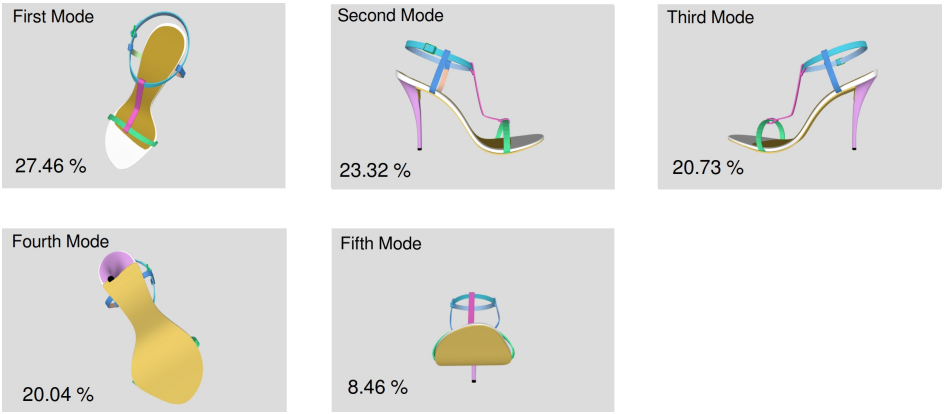

Figure 4.5: Top 3D modal views obtained from spherical k-means clustering of the distribution of views at iteration 10 for the Shoe. Clockwise from left to right: The ranked perspectives that correspond to the centroids of the spherical k-means clustering analysis (here 5 clusters were used). These views also correspond to diagrammatic views of each of the sides of the object—the right and left sides, front, top and bottom of the shoe. The modes were ranked according to the proportion of samples in each cluster, also shown for each view.

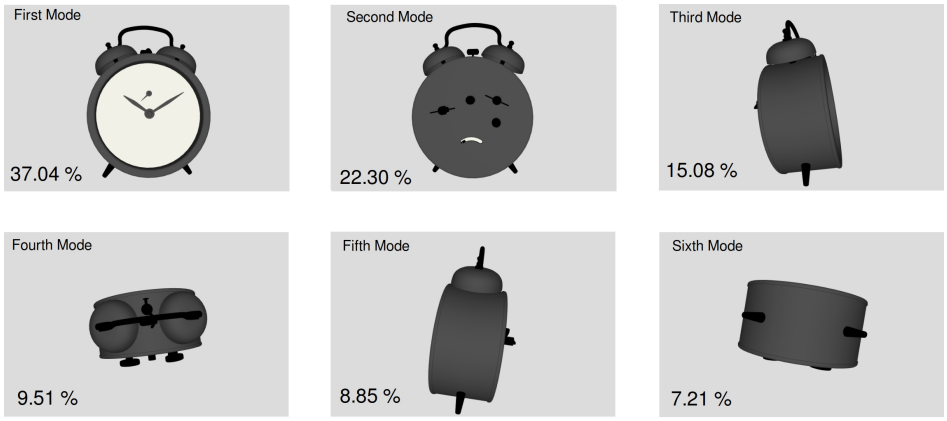

Figure 4.6: Top 3D modal views obtained from spherical k-means clustering of the distribution of views at iteration 10 for the Clock. Clockwise from left to right: The ranked perspectives that correspond to the centroids of the spherical k-means clustering analysis (k = 6). These results als clearly reveal a strong bias towards orthoganal diagrammatic perspectives. The first two show the back and front of the clock, followed by the right side, and the left side, the bottom, and the top. The modes were ranked according to the proportion of samples in each cluster, also shown for each view.

Figures 4.1, 4.2, 4.3 show, for each object, the initial uniform distribution of camera perspectives (in panel A), as well as the distributions of the same views at iteration 10, color-coded according to cluster membership of the points following a spherical k-means clustering (in panel B). The picture that emerges for all three objects is a bias for 3D perspectives to be remembered towards a small set of axis aligned orthogonal perspectives in the standard basis, where the views correspond to the six faces of the objects.

Using the distributions of camera perspectives obtained for the final iteration of the serial reproduction process, we computed spherical k-means clusters for all the objects and include images of the objects viewed from these perspectives. Remarkably, k-means clustering produced centroids that were located close to the ends of the axes in the standard basis. Figure 4.4 shows the centroid perspective views corresponding to the k-means clusters (for six clusters), sorted according to the proportions of samples in each cluster. Figure 4.5 shows the centroid views for the shoe, and Figure 4.6 shows the same for the clock. In all cases, the same pattern emerges: 3D orientation memory representations appear to be low-dimensional, orthogonal multiple-viewpoint specific representations, like diagrammatic projections of the object faces onto the planes spanned by pairs of axes in the canonical standard basis over which the objects are aligned.

