Object Recognition in Spatial and Temporal Context:
Crowding and Serial Dependence

By

Wesley James Chaney

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Vision Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor David Whitney (chair)
Professor Dennis M. Levi
Professor Richard Ivry

Spring 2017

Abstract

Object Recognition in Spatial and Temporal Context:
Crowding and Serial Dependence

by

Wesley James Chaney

Doctor of Philosophy in Vision Science

University of California, Berkeley

Professor David Whitney, Chair

Object recognition is a fundamental function of the human visual system. Our visual system must recognize objects under a wide variety of environmental conditions and despite inherent noise in the nervous system. How the brain accomplishes such accurate and reliable performance on object recognition tasks is widely studied but poorly understood. Less understood still is how the brain accomplishes object recognition in a broader spatial and temporal context, as many studies focus on studying the recognition of objects in isolation. Objects in the real world are more often seen and recognized in both a surrounding visual environment, and for more than one moment in a dynamic world. Here we address these questions by examining the phenomena of crowding and serial dependence. Crowding is an interaction of spatial context; it is the inability to recognize an object when it is surrounded by other similar objects, which would otherwise be resolved in isolation. Serial dependence is an interaction of temporal context; a target appears more similar to a previously seen stimulus. First, we ask what types of information survive crowding and what types of representations can crowd one another. Through computational modeling, we present evidence that crowding is a perceptual decision level phenomenon that occurs when we must make examine a distributed representation with limited attentional resources. Next, in a behavioral experiment, we show that lighting information in the context of a scene can influence how nearby object representations interact in crowding. Finally, we show that temporal context can influence the perception of even high level dynamic stimuli like point-light walkers through the mechanism of serial dependence. Together, these experiments show that both spatial and temporal context play an important role in object recognition and that the brain uses high level object representations that can interact in both space and time to accomplish recognition.

Acknowledgments

This work would not have been possible without the support of my family, friends, labmates, advisor, and especially my wife, Kellie Young.

Table of Contents

Chapter 1: Introduction

Object recognition is one of the most fundamental tasks of the human visual system. We effortlessly perceive and recognize objects in the world around us across a wide variety of environments and conditions, such as lighting changes, object movement and rotation, and the presence of distracting information in a scene. The complexity of this task is masked by the apparent ease with which our brain is capable of recognizing objects, from simple objects such as the characters in this text to the very precise high level discrimination, such as identity of human faces.

Accurate discrimination of an object's component features is crucial for recognition and many factors affect the discriminability of these features, from simple properties such as contrast and orientation to more cognitive processes such as attention. Generally, object recognition is thought to primarily consist of a sequential processing stream that builds increasingly complex features from visual input, from occipital to temporal cortex. However, the processes underlying the recognition of single objects are still not fully understood, and research in this domain has been made even more complicated by the fact that spatial and temporal context influence object recognition at many different levels of this processing stream. The influence of context on object recognition includes effects of scene context, negative aftereffects, temporal effects such as recency, and long term adaptation. Here, I investigated two specific ways in which object recognition varies in spatial and temporal context: crowding and serial dependence.

Crowding occurs in the spatial domain. It is the "deleterious effect of surrounding clutter on object recognition" (Whitney & Levi, 2011) and it is a fundamental limit on our ability to recognize objects. Crowding occurs over the vast majority of our visual field and places limits on recognition in the periphery that go far beyond the limits of visual acuity. Objects that can easily be recognized in isolation at a particular location in the visual field can become completely unrecognizable in the presence of nearby flankers. Crucially, crowding impairs only the identification/discrimination of features and objects, leaving detection unaffected. Hence, it is considered an important tool for investigating visual processing, as it impairs only a specific, intermediate step of object recognition. Understanding crowding is also critical for understanding the types of computation that happen over large parts of the human visual field.

Serial dependence, on the other hand, occurs in the temporal domain. It is the systematic attraction of the current percept to a similar recently seen stimulus. Like crowding, serial dependence leads to illusory changes in the appearance of stimulus features that are used for object recognition. However, serial dependence is different from crowding in that it may serve to facilitate object recognition. Objects in the world are highly autocorrelated, as they generally do not disappear at random and rarely undergo dramatic changes in their characteristics. Therefore, a systematic attraction to what was previous seen, tuned by similarity, helps to generate perceived stability from otherwise noisy and rapidly changing visual inputs. Serial dependence provides evidence for a continuity field, a spatio-temporal operator in the human system that serves to integrate object representations over time in a manner that promotes stability of visual perception.

Many theories of crowding posit a loss of information and representational power that occurs in low level visual processing. However, these models cannot account for crowding between high level holistic representations, such as faces. In Chapter 2, we present a computational model of crowding called the Hierarchical Sparse Selection Model (HSSM) that can resolve these conflicts between existing models and the known data. The HSSM makes testable predictions and shows how crowding occurs at multiple levels in the visual system during the read-out process of a population code. This allows a high-fidelity population level representation to pass through successive layers of visual processing, preserving the necessary information for high-level object recognition, while also predicting the crowding effect for stimuli at each level of visual processing.

In Chapter 3, I present evidence that lighting interpretation in a simple scene can affect the magnitude of crowding. This presents a further problem for low level crowding models: I show that crowding of a simple feature (orientation) can be affected by the interpretation of 3D shape from shading under a specific assumption of a single light source from above. This effect has several hallmark properties of visual crowding and cannot be explained by most competing models of crowding.

In Chapter 4, in the temporal domain, I show that serial dependence can also occur between perceived biological motion from point-light walkers. Serial dependence has been previously shown to attract the perception of a current stimulus to the perception of a previous stimulus across many levels of the visual system from orientation, to numerosity, to face identity, emotion, and attractiveness. Recently, serial dependence was also shown for motion direction. However, this is the first work to provide evidence for serial dependence in dynamic, global stimuli. Our work extends the domain of serial dependence to include object representations in the visual system that require the integration of high level form over time.

In sum, I provide evidence that both serial dependence and crowding are general phenomena which operate on a wide variety of object representations, including 3D form and temporal dynamics. In order to understand object recognition, we must take into account spatial and temporal context at multiple levels from low level features, such as orientation, to high level features like lighting interpretation and biological motion, and allow for the top-down influence of our prior assumptions about lighting and form.

Chapter 2: The Hierarchical Sparse Selection Model of Crowding

Introduction

Peripheral vision is not what it seems. Despite the subjective experience of seeing rich detail throughout the visual field, if we are pressed to report the identity of one individual object among others in the periphery, we are very often unable to do so due to the phenomenon of crowding (Levi 2008) (Fig. 1). Crowding occurs when an object appears among clutter; we lose individual access to the identities of objects spaced too closely together. Access to individual objects is replaced with access to textures of objects – we have an impression of the kind of "stuff" that occupies different regions of space, but no awareness of individual items (Cavanagh 2001; Tyler and Likova 2007; Balas, Nakano et al. 2009; Greenwood, Bex et al. 2009; Freeman and Simoncelli 2011). Crowding imposes a fundamental limitation on our ability to identify objects in everyday life (Whitney and Levi 2011).

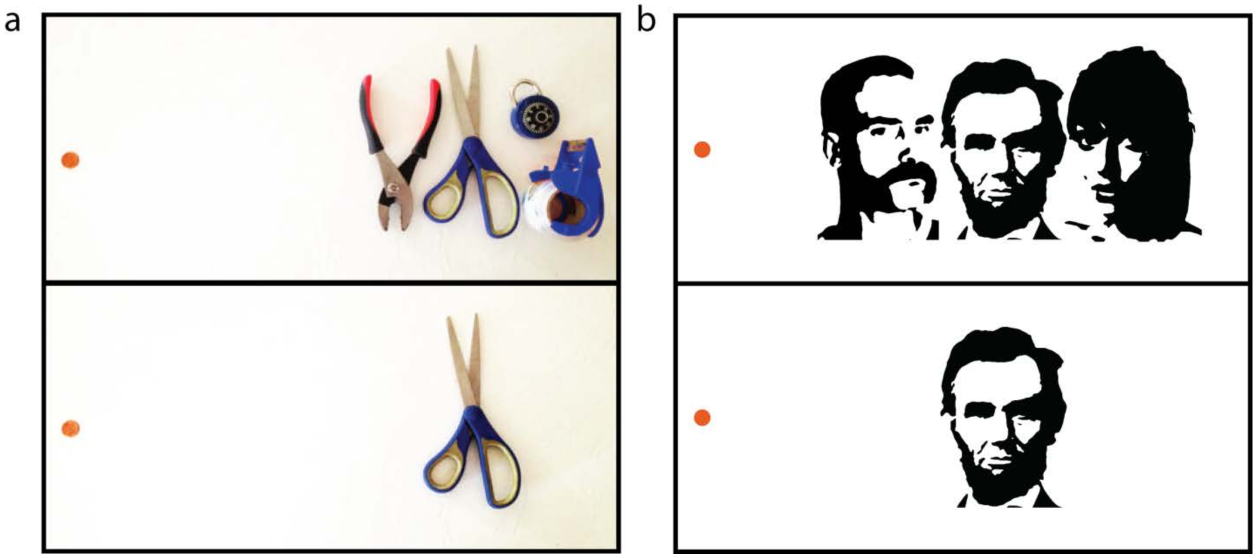

Figure 2.1: Visual crowding. **a)** When fixating the penny on the left in the upper panel, the center object on the right is difficult to identify, although it is clear that something is present in the center. In the lower panel, in the absence of surrounding clutter, identifying the same object at the same eccentricity is much easier. Crowding impairs the ability to recognize (but not the ability to detect) objects amidst visual clutter. **b)** Faces crowd each other. While fixating the orange dot in the upper panel, it is relatively difficult to recognize the identity of the central face due to crowding from the flanking faces. In the lower panel, in the absence of flanking faces, the central face is easier to identify. Crowding between faces is not simply due to crowding of low-level features such as edge information – inverting or scrambling the flanking faces, which preserves low-level features but disrupts holistic face information, reduces crowding between faces (Louie, Bressler et al. 2007; Farzin, Rivera et al. 2009; Fischer and Whitney 2011).

Yet there is another sense in which our visual experience in the periphery is misleading: the experience of crowding seems to imply that the brain simply lacks the bandwidth to represent individual objects outside of those that we scrutinize at the fovea; indeed, nearly all current models of crowding posit that the experience of crowding reflects an underlying irreversible loss of information due to a visual processing bottleneck (He, Cavanagh et al. 1996; Levi 2008; Pelli

2008; Balas, Nakano et al. 2009; Greenwood, Bex et al. 2009; Freeman and Simoncelli 2011; Nandy and Tjan 2012). However, emerging research shows that much more information survives in the periphery than previously thought, albeit sometimes outside the reach of conscious awareness. One clue comes from the fact that we can readily recognize objects that require configural processing, such as faces, when we see them in the periphery (McKone 2004; Louie, Bressler et al. 2007), despite the fact that the features of a face in the periphery crowd each other (Martelli, Majaj et al. 2005). How does the brain reconstruct the arrangement of the features of a face after those features have been jumbled together? That crowding happens at many different representational scales, occurring between basic features (Levi 2008), object parts (Martelli, Majaj et al. 2005), and whole objects (Louie, Bressler et al. 2007; Farzin, Rivera et al. 2009; Wallace and Tjan 2011), is paradoxical if crowding at early stages of visual processing destroys the information required by higher-level stages.

We recently directly tested the degree to which object-level information can survive crowding for use in subsequent visual processing (Fischer and Whitney 2011). We presented sets of faces in the periphery and asked observers to report either the expression of an individual (crowded) face from the set, or the average expression of the set as a whole. We found that even for sets of faces where observers were at chance in discriminating the expression of the crowded face, that particular face contributed with high precision to the perceived average of the set, an effect that cannot be explained by a contribution of low-level features alone. Individual object information is not lost amid the clutter in the crowded periphery, it is simply inaccessible to perception. In support of these findings, another recent study found that illusory contour formation, a process that relies on the configuration of the inducer stimuli, can also survive crowding of the individual inducers (Lau and Cheung 2012) (but see (Banno and Saiki 2012) for data suggesting that size information does not survive crowding). Further, crowded objects can unconsciously influence behavior by priming subsequent responses (Faivre and Kouider 2011; Yeh, He et al. 2012) and biasing preferences (Kouider, Berthet et al. 2011).

Thus, a satisfactory theory of crowding must account for not only for the perceptual degradation that crowding produces, but also for how certain information survives crowding and can contribute to downstream processes. The most prominent current models of crowding posit that crowding results from excessive integration of information appearing in the periphery, due to the number of neurons representing a given location in space (Pelli 2008; Pelli and Tillman 2008), lateral connections shaped by image statistics during development (Nandy and Tjan 2012), or the resolution of visual attention (He, Cavanagh et al. 1996). Some over-integration models can successfully account for most or all of the classical properties of crowding, but all posit information loss due to a resolution bottleneck, and thus cannot explain how crowded visual features or objects can be available with high fidelity to downstream processes. Another related model of crowding, the positional averaging model (Greenwood, Bex et al. 2009), posits that crowding results from pooling position information to reduce positional uncertainty. Positional averaging may also account for object-level crowding (Dakin, Cass et al. 2010), but it still posits information loss, and cannot account for how holistic object information survives crowding and influences ensemble perception (Fischer and Whitney 2011). Thus, while the general idea of involuntary pooling captures many aspects of crowding and likely plays a role, overintegration is not the whole story. Other models of crowding, including substitution (Wolford 1975; Chastain 1982) and contrast-gain or masking based models (Krumhansl and Thomas 1977; Chastain 1981; Petrov and Popple 2007) are not more successful; they similarly require that information about crowded objects is lost or substantively modified, a prediction that has been overturned (Faivre and Kouider 2011; Fischer and Whitney 2011; Kouider, Berthet et al. 2011; Lau and Cheung 2012; Yeh, He et al. 2012).

Here we propose a new model of visual crowding, the *hierarchical sparse selection (HSS) model*, in which unconscious object processing continues unencumbered by clutter in the scene. Our model accounts for the known characteristics of crowding, and generates several predictions for future tests (Box 1).

Box 1: HSS Model Predictions

The HSS model makes a number of concrete predictions at both the behavioral and neural levels for future testing:

1) The HSS model predicts that crowded stimuli are represented robustly in the brain even though they are blocked from conscious individuation. Thus, it should be possible with both fMRI pattern analysis and neurophysiological recordings to find precise representations of crowded objects in the brain.

2) The HSS model predicts that the critical spacing of crowding is different for different stimulus categories (e.g., gratings, faces, bodies, objects, etc.) because crowding is a function of receptive field size within the cortical map in which the stimulus is represented. There is already some evidence that critical spacing differs across stimulus categories (see (Whitney and Levi 2011) for a review), but the spatial extent of crowding has not yet been precisely characterized for a wide variety of stimuli, nor has there been a test of the relationship between receptive field size and the extent of crowding across stimulus categories.

3) In the HSS model, precise information about crowded objects persists in the visual processing stream despite the perceptual experience of crowding. Thus, information about crowded targets may be available to other processes in addition to ensemble perception and priming. For example, action may not suffer from crowding as much as perception (Bulakowski, Post et al. 2009).

4) A prediction of the HSS model is that with extensive experience viewing a particular stimulus category at a particular position in the visual field, it may be possible to reduce crowding through training. If information about a crowded target is present but requires fine-tuned connections to decode, it may be possible to train up the required connections. However, such training should not transfer to other sufficiently different stimulus categories even at the same spatial location because crowding depends on connections to the particular map that the stimuli are represented in. There is indeed evidence that training can reduce the strength and extent of crowding (Wolford, Marchak et al. 1988; Chung 2007; Hussain, Webb et al. 2012), but the specificity of the reduced crowding to object category remains to be tested.