4.1.3 Discussion

Our findings suggest that 3D orientation memory for everyday objects are largely *not* the same as "canonical perspectives." In fact, multiple iterations of the serial reproduction process reveal that people's orientation priors for a small set of everyday objects correspond to views orthogonal to the projections of those objects onto the planes of a standard basis. While we only used three everyday 3D objects so far (a teapot, shoe, and clock), in all cases we find that memory for the 3D orientation for these objects is strongly biased towards simplified, and what might be called "diagrammatic" views, similar to those that one might find in architectural or designer blueprints. They are characterized by sharp profile views, views from the front and back, and views from the top and bottom. Remarkably, this pattern is the same across the three objects, and in all cases the uniform distribution of views is quickly altered in people's collective memory into one of these diagrammatic perspectives. So although we had initially expected to find multi-modal "canonical perspectives," the perspectives that bias memory do not seem to be those that people would necessarily spontaneously imagine when picturing one of these objects in their heads, or even the views that would facilitate visual recognition, although some seem consistent with "canonical perspectives" at first glance. Had we employed experimental techniques that would have yielded only a single view, we might even have been led to believe that 3D orientation memory is biased towards well-known "canonical perspectives" such as the first modal teapot view, or that of the shoe (although in the case of the clock, the back view is marginally more probable under the prior we estimated than the frontal view, although they are probably equally likely). However, the characteristics of the additional modes, although they decrease in probability under the prior, reveal a systematic pattern: 3D orientation memory is biased towards the ends of the x, y and z axes in an orthogonal standard basis—with the object centered at the origin, and oriented upright, rather than slanting at odd angles.

5
Conclusion

The original objective of this body of work was to probe for visual memory biases using techniques that might better reveal their detailed structure. Through several case studies in multiple visual domains, we were able to both confirm, and often amend, previous findings in spatial memory, and memory for the 3D orientation of objects. In the domain of spatial memory, preliminary work using serial reproduction chains of a simple memory task revealed highly structured memory priors for the remembered position of a dot presented over an image, and further work confirmed multi-modal canonical perspectives in 3D orientation memory for different objects. The increased resolution afforded by an improved signal-to-noise ratio, combined with non-parametric modeling approaches that make fewer theoretical commitments than traditional models (such as the Category Adjustment Model) allowed us to make several novel contributions. The first is to illustrate a means for uncovering human visual biases in much greater detail (a "higher resolution" picture). The second is to challenge some current theoretical explanations that were based on noisier empirical observations, and to spur new theoretical insights, although a definitive account of what explains the patterns we observed remains elusive.

In the first chapter, we present novel findings that indicate spatial composition biases in aesthetic preferences for the position of objects within a frame. These biases replicate the well-known inward bias, but reveal systematic effects of manipulating the implied speed and trajectories of objects on their preferred placement within a background frame. In particular, faster appearing leftward

and rightward facing objects are shown to be preferred placed nearer the centers of the frame than stationary ones, as if participants simulate their motion forward in time. Additional experiments using serial reproduction chains probed whether the inward bias in human aesthetic preferences are present in spatial memory biases (e.g. can compositional preferences be predicted from spatial memory priors?). Findings reveal clear off-center biases in the memory for the position of a rightward and leftward facing object, although not an inward bias.

In the domain of spatial memory, we outlined two main bodies of work—the first investigated distortions in the remembered position of a dot positioned over controlled simple images (curved shapes and regular polygons), and the second investigated these distortions when simple shapes were substituted for complex natural images. In both cases, we uncovered highly replicable, and clearly structured results that would have been difficult, if not impossible to reveal using traditional experimental methods. Building on prior work that confirmed the presence of "prototypes" biasing spatial memory, we estimated distributions of the prior (rather than just point estimates) and investigated the effect of manipulating some simple visual features on the prior's structure. The presence of vertices in the geometric shapes had a predictable effect on spatial memory—leading to a bias towards paired attractor regions on either side of each vertex. However, we determined that given a fixed set of experimental parameters (the duration of the original display, as well as the duration of the retention phase in the task) there appeared to be an upper bound in the complexity of the prior, with bi-modal clusters around the vertices merging into the four quadrant modes of a circle as we multiplied the number of edges in the regular polygons we presented to participants. We also determined that, at least for the circle shape, there is a remarkable scale invariance in the prior's structure. Altering the size of the circle did not change the emergence of four modes centered inside each of the circle quadrants. In addition, manipulating a square or triangular frame of reference around the circle did not yield any significant differences, but appeared to interact when we introduced a 45 degree rotation applied to the bounding frames: For the circle bounded by a square frame, this yielded a prior with eight distinct modes rather than four, and for the circle bounded with a rotated triangle, six modes appeared to emerge. Finally, altering the exposure time of the timed display for a regular polygon with many (21) vertices yielded a pattern similar to the one observed with a simple circle. While a definitive theoretical account remains elusive, these results demonstrate the possibility for measuring even the slightest deviations that result from manipulating a variety of visual features. Further work loosening the spatial domain constraints with the shapes reveal a similar structure, although additional modes appear on either side of the shape boundaries. Convergence analyses indicate in all cases that using 500 chains with 20 iterations yields a stable estimate of the prior by the 20th iteration of the process, and often earlier. All in all, experiments using the simple shapes often

raised many questions, and the need for definitive explanations for the patterns we observed remains. However, the intricate patterns that were uncovered provide fertile ground for new thinking on the nature of spatial memory biases, and provide an improved empirical basis for this purpose.

Using natural images in our task yielded estimates with even greater complexity than those that we obtained for simple shapes. While prior work presented evidence for spatial memory biases towards the centers of mass of segmented image regions, our results reveal a more complex picture, although with clear regularities. For instance, images of faces all show memory biases towards salient features—the eyes, nose, and mouth. Having the possibility to estimate these complex priors using naturalistic stimuli afforded the opportunity to investigate which image features are predictive of spatial memory biases. While the experiments with simple shapes suggested that edges near vertices are clear attractors, the patterns in the priors using natural images were not easily predicted with low level edge and corner feature detectors (following a grid search over the parameter settings of the Canny edge detector, Harris, and Shi-Tomasi corner detectors). However, we found that obtaining priors on the same task using the segmentation maps yielded similar patterns to those that we estimated from the original images. The spatial memory priors obtained from using the image segmentation maps turned out to be largely more predictive of the natural image priors than the low-level edge and corner features, suggesting that biases in spatial memory are due to higher-order visual features, or "semantic corners". This result suggests new avenues for using contempory computer vision feature learning models, such as convolutional neural networks (CNNs), or sparse auto-encoders that are trained to reproduce an input image from a lower-dimensional latent representation. To the extent that the biases in spatial memory are an incidental bi-product of irregular visual memory representations, can compressed representations of these images learned from auto-encoders predict these patterns? Another approach would be to use a pre-trained CNN with a Lagrangian constraint at the input layer that restricts the number of input pixels that may be used by the network in a classification task. The network would then need to derive additional weight parameters for each of the pixel intensities, and learn the optimal distribution of a limited budget of pixel-wise activations while optimizing classification accuracy. While arriving at a clear understanding of the features that drive these visual memory biases will take more time, the quality of our empirical measures allowed us to challenge the "center of mass" theory, as well as other accounts based on simple low-level visual features.

The notion of "semantic corners" may be related to the notion of visual features that are diagnostic of an image or object's identity. Are "semantic corners" the visual features that are the most characteristic of "prototypes"? or the most diagnostic of some category membership? We conducted several experiments in which we sampled Gaussian windows from the spatial memory priors we de-

rived for a small set of images. We then generated sets of images that were either masked by windows sampled from their own prior (correct masking), or from priors corresponding to other images in the set (incorrect masks). We predicted that if our priors are correlated to "semantic corners" that are indeed diagnostic of the image's identity, that we would measure improved recognition accuracy for images that were correctly masked when compared to the same images if they were incorrectly masked. Using different presentation times in the recognition task, we confirmed and replicated significantly higher d' recognition values for the correctly masked images compared to the d' recognition values for the incorrectly masked images. We repeated the recognition task with a subset of the simple shapes, with similar results. These recognition experiments suggest that the spatial memory prior convergences towards regions that happen to be diagnostic of an object's identity. We followed the series of recognition experiments with some preliminary experiments looking at whether people show increased discrimination sensitivity in regions that have higher probability under the spatial memory prior. Preliminary results using the circle prior suggests that while discrimination is improved over regions that have a higher probability under the spatial memory prior, they are also improved in other regions, suggesting that our spatial memory priors are not necessarily measures of regions where visual discrimination is highest. Additional experiments are needed in order to establish whether a correlation really exists between the spatial memory priors we estimate, and visual areas linked to higher discrimination accuracy.