The Hierarchical Sparse Selection model of visual crowding

Our proposed model rests on two principles. First, large receptive fields or integration regions do not imply the loss of fine-scaled information. While it is true that the output of a single neuron with a large receptive field will carry highly integrated, spatially- and featurally-ambiguous information in the presence of visual clutter, a population of many such neurons can carry sufficient information to resolve details on a scale far smaller than the receptive field size. Indeed, the feature or object at a precise location can be isolated from amongst clutter by combining the outputs of many highly overlapping receptive fields, as has been described in detail in the ensemble- and coarse-coding literature (Eurich and Schwegler 1997; Pouget, Dayan et al. 2000; Purushothaman and Bradley 2004), and large receptive fields may in fact be a more efficient means of carrying fine spatial information than small receptive fields (Baldi and Heiligenberg 1988; Snippe and Koenderink 1992; Eurich and Schwegler 1997). Figure 2.2a depicts this concept: neurons tuned to facial features have receptive fields that cover many features at once for a face seen in the periphery. Each individual neuron signals ambiguous information about the features present at a given location, yet with a proper decoding scheme, a combination of the outputs of many neurons can resolve the feature present at a given location. Thus, object processing can proceed unencumbered by clutter given precise enough wiring from one stage to the next. This notion is consistent with the fact that higher-level visual areas that are closely tied to the perception of object identity and position (Williams, Dang et al. 2007; Fischer, Spotswood et al. 2011; Maus, Fischer et al. 2013) have large receptive fields even in central vision (Raiguel, Hulle et al. 1995; Amano, Wandell et al. 2009), yet we can resolve and identify closely spaced objects in central vision.

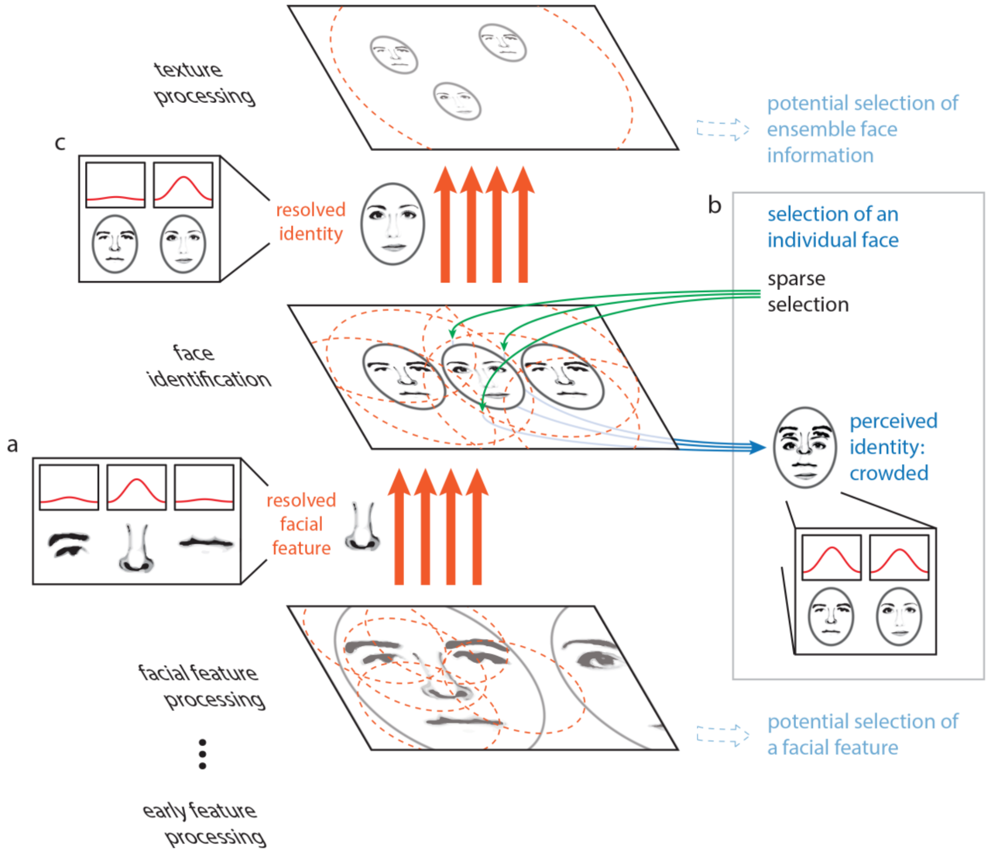
texture processing
potential selection of ensemble face information
c
resolved identity
b
selection of an individual face
sparse selection
face identification
a
perceived identity: crowded
resolved facial feature
facial feature processing
potential selection of a facial feature
early feature processing

Figure 2.2: The hierarchical sparse selection model of visual crowding. Unambiguous information about features or objects is passed between processing stages via an ensemble- or coarse-coding scheme, using a sufficient number of receptive fields with dense connections to avoid information loss through over-integration **(a).** Perceptually accessing an object from a given map relies on a sparse selection of a subset of the receptive fields or connections from that particular map only, resulting in the read-out of an ambiguous conscious percept **(b)**. Thus, an object that is perceptually crowded can nonetheless be passed, intact, to a subsequent texture processing stage **(c)**.

If high-fidelity information can be transmitted through a neural system with large receptive fields, why does crowding occur? The second component of our proposal is that while the feed-forward cortical object processing hierarchy possesses the copious and fine-tuned connections necessary to resolve the relevant features at every stage, the operation which "reads out" selected cells' outputs to conscious perception does not. Key to this notion is that within a coarse coding framework, unambiguous features and objects need not be explicitly represented by individual neurons at any stage of processing. Rather, information about an individual visual feature is encoded across a population of cells, and this information is decoded *between* stages of processing by the precise pattern of connections between neurons in one stage and the next. In the example in Figure 2.2, no single neuron at the facial feature processing stage unambiguously represents the nose, nor does any single neuron at the face identification stage. However, the presence of the nose at its precise location is conveyed between the facial feature processing and face identification stages by a specific and finely-tuned pattern of connections. If the selection of information from a given map for perception relies on connections to a subset of the units in this map (a "sparse selection"), there may be insufficient information available to unambiguously decode the selected feature (Fig. 3b). Thus, it is only possible for an observer to perceptually individuate an object when it can be unambiguously decoded from this limited sparse selection of the information in the neural population representing it, and this requires that the object is sufficiently separated from the clutter around it. However, object processing carries on regardless of whether this condition of sufficient separation is met (Figure 2.3c). It is important to differentiate sparse selection from the unrelated notion of sparse coding. Here, by "sparse selection" we mean capitalizing on information from a limited and sometimes insufficient number of units, whereas "sparse coding" refers to a sufficient coding scheme that favors having the smallest number of active units possible.

Why would perceptual selection only sample a subset of the relevant information available for resolving objects in the periphery? There are two likely reasons: First, attention must be highly flexible, able to select any feature from any position in the visual field. The number of connections required to perfectly sample information from any visual map in the brain is prohibitive. Putative attentional regions in the fronto-parietal network (Corbetta, Miezin et al. 1993; Buschman and Miller 2007) and the pulvinar (Petersen, Robinson et al. 1987; Fischer and Whitney 2012) possess widespread connectivity throughout the brain, but connect with only a subpopulation of the cells in a given brain region (Curcio and Harting 1978; Schall, Morel et al. 1995; Kaas and Lyon 2007). Second, the integrated ensemble information that we perceive in the periphery is useful for providing a rapid gist of the scene (Oliva 2005), as well as guiding attention and saccades (Torralba, Oliva et al. 2006). Trading off individual object information for ensemble representations in the periphery might be a benefit rather than a hindrance.

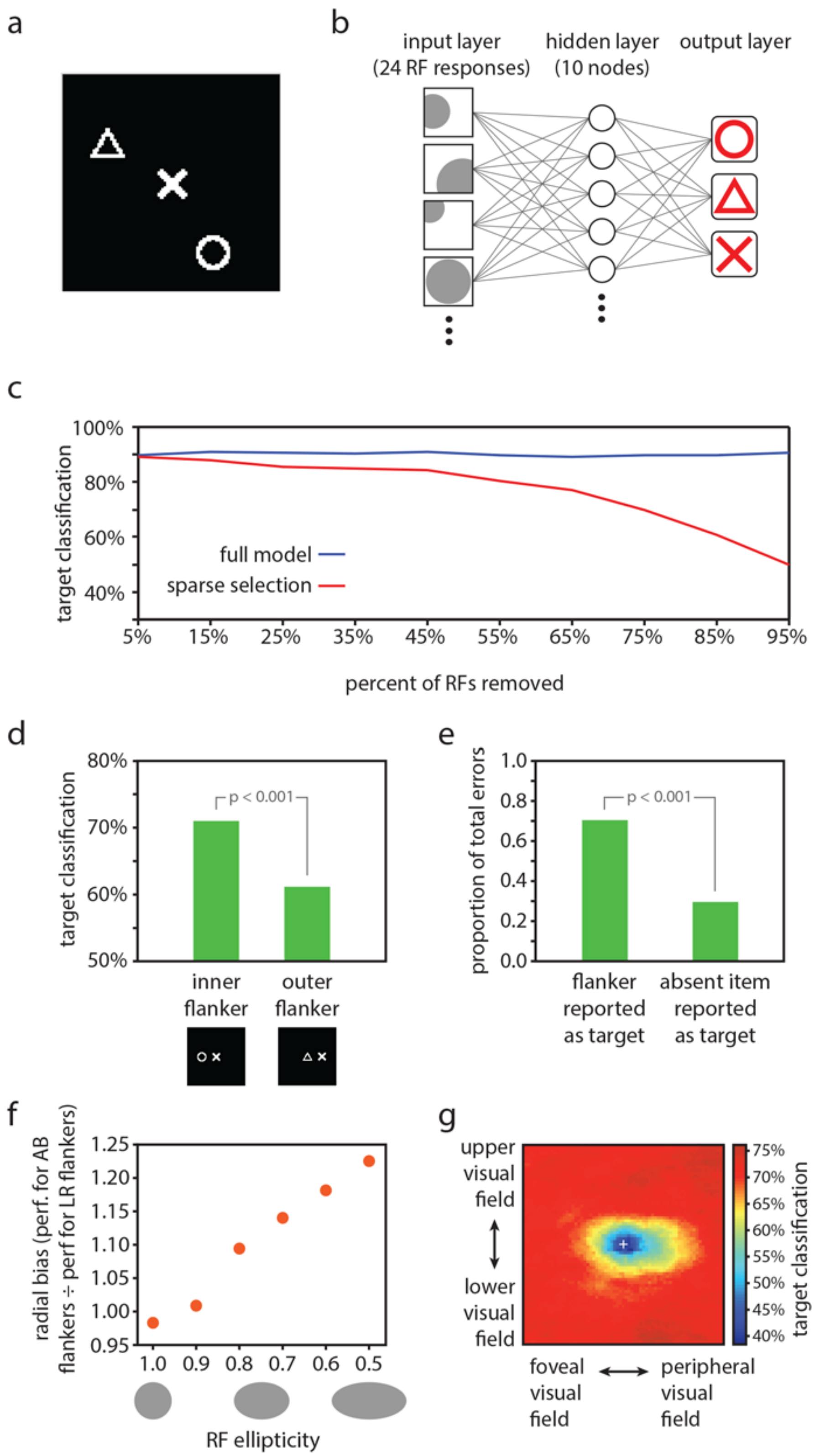

Figure 2.3: Computational model results. **a)** Example display for model training and testing. One symbol (the target item) was positioned in the center of the image, and two other randomly

selected symbols (flankers) were positioned in random locations elsewhere in the image. **b)** Schematic depiction of the neural network decoding model. Twenty-four receptive fields were tiled in random locations over the image. RFs had random tuning functions, but an equal number of RFs were optimally tuned to each of the three symbols (8 RFs optimally tuned to each of circles, triangles, and Xs). The model contained a ten unit hidden layer and a three unit output layer. Each node in the hidden layer was connected with each of the 24 RFs, and with each of the three nodes in the output layer. The three output layer nodes corresponded to the three symbols, and stimulus decoding was determined by taking the maximally responsive node in the output layer in a winner-take-all fashion. **c)** Comparison of performance for the full model vs. the sparse selection model. While the full model classified the target symbol with high accuracy (90.4% correct) despite the presence of flankers (blue data), classification performance of the sparse selection model decreased monotonically as more connections were removed (red data). Thus, although fine-scaled information can be decoded from a population of neurons with large receptive fields, robust decoding of crowded stimuli relies on a full sampling of the information present in the neural population. **d)** RF scaling with eccentricity yielded the inner-outer asymmetry characteristic of visual crowding. We tested the model performance using images with one flanker positioned either on the foveal side of the target symbol (here shown to the left of the target symbol) or on the eccentric side of the target symbol. Classification was significantly worse when an outer flanker was present vs. an inner flanker ($p < 0.001$), mirroring effects found in human performance (Bouma 1973; Petrov, Popple et al. 2007). **e)** Using the same test images as those used to test the inner-outer asymmetry, we found that when the model made a classification error, it was significantly more likely to report the flanker as the target than to report the symbol that was absent from the display ($p < 0.001$). This is consistent with human performance; observers frequently substitute a flanker for the target in a crowded display (Wolford 1975; Chastain 1982). **f)** We varied RF ellipticity in the model from 1.0 (circular) to 0.5 (half as tall as wide). For each value of RF ellipticity, we tested model performance with images in which flankers were positioned either to the left and right of the target (LR flankers; positioned along the radial dimension relative to the fovea) or above and below the target (AB flankers; positioned along the tangential dimension). We computed the ratio of performance when AB flankers were present to performance when LR flankers were present as a measure of radial bias in model performance. Radial bias increased monotonically as RFs became more elliptical, demonstrating that asymmetrically shaped RFs are a plausible source of the radial bias in crowding. However, the validity of the HSS model does not hinge on elliptical RFs. Other potential sources of the radial bias in crowding such as saccadic influences on the development of lateral connections (Nandy and Tjan 2012) could be similarly integrated into the HSS model. **g)** A visualization of the crowding zone based on the neural network model performance (the region of space around the target within which the presence of a flanker crowds the target). The white cross marks the location of the target; model performance was tested with a single flanker positioned at every possible location within the display. Here, we expanded the size of the display space by 50% relative to previous tests of model performance in order to visualize the full extent of the crowding zone. The visualized crowding zone is reminiscent of the elongated spatial interaction zones found by Toet and Levi (1992).

Importantly, our proposal is not that crowding results from the same limit on the spatial resolution of attention proposed by He and Cavanagh (He, Cavanagh et al. 1996). Their model asserts a smallest area of the visual field over which attention can operate; our model is about the sparsity of sampling within that region. Attentional sampling could be highly spatially specific, yet if attention samples from a limited number of receptive fields at the selected location, the object at that location cannot be resolved. Further, the sparse selection we propose can happen at any level of processing and is not limited by a single resolution of attention. It is the size of the receptive fields at a specific level of analysis, coupled with a sparse sampling of the information represented at that level of analysis for perceptual access, that causes crowding. The HSS model predicts that the critical spacing for crowding (the maximum distance at which a flanker can be positioned from the target and still cause crowding, as a function of eccentricity) differs for different stimulus classes (see Discussion), whereas the attentional resolution model predicts a single critical spacing for all stimuli based on the smallest possible attentional window at a given eccentricity.