In chapter 3, we presented some preliminary findings looking at the remembered 3D orientation of a small sample of everyday objects. To the extent that memory for 3D orientation corresponds to "canonical perspectives", we find that although 3D orientation priors for bilaterally symmetric objects are strongly biased towards one privileged view, in all cases they show the clear presence of additional modes, which typically include the mirror image of of the first modal perspective, which we predicted might be present due to the structural symmetry of the objects themselves. However, the additional modes appear to be "diagrammatic" views, similar to the views that one finds in engineering blueprints of objects, or mugshots—they show strong side views (For the teapot, these are the left side (with the handle on the right), the bottom view, and the view from the right side. These are followed by the front view of the spout, the back view, and the top view in a spherical k-means clustering analysis). For the 3D shoe, the modal views that were revealed, also through spherical k-means clustering, appear to be sharp lateral views (top view, right and left profile views, bottom and front views). The same holds for the clock, with the first modal views being the front and back faces, followed by the right profile view (with the clock face facing to the right), the top view, left profile view (with the clock face facing to the left), and the bottom view.

What transpires from a close look at the 3D orientation memory priors for the everyday objects

we selected, is that the prior in memory does not correspond to "canonical perspectives." This came as a surprise, as we expected to see canonical views emerge. Although seminal work on "canonical perspectives" never defined or tested these views in a memory task, later work investigated the effects of object perspective in a memory task, and found that objects that had been studied from a non-canonical perspective produced lower recall rates than objects studied from the canonical perspective discovered by Palmer, Rosch and Chase (Gomez, Shutter, & Rouder, 2008), suggesting that canonical perspectives may be privileged views in 3D orientation memory. Our findings indicate that people reconstruct the 3D orientation of objects in terms of representations of these objects that are aligned to an orthogonal standard basis, and relative to views along each of the axes in that standard basis—views that maximize the visibility of only one face of the object at a time, unlike "canonical perspectives," which often include semi-profile views, and which have been described in terms of the views that maximize information about the overall 3D structure of the object (the "maximal information hypothesis"). The modal views we measured neither seem to be the views that would be predicted by the "frequency hypothesis," or the "maximal information hypothesis." They are abstract entities, which may be shaped by semantic or linguistic knowledge of how an object may be visually partitioned into a discrete number of faces. They appear to show multi-view 3D representations of objects in memory that are relatively low-dimensional, a finding that is consistent with some theoretical and experimental findings examining cognitive representations of 3D object structure (Bülthoff, Edelman, & Tarr, 1995).

While many questions remain, this dissertation illustrates the promise of using serial reproduction chains to study biases in visual perception and memory. In a sense, it is because of the remarkable precision and fidelity with which we could make our empirical observations that we don't yet have definitive answers that can fully account for them, since a high-resolution picture of these phenomena reveal complex patterns that are challenging to model, predict, and understand. Yet that precision and fidelity allowed us to amend prior work, both empirically and theoretically, by showing that explanations based on noisier estimates did not always explain the biases in people's visual memory once these were more clearly revealed. We are hopeful that this empirical approach will spur new theoretical insights and lead to a better understanding of human spatial memory biases. We ended with a foray into the domain of 3D orientation memory, investigating the structure of human 3D object representations. While this work is still in its infancy, it is already revealing remarkable structure, which, like our results in the 2D memory domain, will inform our current understanding of the structure of human visual memory representations.

References

Alexander, C. (2002). *The nature of order: the process of creating life*. Taylor & Francis.

Amheim, R. (1974). *Art and the visual perception: A psychology of the creative eye*. University of California Press.

Arnheim, R. (1954). 1974. *Art and Visual Perception: A Psychology of the Creative Eye: The New Version*.

Arnheim, R. (1983). *The power of the center: A study of composition in the visual arts*. Univ of California Press.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes1. In *Psychology of learning and motivation* (Vol. 2, pp. 89–195). Elsevier.

Bartlett, F. C. (1932). Remembering: An experimental and social study. *Cambridge: Cambridge University*.

Bertamini, M., Bennett, K. M., & Bode, C. (2011). The anterior bias in visual art: The case of images of animals. *Laterality: Asymmetries of Body, Brain and Cognition*, *16*(6), 673–689.

Bode, C., Bertamini, M., & Helmy, M. S. (2016). Left–right position in moving images: An analysis of face orientation, face position, and movement direction in eight action films. *Art & Perception*, *4*(3), 241–263.

Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of vision*, *11*(5), 4–4.

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, *105*(38), 14325–14329.