Computational model

To test the outcome of drawing a sparse sample from coarse-coded visual information, we constructed a computational model aimed at decoding crowded visual features based on the output of randomly tiled receptive fields.

i) Model construction

The display images were 101 x 101 pixel images consisting of white symbols drawn on a black background (Fig. 3a). There were three possible symbols: a triangle, an X, and a circle, each drawn within a 20 x 20 pixel area in the images. In all display images, one symbol was placed at the center of the image; this center symbol was the crowded item that the model aimed to decode. In training images, two additional random symbols (flankers) were placed at random locations within the image; the training set comprised 120 such images – 40 images with a triangle at the center, 40 images with an X at the center, and 40 images with a circle at the center. Model testing was conducted on an independent set of 60 images constructed in the same fashion for basic model testing or with the flankers placed at specific locations for testing of asymmetries and substitution errors (described below).

The model consisted of receptive fields tiled over the image space (the input layer) whose outputs were fed into a neural network with one ten unit hidden layer and a 3 unit output layer (Fig. 3b). On each iteration of model training and testing, we tiled 24 receptive fields over the image space in random locations. Receptive fields had a mean diameter of 50 pixels. The left side of the image was treated as being more foveal and the right side of the image more eccentric, such that the image represented a patch of the right visual field. Receptive field size scaled linearly with eccentricity with a slope of 0.7, consistent with the scaling in extrastriate object-selective cortical regions (Amano, Wandell et al. 2009). Each receptive field was preferentially tuned to one of the three symbols but responded to some degree to each of the symbols. The response of a receptive field was computed by convolving a filter (a 20 x 20 image of the symbol that the RF was maximally tuned to) over the entire image and then taking the maximum of the convolution output within the region of the display image that the RF covered. Thus, when the optimal stimulus was present anywhere within an RF, the RF response was 1.0; if the preferred stimulus was partially within the receptive field or a nonpreferred stimulus fell within the receptive field, the response was less than 1 but greater than zero. We applied a rectification that mapped negative convolution values (possible if two stimuli fell close together within the RF) to zero. If no stimulus fell within a receptive field, its response was zero.

The set of 24 receptive fields comprised the input layer to the neural network; each RF had a connection to each of ten units in the hidden layer, and each unit in the hidden layer had a connection to each of three units in the output layer (Fig. 3b). The three output layer units corresponded to the three stimulus categories; stimulus decoding was determined in a winner-take-all fashion on the three output units. Training of the model weights was conducted with scaled conjugate gradient backpropagation implemented with the Matlab Neural Network Toolbox (MathWorks, Natick, MA). Model performance was then taken as the proportion of 60 independent test images correctly classified by the model. We conducted 1,000 iterations of model training and testing, randomizing the stimuli, RF locations, and RF tuning on each iteration, and we report the average model performance across all iterations. To test the significance of the model performance, we generated an empirical chance distribution by shuffling the stimulus labels prior to model training, then testing on an independent set of images with the correct labels. Repeating this shuffling procedure 1000 times produced a distribution of performance estimates that would be expected by chance; the significance of the model performance was taken as the proportion of the chance distribution that was larger than the actual estimated model performance.

To test whether the same model predicts crowding at the fovea, we adjusted the overall display size to 61x61 pixels from 101x101 pixels to keep target, flankers, and receptive fields within a smaller eccentricity range and closer to the fovea. The same three images were used (white circle, X, and triangle on black background) at the same sizes as before (20x20 pixels each). The target was presented in the center of the display image with two flankers randomly placed in nonoverlapping positions. The number of receptive fields in the model was increased from 24 to 45. This increase combined with the reduction in overall display image size lead to an increase in RF density (ratio of number of RFs to pixel area) by a factor of 5, consistent with an estimate of cortical magnification from V1 (Sereno, Dale et al. 1995; Engel, Glover et al. 1997; Qiu, Rosenau et al. 2006) assuming target eccentricity of 5 degrees in the previous model and 1 degree or less in the foveal model. This is a conservative estimate because cortical magnification is greater in extrastriate visual cortex than in V1 (Harvey and Dumoulin 2011), and object crowding likely occurs beyond V1 (Whitney and Levi 2011; Farzin and Rivera 2009). The remainder of the model was left unchanged: we used 10 hidden units, 120 training images, and 60 test images to run 1000 iterations of the model.

Finally, in order to further illustrate and clarify the hierarchical nature of the model, we present simulations of performance on two additional tasks, identifying either features or an object constructed from those features, using the same feature-tuned receptive fields in the input layer. In these simulations, there were two possible tunings for receptive fields, a horizontal line and a vertical line. The display images were again 101x101 pixel images with a target at the center. For the feature task, the target and flankers were either a horizontal or vertical lines. For the object task, the receptive field tuning remained the same, but the target and flankers consisted of “tumbling Ts”: the letter T oriented in one of the four cardinal directions. The size of the receptive fields was reduced to an average of 20 pixels diameter and the number of receptive fields was increased to 48, modeling a region with selectivity for lower-level features. All other aspects of the model were identical to the original implementation and we tested the model by performing 1000 iterations with randomized target and flanker identities, flanker locations, and receptive field locations within the 101x101 display image.

ii) Model performance

Target shape decoding performance was 90.4% correct, significantly greater than chance (chance performance = 33.3% correct; $p < 0.001$). This result establishes that target identity in a cluttered array can be resolved from the pooled output of a population of RFs, even when no individual RF is small enough to encompass the target alone. To test the effect of sparse sampling from the simulated neural population, we repeated the above analysis, this time removing a portion of the receptive fields from the network and then retraining (assigning new connection weights) after the removal of units and prior to testing. This procedure simulates the case where decoding of stimuli for conscious perception relies on a network of connections entirely distinct from that of feed-forward processing, connected to a sparsely selected subset of units. The results of this analysis are shown in Figure 2.3c: reducing the number of units sampled for the readout of the crowded central target led to a monotonic decrease in model performance, with performance dropping to 90% of the full model performance when 85% percent of the input units were sampled. Removing a portion of the receptive fields from the trained network without retraining prior to testing (simulating the case where attentional selection taps into the same network that robustly represents the target identity, but only has access to a subset of the units in the network) produced a comparable pattern of results. Similarly, removing individual connections rather than entire RF units from the model also resulted in a monotonic decrease in performance, though at a slower rate than removing entire receptive fields. The principle of "sparse selection" therefore holds irrespective of whether it is entire units or individual connections between units that are selected. In short, decoding target identity from a population of cells requires connections with a sufficient proportion of the cells to resolve those stimuli that are spaced closer together than the size of a receptive field.

We next asked if model performance followed the well-established property of inner-outer asymmetry: a flanker presented in a more eccentric location relative to the target produces stronger crowding than a flanker presented at the same distance from the target but in a more foveal position (Bouma 1973; Petrov, Popple et al. 2007). To test for an inner-outer asymmetry, we trained the model in the same fashion as above, but tested on images with just one flanker, positioned either 25 pixels to the left or 25 pixels to the right of the target. In this case the flanker was not allowed to be the same symbol as the target; thus, there were twelve total images in the test set. The sparse selection model for this and subsequent tests was generated by dropping a random selection of 50% of the RFs in the full model post-training. A comparison of model performance for test images where the flanker was more foveal than the target (positioned to the left) vs. the images where the flanker was more eccentric revealed an asymmetry in line with psychophysical results: the presence of an eccentric flanker yielded significantly worse model performance ($p < 0.001$; Fig. 3d). This asymmetry was absent without sparse selection – the inner/outer asymmetry emerges from the model as a result of the interaction between receptive field eccentricity scaling and sparse selection.

Another well-established aspect of crowding is that when observers make errors in reporting a crowded target, they report a flanker rather than another potential symbol with above-chance frequency (substitution errors; (Wolford 1975; Chastain 1982)). Using the same set of test images as described above for testing the inner-outer asymmetry, we asked whether the model more commonly reported the flanker, rather than the third symbol which was not present in the display, when it made an error. This was in fact the case: 70.4% of errors arose from reporting the flanker as the target, rather than reporting the symbol that was not present (Fig. 3e).

In behavioral tests, flankers positioned radially in relation to the target (eg, to the left and right of the target for a target appearing on the horizontal meridian) crowd more strongly than flankers positioned tangentially (above and below the target in the same example), an effect known as a radial bias (Toet and Levi 1992). A simple addition to our model could account for the radial bias in crowding: if receptive fields are elliptical rather than circular (Motter 2009), elongated in the radial direction, a radial bias emerges in the model performance. We tested this effect by using test images with two flankers either 25 pixels to the left and right of the target or 25 pixels above and below the target. We then varied the ellipticity of the receptive fields in the model from zero (perfectly circular) to ½ (half as large in the vertical direction as in the horizontal direction). The relative performance for test images with left/right flankers vs. images with up/down flankers decreased monotonically with increasing RF ellipticity. That is, the radial bias in model performance increased with more elliptical RFs, and was significant (a significant departure ($p < 0.05$) from a left/right vs. upper/lower performance ratio of 1, which reflects no bias) with ellipticity values of 0.8 or smaller (Fig. 3f).

There is strong evidence for elliptical receptive fields throughout the visual processing stream in mammals, for example in V4 of rhesus monkeys (Motter 2009), in macaque ventral visual areas (Op De Beeck and Vogels 2000, Pigarev et al. 2002), in areas 7, 21a, and claustrum of cats (Sherk and Levay 1981, Rodinova et al. 2004) and in RF subregions in mouse visual cortex (Smith and Häusser 2010). As such, it is important to incorporate elliptical receptive fields in a computational model of crowding in ventral cortical regions. Ellipticity is one possible explanation for the radial bias in crowding , and it would dovetail with the aforementioned neurophysiological literature. However, there are other potential contributors to the radial bias in crowding such as saccadic influences on the development of lateral connections (Nandy and Tjan 2012) that could be similarly integrated into the HSS model. Even without elliptical receptive fields, cortical magnification factor in the random placement of the RFs and eccentricity-dependent size scaling introduced some radial bias into our model. Our model does not hinge on any particular mechanism for the production of a radial bias; rather, the HSS model can be thought of as a module that can be added to many current models of crowding in order to extend them to account for how high fidelity information can survive crowding.

Next, we generated a visualization of the spatial extent of crowding produced by the HSS model (Fig. 3g). We used training and test images that were 150% of the size used in previous model testing (now 151 x 151 pixels); symbols were still 20 x 20 pixels. To accommodate the larger display image space, we increased the number of RFs in the model to 48, and the number of training images to 240. The ellipticity of RFs in the model was set to 0.5. On each of 100 iterations, we trained the model using the 240 training images (each had a target at the center of the image and two randomly positioned flankers), and then tested the model performance on a series of test images in which a flanker was positioned at every possible location in the display image. For each possible flanker location, there were six test images corresponding to all pairings of one symbol type as the target and a different symbol type as the flanker. Within a given flanker location, overall model performance was the % of the 6 test images correctly classified. In Figure 2.3g, the color at a given location in the image corresponds to the model performance when a flanker was positioned at that location and a target was positioned at the center of the image. The performance shown in Figure 2.3g is average performance over 100 iterations. The resulting visualized "crowding zone" is reminiscent of the elongated spatial interaction zones found by Toet and Levi (1992), and additionally shows an inner/outer asymmetry: the region within which a flanker degrades performance extends further into the periphery than toward the fovea.

Evidence for whether crowding occurs in central vision is mixed (Levi, 2008), but crowding is generally thought to be at least weaker near the fovea than in the periphery. Our foveal model (Fig 4a) with a modest increase in RF density and a bias toward locating RFs at lower eccentricities in accordance with the V1 cortical magnification factor (Sereno, Dale et al. 1995; Engel, Glover et al. 1997; Qiu, Rosenau et al. 2006) showed higher overall target identification performance, correctly identifying a target in 98.18% of trials, significantly greater than chance (chance performance = 33.3% correct; $p < 0.001$). Furthermore, performance in the foveal model required removing 75% of the RF units to reach 90% of the full model performance, 20% more than the peripheral model and equivalent to a 44% reduction of the RFs remaining in the peripheral model. We do not, however, want to stress too strongly the specific values we obtain. The parameters used here reflect extrapolations of cortical magnification and receptive field scaling into the most foveal portion of the human visual field, which affect the performance of the model. Rather, the results should be taken to qualitatively show that increased density of receptive fields and reduction of the size of the receptive fields could explain why sparse selection at the fovea would not result in crowding or would cause much weaker crowding than in the periphery.

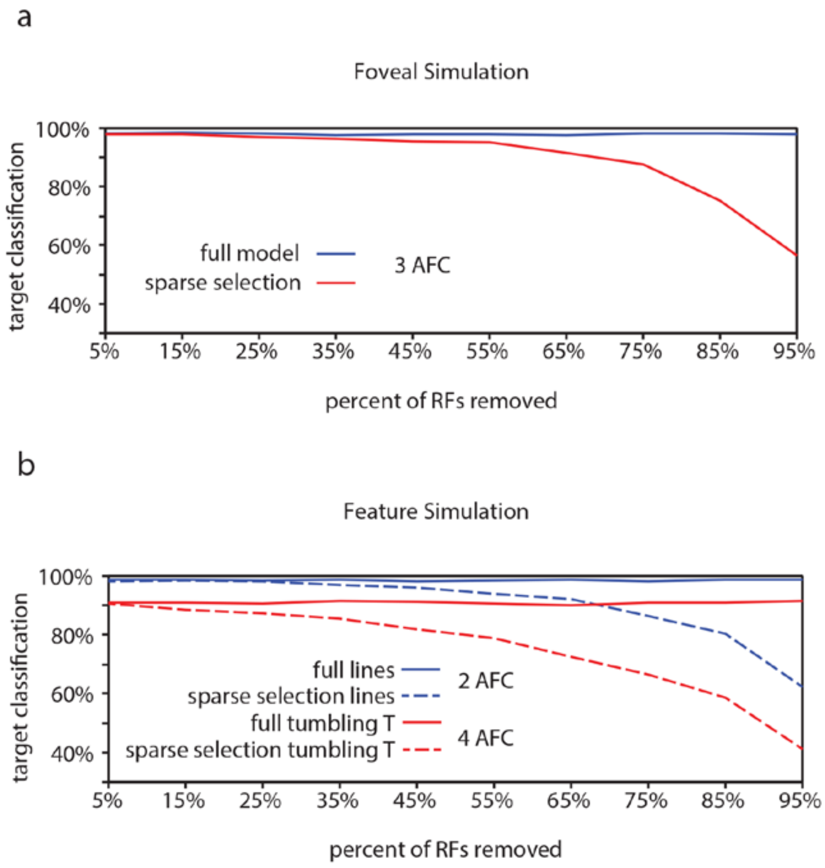

Figure 2.4: Computation model results. **a)** Comparison of performance for the full model vs. the sparse selection model at the fovea. Increased receptive field density yields higher full model performance than in the periphery (98.18%). Sparse selection of receptive field units leads to a much slower performance decrease demonstrating reduced crowding in the fovea. Increased receptive field density in foveal regions can reduce or eliminate the effect of sparse selection. **b)** Demonstration that feature-tuned receptive fields can perform well on both a feature discrimination task (98.73%) and an object discrimination task (91.06%) in the presence of flankers (solid lines). Sparse selection of feature-tuned receptive fields leads to a monotonic decrease in performance and crowding of features (dashed blue), showing that the model predicts crowding at any behaviorally relevant level of processing. Sparse selection occurs at only one level, avoiding the degradation of object level encoding that would occur if object representation depended on the sparse selection of features (shown in dashed red).

Finally, to demonstrate the hierarchical aspect of the model, we conducted a simulation of crowding performance using feature-tuned receptive fields, as opposed to objects or letters. In order to show that the model predicts crowding for features as well as objects, we first trained and tested the model with target and flankers that were horizontal and vertical lines (Fig 4b). Overall model performance was 98.73%. The model dropped below 90% of full model performance (88.86%) when 75% of the receptive fields were removed before retraining, indicating that crowding would occur in the identification of horizontal and vertical lines, if there were sparse selection of feature-level information, a simple task and only 2AFC as opposed to the 3AFC tasks in previous simulations. This demonstrates the hierarchical aspect of the model: the model can account for crowding of both features and whole objects when it is applied at any behaviorally relevant level.

The HSS model states that attention sparsely selects from the behaviorally relevant level of the visual hierarchy (Fig 2b), not that there is a cumulative effect of sparse selection at each level of the hierarchy. To show why, we trained the full model of this same network with feature detector receptive fields to identify “tumbling Ts” at a surprisingly high 91.06% correct performance (Fig 4b). This is a 4AFC task where every target and flanker contains both of the possible features that any given receptive field is tuned to and only relative location information is useful for the task. At 75% removal of feature tuned receptive fields, enough to cause crowding of features, “tumbling T” performance dropped to 66.6% correct. This scenario shows what would happen if degraded feature information was passed forward to subsequent visual processing stages – object-level information would be severely degraded. This contradicts many studies that have demonstrated that object level information gets through the bottleneck of crowding (Fischer and Whitney 2011). That is, if sparse selection occurred cumulatively at each level in the hierarchy (which is not what we are proposing), it would suffer from the same weaknesses as other crowding models: it could not account for the preservation of object information evidenced by object ensembles, priming and other effects (Faivre and Kouider 2011; Kouider, Berthet et al. 2011; Lau and Cheung 2012; Yeh, He et al. 2012)

Because the HSS model of crowding posits that sparse selection occurs only at the behaviorally relevant level of representation (selection occurs at the feature level of representation when the task is to identify a crowded feature and at the object level when the task is to identify the object), the object representation is preserved in the full feed-forward hierarchy.

Crowding can still occur at either level through a sparse selection of either feature or object level information for conscious awareness.

The model performance demonstrates that, in principle, the hierarchical sparse selection model can give rise to the known properties of visual crowding while supporting the transmission of high precision information within the cortical object processing hierarchy. This computational model is not intended to provide quantitative predictions about the number of neurons required in a coarse-coding framework or the number of neurons sampled by attention, but rather to provide a conceptual verification that: 1) fine-scaled information can be decoded from a population of neurons with large receptive fields, 2) a sparse selection from a neural population with large receptive fields results in rapid degradation of target identification and flanker-target confusions in the periphery but not in the fovea, 3) sparse selection at the behaviorally relevant level of processing nonetheless leaves high-fidelity stimulus information intact in the feed-forward visual processing stream, and 4) properties of receptive field scaling (in this case, larger receptive fields in more peripheral locations) can give rise to the asymmetries that are diagnostic of crowding.

Discussion

The hierarchical sparse selection model accounts for both the broad array of previously known characteristics of crowding and for recent findings that information can survive crowding, influencing ensemble perception (Fischer and Whitney 2011; Lau and Cheung 2012), priming behavior (Faivre and Kouider 2011; Yeh, He et al. 2012), and biasing preferences (Kouider, Berthet et al. 2011). The computational implementation of the HSS model described above deals with the simple case of decoding target identity from a small, discrete set of stimuli. The computational model itself is not intended to provide an exhaustive account of how sparse selection leads to crowding, but rather to provide a proof of concept that simply reducing the amount of information sampled for perceptual readout at any particular level of analysis gives rise to many of the known characteristics of crowding.

A hallmark of the HSS model is that it posits that crowding occurs between stimuli that are represented in the same cortical maps but not between stimuli that are represented in distinct maps (here, by “map” we mean an organized representation of visual space and/or basis dimensions *within* an object category). This feature of the HSS model accounts for why flankers of a different object category than the target are not effective crowders (Louie, Bressler et al. 2007; Farzin, Rivera et al. 2009). Since categorically different objects and features are coded in separate maps in the cortex (Op de Beeck, Haushofer et al. 2008), a target will be isolated in its cortical map and thus recognizable if the surrounding flankers are sufficiently different to be represented in a different cortical region. Likewise, this feature of the HSS model explains how grouping the flankers into an object can break down crowding (Livne and Sagi 2007; Saarela, Sayim et al. 2009) by causing the object formed by the distracters to be processed in a different cortical map than the target. Even when the target and flankers are of the same object category (e.g. a Gabor crowded by Gabors or a letter crowded by letters), a large difference between the target and flankers along dimensions such as color, orientation, and spatial frequency, and others can attenuate crowding (Andriessen and Bouma 1976; Nazir 1992; Kooi, Toet et al. 1994; Chung, Levi et al. 2001; Põder 2007). This could also be the result of compulsory grouping of the target and flankers into separate objects (Kooi, Toet et al. 1994), but another possibility exists: when the target and flankers differ markedly along one of these dimensions, even a sparse sample may be sufficient to successfully resolve the target from the flankers because of the large target/flanker signal difference. The fact that visual “pop-out” can alleviate crowding (Põder 2007) may simply be due to the target and flankers being different enough to resolve from the sparse sample of neural outputs available to conscious perception.

The HSS model also naturally accommodates the finding that a crowded target can produce adaptation and aftereffects despite being perceptually inaccessible (He, Cavanagh et al. 1996; Aghdaee 2005; Whitney 2005; Harp, Bressler et al. 2007; Bi, Cai et al. 2009): a crowded object fatigues the same population of cells that it would if it was presented in isolation – the perceptual phenomenon of crowding does not interfere with the underlying stimulus representation.

In sum, we present a novel model for visual crowding which posits that crowding occurs at multiple levels throughout the visual processing hierarchy, rather than at a single bottleneck. Counterintuitively, information about crowded objects is represented robustly in the brain, but may be inaccessible to conscious perception due to a sparse selection of information on which perception relies. The model is not intended to replace all existing models of crowding, but it could be a complementary component of any existing model; the HSS model does help account for many puzzling findings in the crowding literature that have otherwise gone unexplained.

Chapter 3: Lighting interpretation within scenes determines crowding

Introduction

Crowding is the impairment of object recognition in closely spaced clutter (Levi, 2008; Strasburger & Wade, 2015). Given its ubiquity in everyday life, it is considered a fundamental bottleneck of object recognition in visual scenes (Pelli & Tillman, 2008; Strasburger, Rentschler, & Jüttner, 2011; Whitney & Levi, 2011). Crowding impairs the recognition of objects that are well above acuity thresholds, jumbling their appearance into indiscriminable high contrast clutter. Crowding occurs over the vast majority of the visual field, where objects that can easily be identified in isolation become unrecognizable in the presence of surrounding flankers.

Crowding is characterized by three main spatial properties. First, the strength of crowding is eccentricity dependent and the region around a target in which flankers will interfere with recognition scales with the eccentricity of the target. This is known as Bouma's window (Bouma, 1970; Pelli & Tillman, 2008). Second, the radial-tangential anisotropy: there is stronger crowding by flankers placed along the same line from fixation than for flankers placed orthogonal to this line (Toet & Levi, 1992). Third, the inner-outer asymmetry: flankers placed at a greater eccentricity than the target tend to cause a larger crowding effect than flankers placed more foveally (Bex, Dakin, & Simmers, 2003; Bouma, 1973; Petrov & Meleshkevich, 2011). These asymmetries can be used to distinguish crowding from other phenomena such as visual masking (Petrov, Popple, & McKee, 2007).

Behavioral experiments have demonstrated that crowding occurs selectively at many levels along the visual hierarchy, from orientation to faces (Whitney & Levi, 2011). It occurs not only for single features and object parts (Andriessen & Bouma, 1976; Levi, Klein, & Aitsebaomo, 1985; Westheimer & Hauske, 1975), but also for whole objects (Wallace & Tjan, 2011). This can be observed in mid level visual stimuli, such as words composed of letters (Huckauf, Heller, & Nazir, 1999), and higher level stimuli, such as faces composed of facial features (Louie, Bressler, & Whitney, 2007; Sun & Balas, 2014). It even occurs selectively for two-tone Mooney faces, which are recognized holistically and are not defined by any set of particular features (Farzin, Rivera, & Whitney, 2009). Furthermore, there is evidence for crowding in the representation of heading direction of biological motion walkers that occurs only when these stimuli can be perceived as whole walkers (Ikeda & Watanabe, 2016; Ikeda, Watanabe, & Cavanagh, 2013).

While behavioral evidence shows that crowding is ubiquitous throughout the visual processing hierarchy, there is limited consensus on the neural locus of crowding. Crowding occurs dichoptically, with the target and flankers presented monocularly to separate eyes, indicating that crowding occurs in the cortex (Flom, Weymouth, & Kahneman, 1963). Indirect arguments have been made for the locus of crowding, such as using estimates of receptive field size from behavioral data (Tyler & Likova, 2007). These findings point to many different early to mid level visual areas, from V1 to V8, depending on the stimulus and task used. Most of these approaches implicitly assume that crowding occurs at a single stage in the visual processing hierarchy (Freeman & Simoncelli, 2011). Evidence for physiological markers of crowding has been found in V1, V2, and other regions beyond these early visual processing areas (Anderson, Dakin, Schwarzkopf, Rees, & Greenwood, 2012; Bi, Cai, and Fang, 2009; Millin, Arman, Chung, & Tjan, 2014).

There are several theories of crowding, including compulsory pooling and excessive integration, averaging, and contrast-gain/masking based models (Balas, Nakano, & Rosenholtz, 2009; Freeman & Simoncelli, 2011; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001; Pelli, Palomares, & Majaj, 2004; Wilkinson, Wilson, & Ellemberg, 1997). Most models of crowding assume either an early stage loss of information or a loss of information at each stage. However, all these theories have been strongly challenged by recent behavioral results. A few others may more easily accommodate top-down effects. One model proposes that crowding is due to a limited resolution of attention. Two models have been proposed that suggest crowding happens during the readout process from population responses and can address crowding at multiple levels of representation (Chaney, Fischer, & Whitney, 2013; Harrison & Bex, 2015). These models have several advantages over lossy early stage models. First, it is difficult to reconcile these early stage models with crowding of higher level categories like faces (Farzin et al., 2009; Louie et al., 2007). Second, despite information loss from crowding, some information survives crowding (Faivre, Berthet, & Kouider, 2012; Kouider, Berthet, & Faivre, 2011; Yeh, He, & Cavanagh, 2012). For example, crowded objects can still cause priming and aftereffects and they still contribute to ensemble percepts (Atas, Faivre, Timmermans, Cleeremans, & Kouider, 2014; Faivre & Kouider, 2011; He, Cavanagh, & Intriligator, 1996). Third, crowding was shown to be modulated by precueing attention to the target location (Yeshurun & Rashal, 2010). Fourth, influence of certain contextual features can break crowding. Crowding can be modulated by many other factors including similarity (Kooi, Toet, Tripathy, & Levi, 1994), perceived depth (Astle, Mcgovern, D.P. & Mcgraw, 2014), pop-out, and grouping (Herzog & Manassi, 2015; Herzog, Sayim, Chicherov, & Manassi, 2015). Similarity often depends on more than simple local edge information and pop-out is the result of top-down influence. Strict lossy feedforward models such as compulsory pooling are unable to explain contextual effects in crowding like grouping. In fact, it was found that grouping between target and flankers can strongly determine the strength of crowding (Livne & Sagi, 2007; M. Manassi, Hermens, Francis, & Herzog, 2015; M. Manassi, Lonchampt, Clarke, & Herzog, 2016; Mauro Manassi, Herzog, Sayim, & Herzog, 2012; Mauro Manassi, Sayim, & Herzog, 2013; Saarela, Sayim, Westheimer, & Herzog, 2009). Grouping often occurs preattentively and automatically (Erlikhman et al., 2013; Roelfsema, 2006), but necessarily requires higher level processing than simple edge detection. Nonetheless, grouping influences the strength of crowding in tasks that require only the most basic of feature comparisons like the Vernier discrimination task (Manassi et al., 2012; Sayim, Westheimer, & Herzog, 2010). Focusing on the order of functional cognitive processes, particularly those such as grouping which precede the information loss associated with crowding, can serve to highlight weaknesses in these models.

Similarly to grouping, there is evidence that the brain interprets and discounts lighting and shadows preattentively (Ramachandran, 1988). Therefore, we might expect that the computations associated with lighting and shadows occur prior to the deleterious influence of crowding, much in the same way as grouping. Ramachandran has shown that the brain defaults to a straightforward explanation of simple scenes favoring an interpretation of a single light source located above the scene. This lighting interpretation can generate popout leading to relatively flat search times for stimuli that are consistent with this lighting-from-above assumption. Further evidence in visual search has shown that the visual system rapidly interprets shadows as shadows and that this interpretation has a subsequent negative effect on the speed of processing for relatively fine discriminations of these shadows (Rensink & Cavanagh, 2004). Because crowding is a fundamental limit to our capacity for object recognition (Whitney & Levi, 2011), and since shadows and shading are ubiquitous in natural scenes, it is important to see whether the visual system is capable of analyzing and utilizing lighting information to avoid crowding from these additional contours or even to alleviate crowding in some circumstances. Various forms of specialized processing occur for lighting and shadows, including discounting the illuminant, preattentively discounting shadows, and perceiving shape-from-shading, amongst others, but it is unknown whether the loss of information in crowding functionally follows or precedes these processes.

The purpose of this paper is to examine whether crowding of orientation can be determined by the lighting interpretation within a scene. Rather than placing crowding at a specific physical location in the visual hierarchy, before or after lighting interpretation occurs, which may well be an ill posed problem to begin with, we test its place in a particular hierarchy of cognitive processes that contribute to perception. Focusing on this more abstract level of functional processes requires no assumptions about a singular neural locus for crowding, about the roles of feedforward and feedback connections, or about the parallel nature of visual processing, but instead emphasizes behavioral understanding of the crowding phenomenon that can also inform our physically based models.

We asked whether specialized processing for lighting information determines crowding. If the visual system computes and utilizes information about lighting direction prior to the deleterious effects of crowding, then shape-from-shading may be used to create conditions which predict a reduction in the strength of crowding. We compared performance between arrays consisting of a single Gabor target and surrounding checkerboard flankers, each superimposed on dots shaded in one of four possible directions, which differed only in the direction of the shading gradient (Experiment 1). To ensure that this result was not confounded with the relation of the lighting direction to the angle of the stimulus from the fovea, we repeated the experiment in the lower visual field (Experiment 2). Finally, we presented a single flanker on either side of the target to determine whether any inner/outer asymmetry was present and whether flanker-flanker grouping could be responsible for any observed effects (Experiment 3).

General Methods

Apparatus

Stimuli were generated on a Macintosh computer running PsychoPy2 and presented on a gamma-corrected CRT Sony Multiscan G500 monitor. The refresh rate of the display was 60 Hz and the resolution was 1024 x 768 pixels. Stimuli were viewed from a distance of 57 cm. Subjects used a keyboard for all responses (left-right arrow keys to indicate -45° and +45° orientation of the target Gabor). Five subjects participated in the experiment (3 male, 2 female). All had normal or corrected to normal acuity. All experimental procedures were approved by and conducted in accordance with the guidelines and regulations of the UC Berkeley Institutional Review Board. Participants were affiliates of UC Berkeley and provided informed consent in accordance with the IRB guidelines of the University of California at Berkeley.

Stimuli

The stimulus consisted of three major elements: the target Gabor patch, the surrounding checkerboard flankers, and the shaded dots that made up the background (Figure 3.1). The sinusoidal wave in the target Gabor had a spatial frequency of 3 cycles/degree and a phase shift of one half cycle. The Gaussian window had a standard deviation of 0.13°. The contrast of the target Gabor was 40% (Michelson contrast). The orientation of the target Gabor differed from trial-to-trial and was either -45° (tilted left) or +45° (tilted right) from vertical. The checkerboard flankers were made of the same two Gabors that were used as targets, superimposed with an increased contrast of 90% (Michelson contrast). The shaded dots each subtended 1° of visual angle and could have one of four possible shading directions. We refer to a light-dark gradient from left (or top) to right (or bottom) as "left shading" (or "top shading"). The shading luminance ranged from 0.36 cd/m^2 at the dark end to 150 cd/m^2 at the brightest end.