Bülthoff, H. H., Edelman, S. Y., & Tarr, M. J. (1995). How are three-dimensional objects represented in the brain? *Cerebral Cortex*, *5*(3), 247–260.

Claidière, N., Smith, K., Kirby, S., & Fagot, J. (2014). Cultural evolution of systematically structured behaviour in a non-human primate. *Proceedings of the Royal Society of London B: Biological Sciences*, *281*(1797), 20141541.

Freyd, J. J., & Finke, R. A. (1984). Representational momentum. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(1), 126.

Gibson, J. J. (2014). *The ecological approach to visual perception: classic edition*. Psychology Press.

Gomez, P., Shutter, J., & Rouder, J. N. (2008). Memory for objects in canonical and noncanonical viewpoints. *Psychonomic bulletin & review*, *15*(5), 940–944.

Gosselin, F., & Schyns, P. G. (2001). Bubbles: a technique to reveal the use of information in recognition tasks. *Vision research*, *41*(17), 2261–2271.

Holden, M. P., Newcombe, N. S., & Shipley, T. F. (2013). Location memory in the real world: Category adjustment effects in 3-dimensional space. *Cognition*, *128*(1), 45–55.

Hubbard, T. L. (2005). Representational momentum and related displacements in spatial memory: A review of the findings. *Psychonomic Bulletin & Review*, *12*(5), 822–851.

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological review*, *98*(3), 352.

Jacoby, N., & McDermott, J. H. (2017). Integer ratio priors on musical rhythm revealed cross-culturally by iterated reproduction. *Current Biology*.

James, W. (2013). *The principles of psychology*. Read Books Ltd.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*(31), 10681–10686.

Langlois, T. A., Jacoby, N., Suchow, J., & Griffiths, T. L. (2017). Uncovering visual priors in spatial memory using serial reproduction.

Lew, T., & Vul, E. (2015). Structured priors in visual working memory revealed through iterated learning. In *Cogsci*.

Leyssen, M. H., Linsen, S., Sammartino, J., & Palmer, S. E. (2012). Aesthetic preference for spatial composition in multiobject pictures. *i-Perception*, *3*(1), 25–49.

Locher, P., Overbeeke, K., & Stappers, P. J. (2005). Spatial balance of color triads in the abstract art of piet mondrian. *Perception*, *34*(2), 169–189.

Locher, P. J. (2003). An empirical investigation of the visual rightness theory of picture perception. *Acta Psychologica*, *114*(2), 147–164.

Locher, P. J., Stappers, P. J., & Overbeeke, K. (1998). The role of balance as an organizing design principle underlying adults' compositional strategies for creating visual displays. *Acta Psychologica*, *99*(2), 141–161.

Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory*, *12*(4), 361–366.

Marr, D. (1982). Vision: A computational investigation into the human representation and processing of visual information. mit press. *Cambridge, Massachusetts*.

McManus, I., Edmondson, D., & Rodger, J. (1985). Balance in pictures. *British Journal of Psychology*, *76*(3), 311–324.

Mezuman, E., & Weiss, Y. (2012). Learning about canonical views from internet image collections. In *Advances in neural information processing systems* (pp. 719–727).

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, *63*(2), 81.

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), 421–449.

Palmer, S. E., & Langlois, T. A. (2017). Effects of implied motion and facing direction on positional preferences in single-object pictures. *Perception*, *46*(7), 815–829.

Palmer, S. E., Schloss, K. B., & Sammartino, J. (2013). Visual aesthetics and human preference. *Annual review of psychology*, *64*, 77–107.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & Team, R. D. C. (2010). *Nlme: Linear and nonlinear mixed effects models, r package version.* R Foundation for Statistical Computing Vienna, Austria.

Rosch, E. (1999). Principles of categorization. *Concepts: core readings*, *189*.

Sammartino, J., & Palmer, S. E. (2012). Aesthetic issues in spatial composition: Effects of vertical position and perspective on framing single objects. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(4), 865.

Wedell, D. H., Fitting, S., & Allen, G. L. (2007). Shape effects on memory for location. *Psychonomic Bulletin & Review*, *14*(4), 681–686.

Wei, X.-X., & Stocker, A. A. (2016). Mutual information, fisher information, and efficient coding. *Neural computation*.

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience*, *5*(6), 598–604.

Wiseman, S., & Neisser, U. (1974). Perceptual organization as a determinant of visual recognition memory. *The American Journal of Psychology*, 675–681.

Xu, J., & Griffiths, T. L. (2010). A rational analysis of the effects of memory biases on serial reproduction. *Cognitive Psychology*, *60*(2), 107–126.