The stimulus configuration on each trial was an array consisting of a central Gabor patch and eight evenly spaced surrounding checkerboards, with each element (Gabor or checkerboard) superimposed on one of nine shaded circles (Figure 3.1). Each dot subtended 1 degree of visual angle and the superimposed Gabor patches subtended approximately 0.8° (6 standard deviations). In order to control for any possible configural, pooling or summary statistic effects, each checkerboard flanker contained both possible target orientations in the array and all possible arrays of target and flankers were displayed equally often. Target orientations of -45° and +45° were chosen so that the difference between the direction of the shading gradient on the background dot and the orientation of the target would be equal under all conditions.

The central Gabor patch was presented at 10° of eccentricity to either the left or right of fixation (fixed for each block) and the surrounding checkerboards were radially arranged and evenly spaced at positions subtending 45° increments at the target location, a total of 8 flanking checkerboards around the central Gabor target. The center-to-center spacing of the target Gabor and surrounding checkerboards varied randomly trial-by-trial from 1.2 to 3.6 degrees in steps of .3 degrees for a total of 8 possible separations, each appearing equally often.

Four different shading conditions were examined by randomly selecting the shading of the central dot under the superimposed target Gabor from one of the four cardinal directions on a trial-by-trial basis. The dot could be shaded up, down, left, or right and we adopt the convention of using "left" to refer to when the bright side of the dot is on the left and the dark side is on the right (see Figure 3.1A and 3.1B). The four conditions were balanced across the block of trials and the shading of the flanking eight dots was always uniform and rotated 180 degrees from the shading of the central dot. If the target was placed on a shaded left dot, the flankers were placed on shaded right dots. If the target was placed on a shaded up dot, the flankers were placed on shaded down dots, and so on. This rotation was held constant across all conditions to equate target-flanker orientation similarity across all four shading directions. In all four stimulus conditions, the difference between target and flanker was a reversal of the shading gradient direction. The two possible target orientations were chosen as +/-45° so that target orientation differed from the shading gradient direction by the same amount for all shading conditions; the difference between target orientation and the orientation of the flanker shading direction is always 45°. Therefore, the stimulus information available in any condition was identical and the only effects of the trial-to-trial manipulations were the changes to implied lighting direction and to the stimulus spacing.

Task and procedure

Subjects were seated 57 cm from the display and were instructed to maintain fixation on a centrally presented dot throughout the experiment. On each trial, the target Gabor was randomly presented in one of two possible orientations (-/+ 45°) and superimposed on one of four possible shaded dots, surrounded by 8 checkerboards superimposed on oppositely shaded dots. On each trial, the stimulus array was presented for 125 ms followed by a 750 ms interstimulus interval after each response. The observers' task was to report whether the target was oriented to the left or right of vertical in a 2AFC task by pressing either the left arrow to indicate -45° or the right arrow to indicate +45°. Each subject participated in 1260 trials over two blocks. Subjects performed the task in either the left or right visual field on each block and the order was counterbalanced across subjects. Therefore, subjects participated in a total of 40 trials for each condition/target-flanker spacing combination over the two blocks.

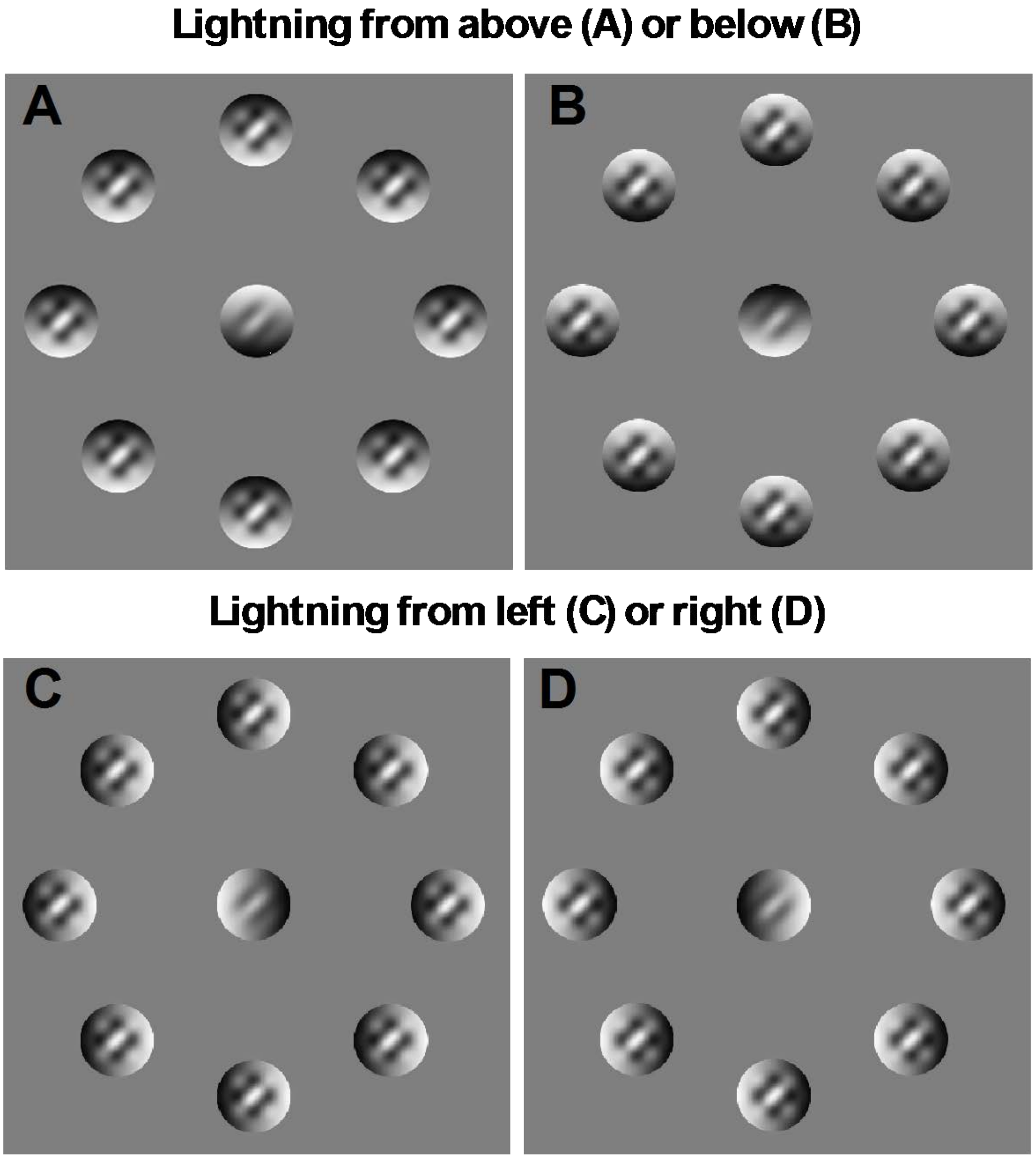
Lightning from above (A) or below (B)
A
B
Lightning from left (C) or right (D)
C
D

Figure 3.1: Stimuli for Experiment 1 and Experiment 2. The four possible stimulus configurations for Experiment 1 and 2 are shown here. A target Gabor at orientation -45 or 45 is superimposed on a shaded dot and surrounded by flanking checkerboards composed of two superimposed Gabors superimposed on dots with shading direction 180 rotated from the direction of the target dot. Target-flanker spacing was manipulated during experiments. A and B refer to Gabors with shading in the up and down direction. These conditions can be interpreted as "lit-from-above," conditions under which the target dot appears to pop-out from flankers due to shape-from-shading. C and D refer to the Gabors with shading in the left and right direction. These conditions can be interpreted as "lit-from-the-side," conditions under which target-flanker apparent similarity increases. Importantly, physical similarity between the stimuli is identical.

Stimuli were shown in the left/right visual field in Experiment 1 and in the lower visual field in Experiment 2.

Experiment 1 Intro

The purpose of this experiment was to directly manipulate the perceived lighting direction within a crowded array to examine its influence on crowding. If shading is computed before the strength of crowding has been fully determined, we expect to see changes in crowding strength due to perceived lighting direction. When the target and flankers are shaded in the up/down directions, the scene is consistent with a lighting-from-above interpretation, and pop-out occurs for the target due to perceived 3D shape. This effect is dramatically reduced in the left/right shaded dots. Therefore, if lighting interpretation precedes crowding, then crowding strength should be reduced for a target with opposite shading direction from the flankers in the up/down shading conditions compared to the crowding strength in the left/right shading conditions.

Experiment 1 Observers

Five adults participated in the experiment (4 male and 1 female). The observers ages ranged from 19 to 28. One observer was an author of the paper. The other 4 participants were naïve to the purpose of the study.

Experiment 1 Data Analysis

Subject accuracy as a function of spacing between the target and flankers was modeled as a logistic function using maximum likelihood parameter estimation with Psignifit 3.0 toolbox. The psychometric function took the form:

$$y\ = 0.5 + (1 - 0.5 - b) * 1/(1 + e^{\frac{x-c}{d}})$$

where y represents the probability of a correct response given a target-flanker separation of x, b is the lapse rate, restricted to the interval 0.00 to 0.05, c is the threshold of the function, and d is a slope parameter for the function.

Hypothesis testing was performed using a parametric bootstrap procedure as described in Efron & Tibshirani (1986) and Wichmann & Hill (2001). After fitting the above function to the data from each of the four experimental conditions (lighting directions) for each subject and verifying goodness of fit, 2000 new samples of sets of 1260 trials (the same number performed by observers) were generated from each fitted function. We then calculated a 95% confidence interval for the separation corresponding to 80% correct performance from the fits in this sample distribution. We considered two conditions within a single observer to be significantly different if the bootstrap samples from one condition yielded a target-flanker spacing threshold corresponding to 80% correct performance that was greater than the same value calculated from a bootstrap sample of the other condition on 95% or more of the bootstrap samples. This generates an empirical p-value for a one-way test from the bootstrap distribution.

Experiment 1 Results

The data from the first experiment are shown here in Figure 3.2. Performance in each of the four lighting directions is shown for each of the four subjects. Horizontal lines represent 95% confidence intervals for the 80% correct performance threshold for each condition.

Among the four shading conditions, only two shading conditions (shaded up and shaded down) are consistent with a lighting-from-above interpretation. The other two conditions (shaded left and shaded right) are consistent with a lighting-from-the-side interpretation. For each observer, there are six possible pair-wise comparisons across the four lighting directions: two comparing lighting-from-above to lighting-from-above or lighting-from-the-side to lighting-from-the-side, and four which compare a lighting-from-above condition to a lighting-from-the-side condition. We first present the four comparisons between lighting-from-above and lighting-from-the-side, as this is the main experimental manipulation of interest.

In order to determine whether a viable lighting-from-above interpretation of a crowded array reduced the strength of crowding as hypothesized, we looked at all of the comparisons of lighting-from-above conditions to lighting-from-the-side conditions. A significant result is indicated when crowding strength is reduced for the up/down shading conditions relative to the left/right shading conditions, resulting in increased subject performance. Across the five observers, there are 20 total comparisons of lighting-from-above to lighting-from-the-side conditions, 14 of which have 80% correct performance thresholds that differ significantly at the 95% confidence level (higher performance in scenes consistent with lighting-from-above relative to lighting-from-the-side). Another 3 were significant at the 90% confidence level. Two observers had psychometric functions in one condition that did not reach ceiling performance and added a substantial amount of noise.

We also looked at whether there was any significant difference between the two lighting-from-above conditions or between the two lighting-from-the-side conditions for each subject. Two subjects showed a significant difference between the lighting-from-the-side conditions, with higher performance in the shaded right condition than the shaded left condition. All other subjects showed no significant difference between these conditions. Two observers showed no significant difference between lighting-from-above conditions, but the three remaining observers did. However, observers were inconsistent in which condition performance was higher with one subject showing increased performance in the shaded up condition compared to shaded down and the two other subjects showing the opposite. The differences between these conditions were smaller than the differences in the lighting-from-above to lighting-from-the-side comparisons.

Since there was no reason to suspect any systematic differences within the lighting-from-above conditions or the lighting-from-the-side conditions, we also performed the same analysis averaging across like conditions. Unsurprisingly, averaging across the lighting-from-above conditions and the lighting-from-side conditions showed a significant difference for all four individual subjects with no bootstrap difference samples crossing 0 ($p < .0005$). These results are shown in Figure 3.2.

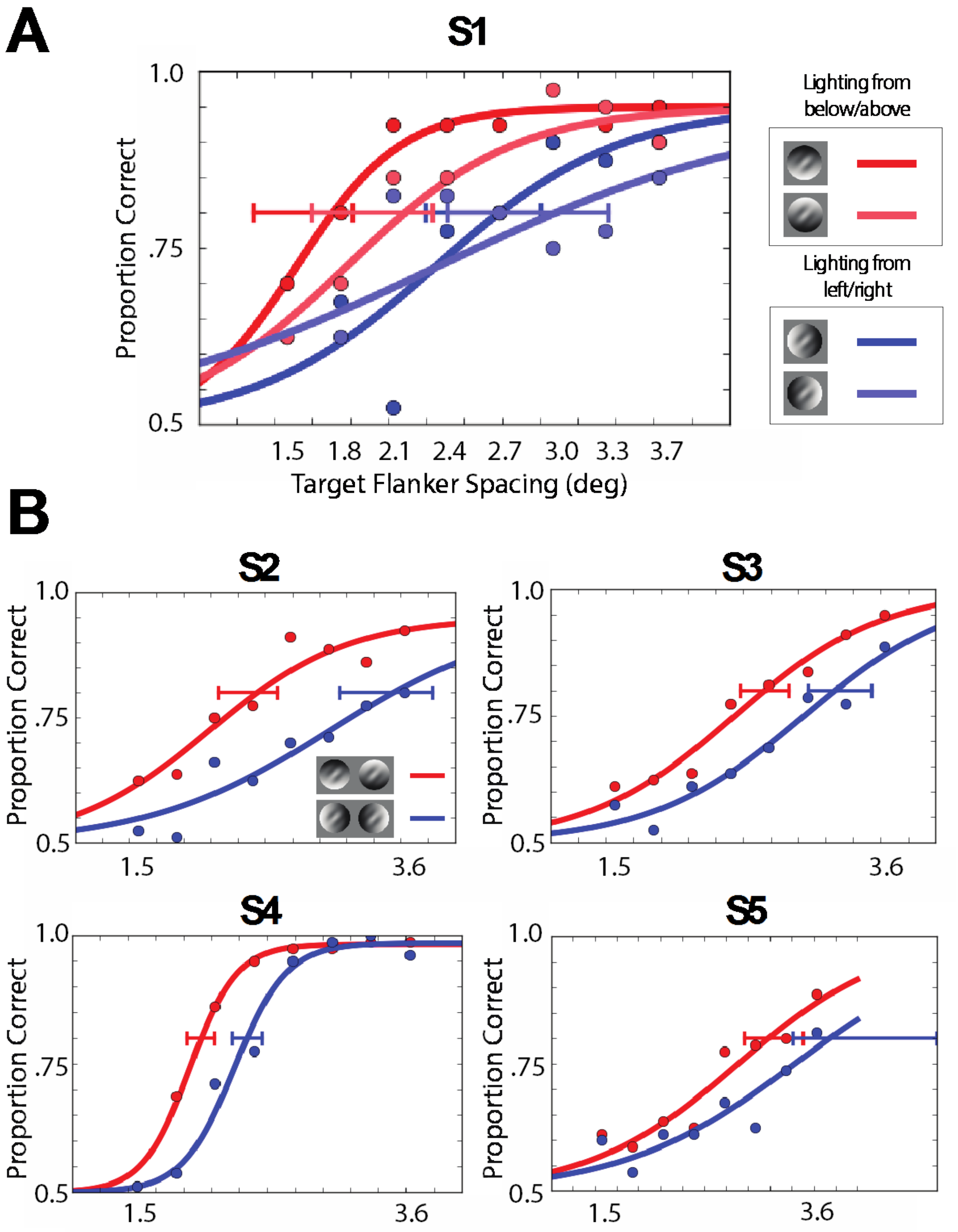

Figure 3.2: Results for Experiment 1. A) Target identification performance as a function of target-flanker spacing is shown for the four experimental conditions for one representative subject (S1). Error bars represent bootstrapped 95% confidence intervals of the 80% correct performance threshold. Other subjects showed similar patterns. We find increased performance under lighting-from-above conditions (up and down) than lighting-from-the-side (left and right). No systematic relationship was found between the two lighting-from-above and lighting from

below conditions, or between lighting from left and lighting from right conditions. B) Results for the other 4 subjects averaged within the lighting-from-above conditions (up and down collapsed) and lighting-from-the-side conditions (left and right collapsed) are shown. Subjects reliably showed increased performance in the lighting-from-above conditions compared to the lighting-from-the-side conditions. This indicates that lighting interpretation is able to modulate a 3D representation and influence the strength of crowding even for a low level orientation discrimination task.

Experiment 2 Intro

In Experiment 1, the two pairs of shading directions being compared also differed in whether their orientation was radial or tangential with respect to the fixation point. For the lighting-from-above condition, the gradient of shading is tangential to the horizontal meridian, and for the other lighting conditions, the gradient of shading lies along the horizontal meridian. Because flankers that are placed radially from a target are known to have stronger crowding effects on the target than tangentially placed flankers with the same spacing and characteristics (Toet & Levi, 1992), it is possible that this anisotropy resulted in the increase in performance in the lighting-from-above conditions we observed in Experiment 1. In order to eliminate this confound, the experiment was repeated with the entire stimulus array presented in the lower visual field.

Experiment 2 Observers

Five observers participated in this experiment (3 male, 2 female). One was an author of the study. The age range of subjects was 19 to 28.

Experiment 2 Method

Experiment 2 was identical to Experiment 1 in task and procedure with only two manipulations made to the stimulus. First, the entire stimulus array was moved to the lower visual field instead of the left or right visual field. Because pilot results revealed increased difficulty with the task in the lower visual field, possibly due to task familiarity in the left and right visual fields, the eccentricity of the target stimulus was reduced to 8°. The size of the stimuli, range of spacings within the array, contrast, and timing were otherwise unchanged. Subjects participated in 1260 trials for a total of 40 trials per lighting condition / target-flanker separation combination.
The statistical analyses performed for Experiment 2 were exactly the same as those explained in the results of Experiment 1. We generated a bootstrap distribution for the target-flanker spacing threshold that corresponded to 80% performance in the task for each condition for each subject. To compare two conditions, we calculate the difference of the bootstrap distributions for the two conditions and construct a confidence interval. Confidence intervals that do not include 0 indicate a significant difference between the means of the distributions.

Experiment 2 Results

Results from Experiment 2 are shown in Figure 3.3. Overall, the results of Experiment 2 are qualitatively similar to those of Experiment 1, replicating the main findings and indicating that the findings of Experiment 1 cannot be explained by an interaction between lighting direction and the direction of the target from fixation. We note that this interaction would predict a reversal of the main results of Experiment 1 if it were the explanation for our results. Again, across lighting conditions there are 4 comparisons per subject for a total of 20 comparisons,14 of which are significant at 95% confidence level. 18 out of 20 comparisons were significant at the 90% confidence level. There was no trend for comparisons within lighting-from-above or within lighting-from-the-side across subjects. Averaging the lighting-from-above and lighting-from-the-side conditions within subject and comparing yielded individual subject empirical p-values: $p < .0005$ (3 subjects), $p < .005$ (1 subject), $p < .01$ (1 subject). In all, Experiment 2 reproduces the main results of Experiment 1 in a different part of the visual field.

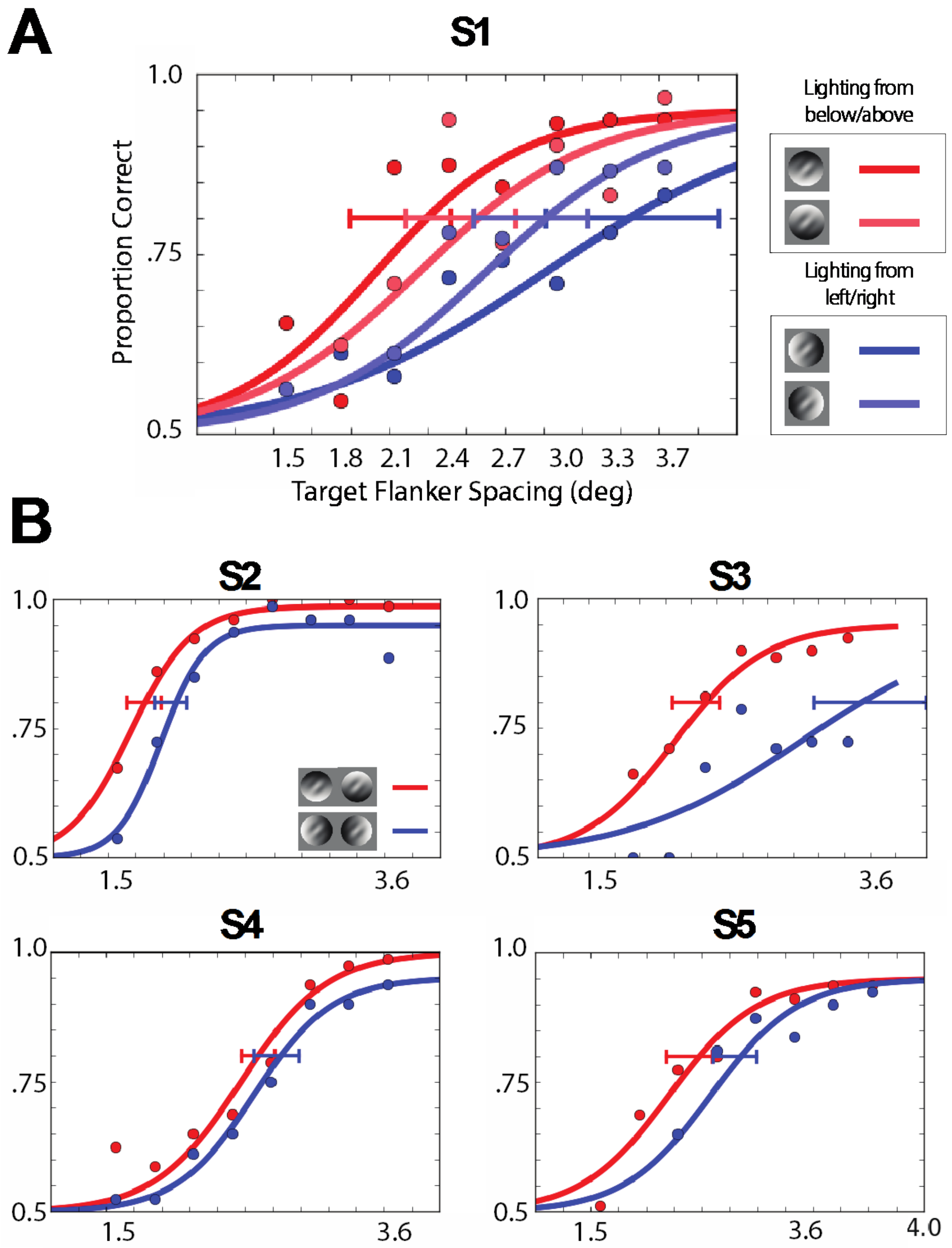

Figure 3.3: Results for Experiment 2. A) Target identification performance as a function of target-flanker spacing is shown for the four experimental conditions for one representative subject (S1). Error bars represent bootstrapped 95% confidence intervals of the 80% correct performance threshold. Other subjects showed similar patterns. We find increased performance under lighting-from-above conditions (up and down) than lighting-from-the-side (left and right). No systematic relationship was found between the two lighting from above and lighting from

below conditions, or between lighting from left and lighting from right conditions. B) Results for the other 4 subjects averaged within the lighting-from-above conditions (up and down collapsed) and lighting-from-the-side conditions (left and right collapsed) are shown. Subjects reliably showed increased performance in the lighting-from-above conditions compared to the lighting-from-the-side conditions. Since the stimuli were presented in the lower visual field, we rule out that the results of Experiment 1 are due to an interaction between the lighting directions or target directions and the direction of the target from fixation.

Experiment 3 Intro

Experiments 1 and 2 showed a shading-dependent modulation of crowding, suggesting that shading may be computed before crowding happens. However, in the previous experiments, all of the flanking dots had the same shading, which could have facilitated a flanker-flanker grouping effect and allowed for improved segregation of the target. Shaded dots which are interpretable as lit from above tend to group more readily than shaded dots which appear to be lit from the side (Kleffner & Ramachandran, 1992) and grouping of flankers is known to influence the strength of crowding (Herzog et al., 2015; Manassi et al., 2012). Therefore, it seems plausible that this grouping is what caused the modulation of crowding that we have observed. To investigate this possibility, we used a single flanker dot, where no flanker-flanker grouping is possible (Chakravarthi & Pelli, 2011). In addition to addressing a possible grouping explanation for the previously observed increase in performance, the single flanker experiment allowed us to test for an identifying feature of crowding; a flanker placed radially from the target with respect to fixation tends to crowd more when it is located at a greater eccentricity than the target (outer condition) than when it is placed the same distance from the target toward the fovea (Bex et al., 2003; Bouma, 1973; Petrov & Meleshkevich, 2011).

Experiment 3 Observers

Five subjects participated in Experiment 3, including 2 females and 3 male subjects, one of whom was an author of the study. Subjects ages ranged from 19 to 28.

Experiment 3 Method

In Experiment 3, only a single flanker, identical to the 8 flankers used in the previous experiments, was presented. To compensate for this decrease in difficulty, the eccentricity of the target was increased to 12° in the right visual field. The target and flanker were presented on the horizontal meridian and in each block of trials the flanker remained either more foveal than the target (inner condition) or more eccentric (outer condition). The spacing varied from from 1.5° to 4.2° in steps of .3° according to an adaptive procedure. A 3-up-1-down staircase procedure was used to adjust target-flanker separation in order to determine the 79% correct threshold in each block of trials . In each block, subjects were shown only inner flankers or outer flankers in four interleaved staircases (one for each lighting condition) for a total of 256 trials per block, 64 trials per condition. Each subjected participated in four blocks. Estimates for each condition were obtained by averaging the stimulus values of the last 20 trials for each condition, yielding two estimates per condition (lighting direction x inner/outer flanker direction) per subject, and these estimates were then combined across subjects for a group analysis.

Experiment 3 Results

The results from Experiment 3 are shown in Figure 3.4. For each of the 8 conditions, we obtained 2 target-flanker spacing threshold estimates from each subject using a staircase procedure. These 2 estimates were combined across the 5 subjects to yield 10 estimates per condition. The means and standard errors are shown in the bar plot. Planned comparisons were conducted using a standard ANOVA procedure to assess the significance of the effect of lighting direction and also the effect of the inner vs. outer flanker. Subjects had significantly lower threshold spacings and thus performed better when viewing stimuli consistent with a lighting from above interpretation (target lit from either below or above) vs. stimuli in which the target and flankers were lit from the side, consistent with the results of Experiments 1 and 2 ($t=4.436$, $p < 0.001$). Furthermore, subjects also demonstrated significantly reduced target-flanker spacing thresholds for displays in which the flanker was placed closer to fixation relative to the target rather than farther from fixation ($t=2.606$, $p = 0.011$). This indicates that the flanker impaired target identification more when the flanker was more eccentric than the target, a hallmark of the visual crowding effect. Finally, the replication of the effect of lighting direction from Experiments 1 and 2 is evidence that the reduction of crowding in scenes consistent with lighting-from-above interpretation is not due to increased flanker-flanker grouping in those conditions.

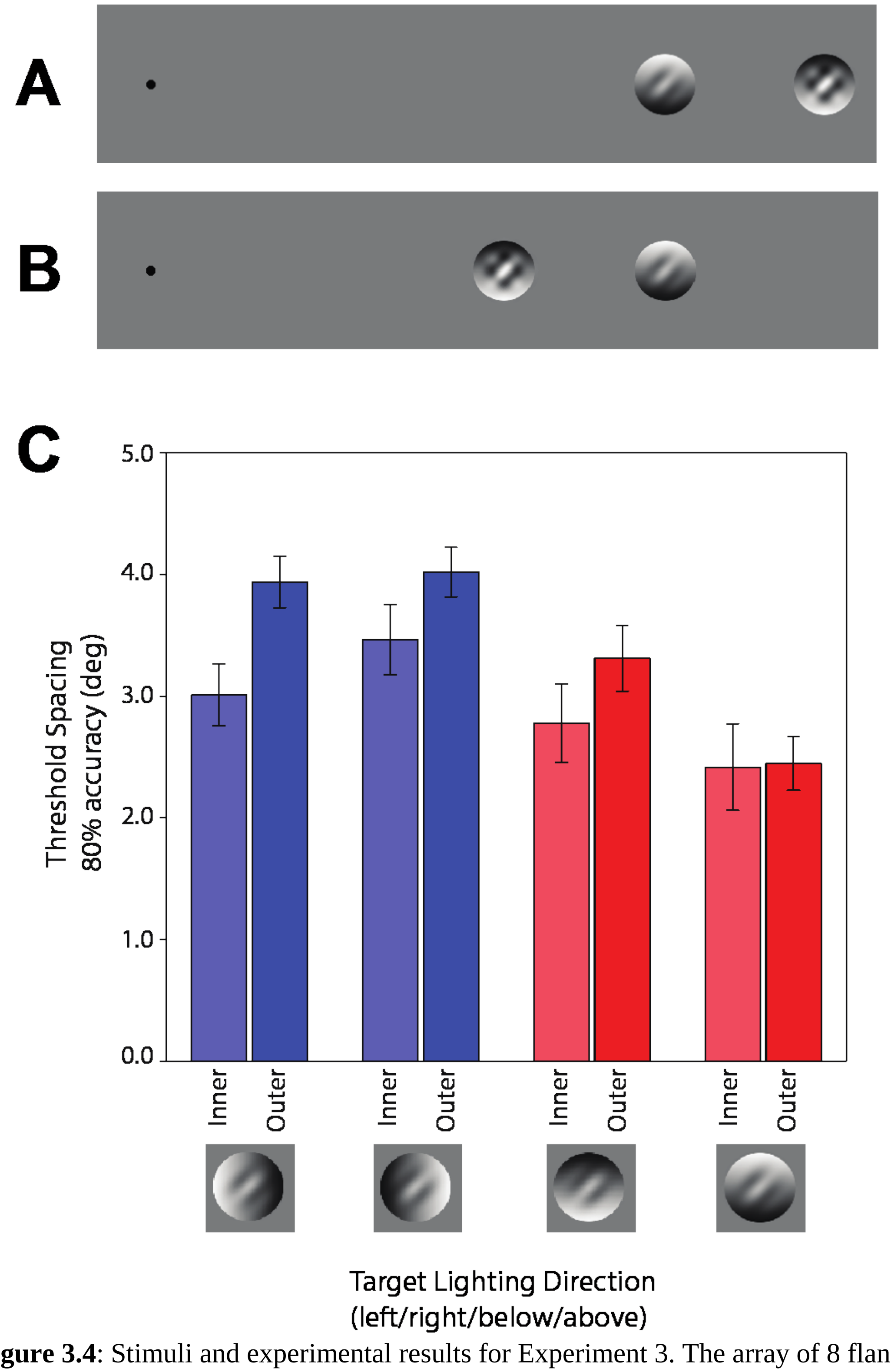

Figure 3.4: Stimuli and experimental results for Experiment 3. The array of 8 flankers was replaced by a single flanker. A) In half of the blocks of trials, subjects saw an "outer" flanker, a flanker more eccentric than the target. B) In the other half of trials, subjects saw an "inner" flanker, a flanker less eccentric than the target. Target-flanker spacing was manipulated in a 3-

down-1-up staircase procedure. C) Group results for Experiment 3. Four 80% accuracy threshold estimates were obtained from each subject for each condition by averaging the last 20 trials for that condition for a total of 10 estimates per condition. Outer flankers crowded significantly more than inner flankers, consistent with a performance impairment due to crowding. Subjects performed significantly better in the "lit-from-above" conditions than the "lit-from-the-side" conditions, despite the presence of only a single flanker, eliminating a flanker-flanker grouping explanation for the results of Experiments 1 and 2.

Discussion

Altering only the direction of interpreted lighting within a scene affects the strength of crowding for an otherwise unchanged array. In Experiment 1, we showed that observers are significantly better at discriminating the orientation of a target when the stimulus array allows for a lighting-from-above interpretation conducive to the perception of shape-from-shading. Importantly, in all conditions tested, target flanker similarity is equated and all other aspects of the stimulus that are known to contribute to crowding strength are controlled. This finding indicates that the computation of lighting and shading information functionally precedes crowding, even for a low level visual feature such as orientation. It also demonstrates that crowding occurs between 3D object representations that are extracted from the 2D retinal image. Oriented shading that would otherwise have no effect on crowding strength, when placed in a context consistent with a lighting-from-above interpretation, gives rise to a perception of depth and 3D shape and this inferred distinction between the 3D representations of the target and flankers is what causes a release from crowding.

Stimulus design considerations contributed to the use of a large orientation difference (90°) in our experiments. While there is some evidence that these large orientation differences are distinguishable at their detection threshold, we do not believe that our results are due to a reduction in detection as opposed to discriminability. A reduction in detection can be explained by contrast masking, rather than crowding (which affects identification), but in our experiments we see impairments at large spacings (3° or more) and exhibiting an inner / outer asymmetry consistent with crowding but not masking. It has been previously argued that a distinction can still be made between detection and identification at the single feature level using a visual search paradigm (Sagi & Julesz, 1984), and our results are consistent with an impairment of single feature discrimination.

There are several possible explanations and mechanisms for this observed release from crowding. Experiment 2 was designed to rule out one 2D image based possibility stemming from a difference in low level stimulus property across the conditions of interest in Experiment 1. In Experiment 1, we observe stronger crowding for horizontal shading gradients as opposed to vertical ones. These gradient directions are confounded with their orientation with respect to fixation; since the stimuli are presented to the right of fixation, the direction of the shading gradient in the lighting-from-above condition is orthogonal to the direction of the target from fixation, while in the lighting-from-side condition the direction of the shading gradient is the same as the direction of the target from fixation. Since there are known radial / tangential asymmetries in crowding, Experiment 2 reverses the confound by placing the stimuli in the lower visual field rather than the right visual field. The similarity of results between Experiments 1 and 2 and consistent release from crowding in the stimulus arrays that imply lighting-from-above suggest that the relation of the direction of the gradient to fixation does not affect our results.

Many of the effects associated with shape-from-shading induced by lighting-from-above are known to also affect strength of crowding. One possible explanation for our results is that the change in lighting direction affects grouping within the presented array. For example, dots which differ in shading from the other dots in a display also tend to group more readily when vertical shading gradients are used (Livne & Sagi, 2007, 2010). Configural grouping effects within arrays of flankers have also been shown to reduce crowding (Manassi et al., 2012). Therefore, it seems possible that the mechanism by which lighting interpretation affects strength of crowding is by altering the strength of grouping between the target and flankers, or between the flankers themselves. However, because the lighting direction effect is still observed in Experiment 3 where we used only one flanker, it seems unlikely that grouping of flankers can explain our findings.

While our data seem to rule out a flanker-flanker grouping explanation for the reduction of crowding in arrays consistent with a lighting-from-above assumption, our primary purpose here is to focus on the importance and role of lighting interpretation on crowding, not to promote a particular intermediary mechanism to explain reduced crowding strength in arrays lit from above. The important point here is that crowding strength within well controlled stimulus arrays which vary only in their potential lighting interpretations is differentiable with respect to interpreted lighting direction. Some other potential intermediary mechanisms are discussed below.

One way in which lighting-from-above consistent arrays might crowd differently from lighting-from-the-side arrays is that dots which differ in shading from the other dots in a display tend to pop-out more when the shading gradients are vertical, resulting in flatter reaction time increases with increased number of distractors than for horizontal gradients in a standard visual search paradigm (Rensink & Cavanagh, 2004). This type of preattentive pop-out has also been shown to reduce crowding in some stimulus arrays (Yeshurun & Rashal, 2010). Therefore, it is possible that this preattentive enhancement at the target location results in the decrease in crowding strength we observe.

Another potential reason is that dots with different vertical shading gradients like those in our lighting-from-above condition create a stronger and less ambiguous perception of depth than horizontal gradients ((Ramachandran, 1988). Kooi et al., (1994) have previously shown that depth differences between target and flankers implied by binocular disparity can modulate the strength of crowding in both magnitude and extent. It is conceivable then that perceived depth due to shape-from-shading (a monocular depth cue) causes the increase in performance that we have observed.

There is also some neurophysiological evidence that may provide some insight into our observed effect. In an fMRI study, neural activity in lower visual areas was shown to be greater for dots with horizontal shading gradients than vertical gradients (Georgieva, Todd, Peeters, & Orban, 2008). This activity may reflect some increased demand in processing these stimuli which could be inhibiting processing of the target in our crowded arrays.

While many attempts have been made to narrow the neural locus of crowding using studies of behavioral responses, doing so is often difficult or impossible in part due to the issue of inseparability of feedforward processing and feedback from higher visual areas. Our data cannot speak to whether the influence of lighting on crowding is a feedforward or feedback process, nor does it constrain the physical locus of where this information loss due to crowding is occurring. However, and perhaps more importantly, our data indicate that crowding is functionally preceded by lighting interpretation and any complete model of crowding must account for the effects of across object scene level characteristics that play a role in determining to what extent crowding will occur. Although some new recent models can easily accommodate for crowding on multiple levels (Chaney, Fischer, & Whitney, 2014) and grouping effects on crowding (Francis, Manassi & Herzog, 2017) they all need to be modified to include the lighting interpretation in a scene. Given our results, it seems likely that the visual system is doing some interpretation of lighting information within a scene before crowding occurs. This leads to the possibility that the visual system may be able to differentially process shading and shadows in a manner that avoids crowding between meaningful pieces of information that are important for segregating and identifying objects, such as object contours, and potentially less useful information for identification, such as shading, which depend on context and lighting of a particular scene.

Chapter 4: Serial dependence occurs in biological motion across different dimensions

Introduction

When recognizing objects, we continuously deal with noisy and rapidly changing input. The visual system itself is one source of this noise and rapid change; our eyes move and blink with high frequency, our head and body are constantly in motion, and there is internal noise at the neuronal level. At the same time, the outer world around us is extremely noisy because of changes in lighting, noise, occlusions and so on. It is then surprising that, despite all these factors, we still perceive a stable representation of the world around us. How does the visual system achieve this stabilization?

A mechanism has been recently proposed that could explain the impressive stability of our visual experience: the object-selective continuity field. Continuity fields promote perceptual stability by inducing serial dependence between similar objects. Because of this perceptual bias, objects and features appear more like similar objects and features that were seen in the recent past (Fischer & Whitney, 2014; Liberman et al., 2014; St John-Saaltink, Kok). Continuity fields shape our perception within a spatiotemporal window. For example, orientation perception is biased toward previous orientations over a large spatial (20° of visual field) and temporal range (10-15 seconds). Interestingly, this bias only occurs for similar orientations, whereas very different orientations remain unbiased. Serial dependence has been shown to influence perception over a wide range of features and objects like orientation (Fischer & Whitney, 2014; Fritsche et al., 2017), faces (Liberman et al., 2014; Taubert, Alais, & Burr, 2016), attractiveness (Kondo, Takahashi, & Watanabe, 2012; Taubert, Van der Burg, & Alais, 2016; Xia, Leib, & Whitney, 2016), numerosity (Corbett, Fischer, & Whitney, 2011), ensemble coding of orientation (Manassi, Liberman, Chaney, & Whitney, 2016) and mapping of numbers onto space (Cicchini, Anobile, & Burr, 2014).

So far, serial dependence has been shown to occur only between *static* 2D images at different levels of visual processing, from orientation. However, in everyday life we are continuously surrounded by an extremely *dynamic* environment, where not only our eyes, head and body move, but also the environment around us is continuously in motion. When people are walking in the street, we ourselves are in motion, and other objects in the scene such as cars are moving, we nonetheless still perceive a stable world.

We theorize that, in order to promote perceptual stability, serial dependence should occur also with dynamic stimuli at different levels of representation. One well studied high level dynamic object representation is evidenced by the perception of biological motion. The visual system infers three-dimensional object movement of a particular form from a cluster of moving dots. We get a vivid impression of a walking human figure from dots which track the motion of a few key points (joints) on a walker. Hence, in this study we used biological motion stimuli as a proxy to investigate whether serial dependence can occur also in the dynamic, 3D environment we deal with in everyday life.

For our purpose, biological motion stimuli are a perfect tool for several reasons. First, they necessarily require integration of information over time into a dynamic structural representation (Lange & Lappe, 2006; Neri, Morrone, & Burr, 1998). Hence, they involve by definition high level feature integration. Second, this class of stimuli conveys a great deal of information using low level features like dot motion. For example, from biological motion we can infer different stimulus dimensions like walking directions, gender, age, identity and even emotions (Atkinson, Dittrich, Gemmell, & Young, 2004; Cutting & Kozlowski, 1977; Dittrich, Troscianko, Lea, & Morgan, 1996; Pollick, Kay, Heim, & Stringer, 2005; Troje, Westhoff, & Lavrov, 2005). Third, in biological motion stimuli our visual system is able to infer a 3D interpretation from a simple sequence of 2D images (Johansson, 1973).

To preview our findings, we found that serial dependence occurs also with biological motion stimuli in (at least) two different dimensions: walking direction and emotion. Our results further strengthen the notion that serial dependence occurs at multiple levels along the visual hierarchy. Taken together, our results show that serial dependence can stabilize our percept even within dynamic, 3D representations. In addition, we also highlight some analysis considerations regarding investigating serial effects when the stimulus space is not circular.

General Method

Apparatus

All experimental procedures were approved by and conducted in accordance with the guidelines and regulations of the UC Berkeley Institutional Review Board. Participants were affiliates of UC Berkeley and provided informed consent in accordance with the IRB guidelines of the University of California at Berkeley. Stimuli were generated on a Windows computer running PsychoPy2 and presented on a gamma-corrected 61cm liquid crystal display monitor in a dark room. The refresh rate of the display was 60 Hz and the resolution was 1600 x 900 pixels. Stimuli were viewed from a distance of 57 cm. Subjects used a keyboard for all responses (left-right arrow keys to adjust the walker, and space bar to confirm bar their response and initiate the next trial).

Observers

Five subjects participated in the experiment (2 male, 3 female, age range 19-31). All participants had normal or corrected-to-normal vision, and all except one were naïve to the purpose of the experiment.

Stimuli

A .05° white fixation dot (150cd/m^2) remained at the center of the screen for the duration of the experiment. The stimulus consisted of a single point-light walker presented to the right of fixation, at 6° of eccentricity. Point-light walkers were composed of configurations of twelve white dots (each dot: 0.11° × 0.11°, 150 cd/m^2) presented against a black background (0.36 cd/m^2). The dots were placed at different locations such that the overall configuration would be perceived as a human body. We created these walkers from a freely available set (Vanrie & Verfaillie, 2004). To create the impression of a walking human body, we generated “videos” from sets of sixty static frames in which the local position of each dot changed from frame-to-frame in a manner consistent with a natural human gait. Each gait cycle (i.e., one step by each foot) lasted 1 second. The application to generate the videos was written in C# and interfaced with OpenGL via the Open Toolkit Library. We generated 180 videos, each with a distinct heading, by rotating the 3D positions of the dots in each frame by a distinct angle around the vertical axis (i.e., the direction of walking). The headings ranged from leftward (−90°) to rightward (90°) in 1° increments (see point-light stimuli in Figure 4.1). We limited the range to forward headings (toward the observer) because backward headings can appear ambiguous (perceived as forward or backward) (Cavanagh, Labianca, & Thornton, 2001). A dot configuration with a completely leftward (−90°) or completely rightward (90°) heading subtended (1.9° × 2.91°) of visual angle at the full extension of the gait cycle (i.e., with ankles maximally extended) and (0.56° × 3.06°) at the minimum extension of the gait cycle (i.e., with ankles crossing the midline of the body). A dot configuration with a completely forward heading (0°) subtended (1.03° × 3.06°) of visual angle. Orthogonal projection was used to create the images of the point-light walkers. On a given trial, a walker was presented for a total of 1000 ms at 60Hz, for one complete gait cycle. The walker started on a random frame in this gait cycle on each trial. A white noise mask was presented for 500ms following the 1000 ms presentation of the walker in the same location on the screen. After the noise mask, a blank screen was shown for 750ms and then a response walker appeared at the same location as the target walker. The response walker was continuously animated and started with a random heading direction. After the subject indicated their response, the trial was followed by a 750ms inter-trial interval.

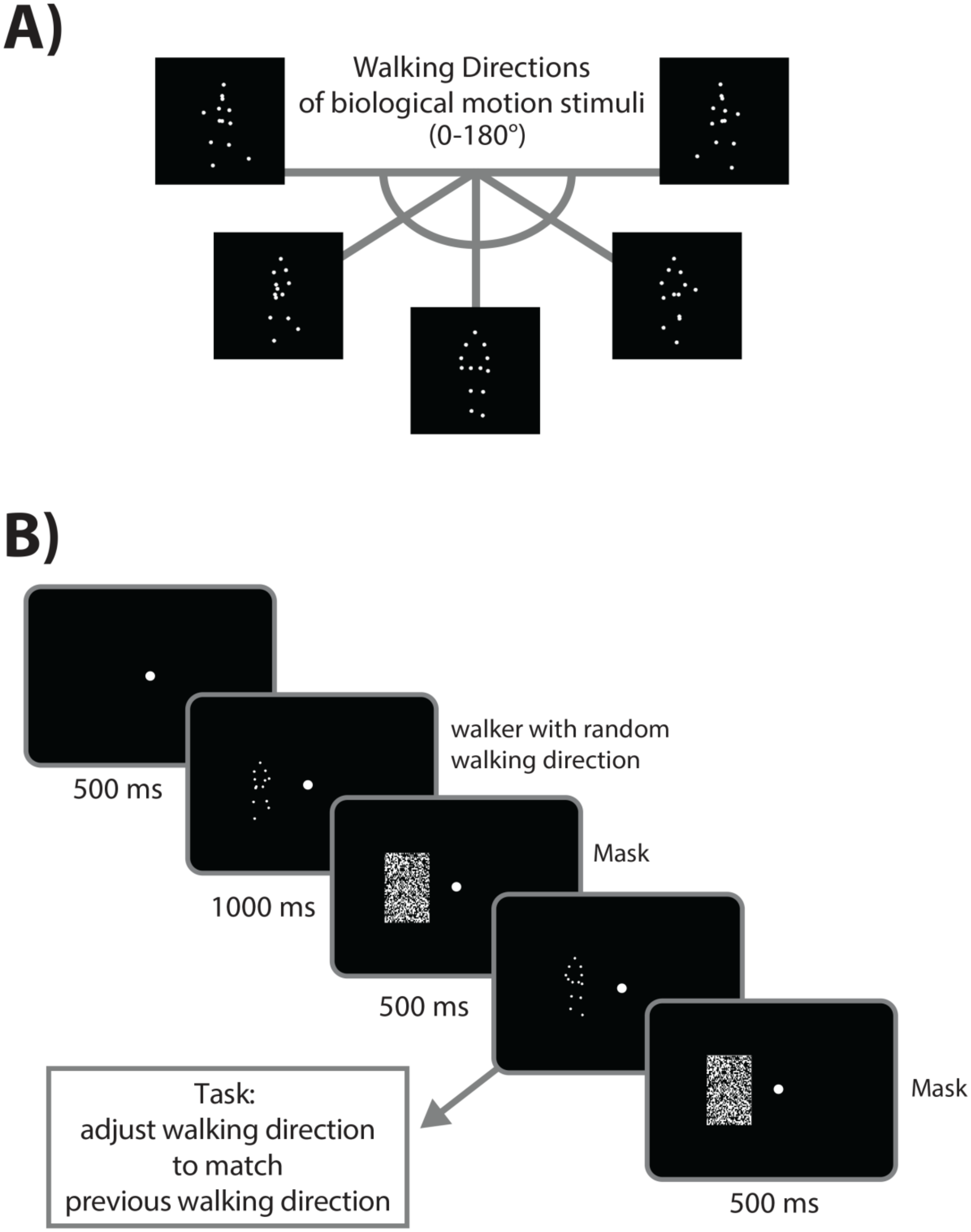

Figure 4.1: The stimuli and presentation sequence used in Experiment 1. A) On each trial, subjects saw a single animated point-light walker facing one of 180 possible directions. The target walker is presented at 6° of eccentricity in either the left or the visual field and subtends an area of approximately 3°x2°. B) The target walker is presented at 60Hz for 1 second, performing one full gait cycle (one step with each foot), beginning on a random frame in the 60 frame gait sequence each trial. The presentation of the walker is followed by a visual mask of white noise for 500ms and then an animated response walker appears at the same location as the target. After a response is entered, there is an ISI of 500ms.

Task and Procedure:

Subjects were instructed to maintain central fixation on the fixation dot throughout the entire experiment. After a brief presentation of a point-light walker in the periphery, followed by a mask, subjects were instructed to match the heading direction of the response walker to that of the target walker. The response walker remained in continuous motion while subjects used the left and right arrow keys to adjust its direction of heading. When the subject felt that they had matched the direction of heading as accurately as possible, they pressed the spacebar key to enter their response. When the response was entered, a noise mask appeared for 500ms, followed by an itertrial interval of 500ms before the next trial. Each subject participated in 3 blocks of 100 trials.

Experiment 1: Serial Dependence In Walking Direction

The purpose of this experiment was to investigate whether serial dependence occurs also for the perceived heading direction of biological walkers. If there is serial dependence for walkers, a walker will be perceived as walking in a direction that is biased toward the heading of the previous walker. Previous results showing serial dependence for visual motion stimuli have not included stimuli which require integration of dynamic information like the movement of the dots that create an impression of biological motion. This experiment investigates whether this perceptual bias also occurs for these higher level representations.

Data Analysis:

Error outlier removal

First, we rejected outlier responses on a per subject basis. This was done by eliminating any trial that has an error greater than 3 standard deviations of the subject's errors in absolute magnitude. This results in a rejection rate of under 5% for all subjects. We then accounted for any systematic left-right biases in the subject's response by subtraction the mean of the response error for each subject. For all subjects, this mean was less than 2 morph units and had no substantial effect on our curve fitting.

Edge effect correction

Examining subjects' responses as a function of the stimulus level reveals an obvious pattern of edge effect influence. In the center portion of the stimulus range, subjects' responses tend to follow the one-to-one line, indicating accurate performance on the task on average. However, near the edges of the range of stimulus values, it is clear to see that the average response becomes compressed toward the center of the stimulus range. These edge effects are biases caused by the minimum and maximum possible response value that result in the mean of responses deviating from the presented stimulus values. In order to remove this artifact, we removed trials which had a stimulus value near the edge of the stimulus range from the analysis.. In our analysis of the walking direction data, we removed the outer 30° of walkers on both ends of the spectrum. Current trials with a walking direction of 0-30° and 150-180° were discarded. We chose 30° because it is equal to 2 standard deviations of response error and therefore most likely includes all of the trials that would likely have their error distributions influenced by the stimulus edge.

Derivative of Gaussian fit

After removing this subset of trials, we then proceed to our analysis as established in Fischer & Whitney, (2014). Response errors are computed by subtracting the direction of heading for the target walker from the final direction of heading of the response walker. The relative direction of the previous trial to the current trial is computed by subtracting the current trial target direction from the previous trial target direction. We plot subjects' response errors [actual response - current stimulus value] as a function of the relative direction between trials [previous stimulus value - current stimulus value]. Since serial dependence is an attraction of the current percept to the perception of the previous percent, we would expect to see subjects' errors biased more leftward when the previous stimulus was walking in a direction to the left of the current stimulus, and vice versa for the right. Furthermore, since serial dependence has been shown to be tuned for similarity in the stimulus dimension being judged, we expect to see that this bias will have a local maximum at some stimulus value near the current stimulus value and then fall off toward unbiased for dissimilar trials. We fit a parametric model to the data, assuming that the data can be modeled by the derivative of a Gaussian function of the form:

$$\boldsymbol{y = a * b * \sqrt{2} * e^{-(bx)^2+.5}}$$

In this formulation, **y** represents response error on a given trial [response direction - target direction] and **x** represents the relative direction between trials [previous target direction - current target direction]. The parameter **a** represents the maximum value of the function and the parameter **b** controls its width. We refer to **a** as the half-amplitude of serial dependence, since the function has range **-a** to **a**. The function is symmetric about **x**=0 and has an **x**-intercept at **x**=0, which represents a trial where the previous and current stimulus have the same value. A positive amplitude **a** indicates that there is a leftward bias when the previous stimulus was to the left, and vice versa for the right, and is evidence of a serial dependence effect. A negative aftereffect would produce an opposite result, a negative amplitude. An amplitude of 0 indicates no bias due to the previous stimulus in either direction. The curve is fit using the Nelder-Mead simplex algorithm for nonlinear optimization to minimize the sum of squared error.

Bootstrap analysis

To obtain an estimate of the magnitude of the serial dependence effect and measure its variability, we fit this function to each subjects' data and generate a measure of confidence using a bootstrap confidence interval. For each subject, we resample the data with replacement 10,000 times and refit the derivative of a Gaussian function. From the distribution of 10,000 amplitudes generated by this procedure, we create an empirical confidence interval for the bootstrap amplitudes by calculating the .025 and .975 percentiles of the data.

Compensating for central tendency bias

Another possible confounding explanation for a finding of positive serial dependence on a linear stimulus space is regression to the mean or a central tendency bias in responses. A more extreme stimulus is more likely to have been preceded by a stimulus closer to the mean of the distribution of possible stimuli. Therefore a central tendency in responses can bias serial dependence to be positive because both the response and the previous stimulus will be in the direction of the mean. In order to control for this possibility, we conduct an analysis keeping only the current trials for which the previous trial was more extreme than the current trial. One main drawback to this control analysis is that it requires discarding 76% of the data. Therefore, we conduct the analysis at the group level, combining all of the accepted trials from each subject into one super subject and following the same fitting procedure as before. The response errors are plotted as a function of the difference between the previous trial stimulus value and the current trial stimulus value and fitted with a derivative of a Gaussian. We construct a bootstrap confidence interval for the amplitude of serial dependence by resampling with replacement the accepted trials collapsed across subjects.

Goodness-of-fit test

In order to ensure that our model was accurately capturing the influence of the previous trial, we also conducted a goodness-of-fit test for each subject. To test whether the influence of the previous trial is an important driving factor in our goodness-of-fit, we tied the stimulus value presented on each trial with the response given by the subject, but permuted the order of the trials. This ensures that if subjects have any particular biases in responding to any particular level of the stimulus that were not controlled by removing the outer ranges of stimulus values, those biases will remain present in the data. However, any systematic effects of the previous trial should be removed. We then fit the data with our normal fitting procedure, and compare the SSEs of the permuted null fits to the SSEs of the fits on the actual subjects' data.

Results:

We measure serial dependence as the half-amplitude of our derivative of a Gaussian fit. A significant result favoring the serial dependence hypothesis is indicated when 95% of half-amplitudes from the bootstrap resampled data lie above a=0. The results from one representative subject is shown in Figure 4.2A. The data shows a bias toward positive errors on the positive x-axis near x=0 and small negative errors on the negative x-axis near x=0, indicating the serial dependence effect. The dotted line shows the average of the twenty nearest neighbors. Figure 4.2B shows the results for the all of the subjects. 4 out of 5 subjects individually show significant serial dependence with 95% bootstrap confidence intervals lying in an entirely positive range. The group mean amplitude is shown in red with an average value of 4.02 and a 95% confidence interval of 1.89 to 6.07. This provides strong evidence for serial dependence in the judgment of heading for perceived biological motion, after adjusting for edge effects caused by the linear stimulus space.

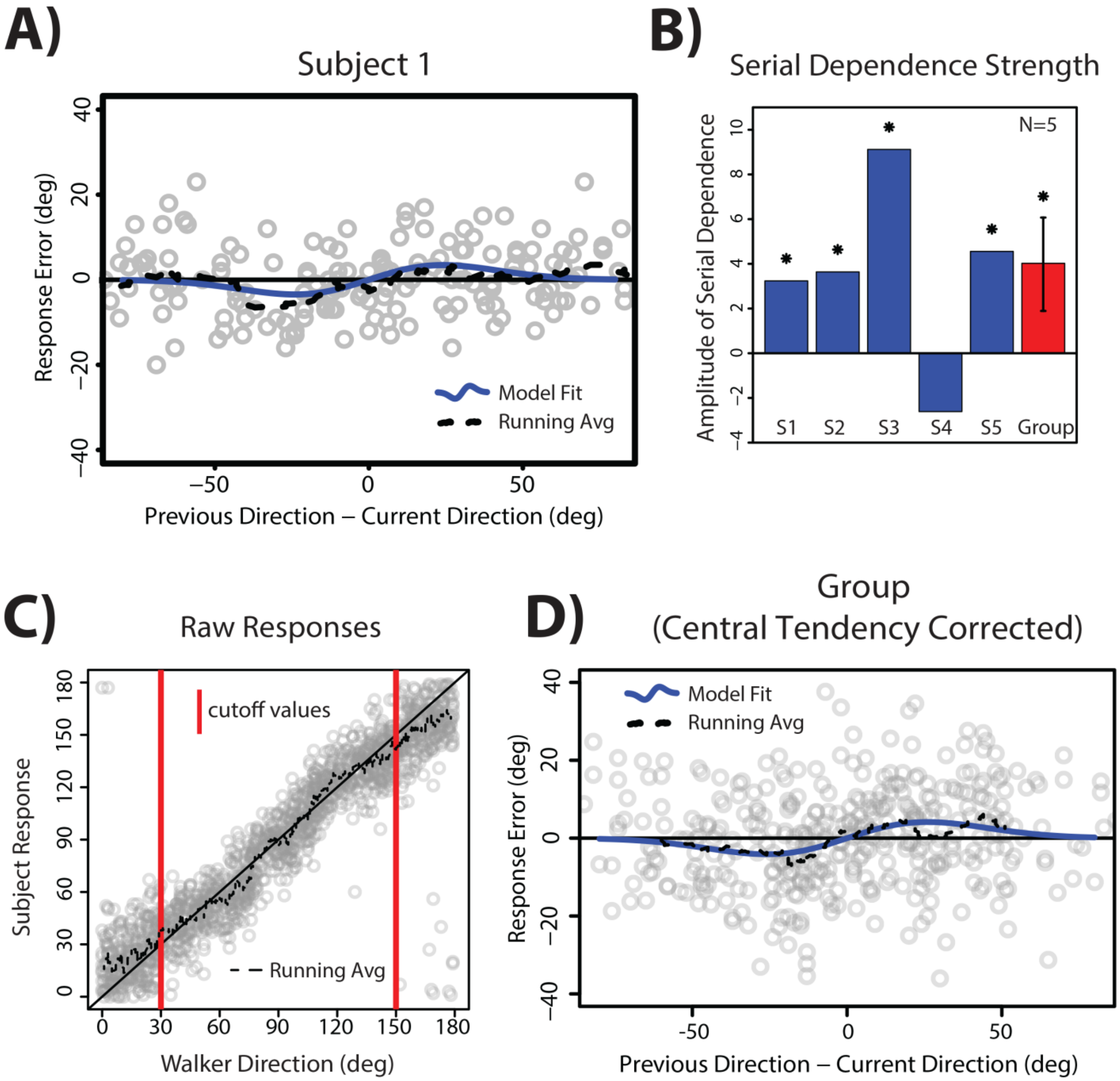

Figure 4.2: Results of Experiment 1. A) Individual results from a representative subject are shown in A. Response errors are plotted as a function of the previous stimulus value minus the current stimulus value and fit with a derivative of a Gaussian. The amplitude of this fit is used as a measure of serial dependence. The fitted line is shown in blue and a running average is shown as a black dotted line. B) The resulting amplitudes for all subjects, as well as the group mean. Asterisks represent significance at the 95% confidence level generated from 10,000 iterations of bootstrap resampling. Four out of five subjects show significant serial dependence. The group mean amplitude of serial dependence is also significant. The error bar represents 95% bootstrap confidence interval. C) The raw responses as a function of stimulus valu. The black line denotes a running average of K-nearest neighbors with K=20. We can see some edge effects caused by our linear response space. We removed trials with stimulus values beyond the red cutoff levels. D) The results of removing trials to account for possible regression to the mean. Only trials with more extreme values than the previous trial are kept. The data is collapsed across subjects and fit

at the group level. We find significant serial dependence using the same bootstrap resampling test as in B.

The results of the regression to the mean analysis are shown in Figure 4.3. This analysis pits the serial dependence effect against any possible influence of edge effects and regression to the mean. Despite this very conservative approach, we find that the amplitude of the fit to this group data, 4.1394, is significantly greater than zero (P<0.0454). This indicates that the evidence for serial dependence is robust to edge effects, regression-to-the-mean effects, and differences in tuning and amplitude of the serial dependence effect across subjects. Using only trials with stimulus values which are more extreme than the stimulus value on the previous trial, there remains strong evidence for serial dependence.

We next compare the sum of square errors of our model fit to the true data with the sum of squares from randomly permuted trial orders. We find that the fit in the goodness-of-fit is significantly better for the individual subjects' responses when the true order of the stimuli and responses is used, as opposed to the randomly permuted trial orders, with the exception of the one subject with a non-significant negative amplitude of serial dependence (p < 0.0144, p < 0.6486, p < 0.0455, p < 0.0215, p < 0.0001). This shows that model is actually accounting for differences in response errors that are due to the stimulus order, and in particular, due to the previous trial. Random permutations which disrupt the information present in the true [previous stimulus - current stimulus] values on the x-axis but maintain the exact relationship between the current stimulus value and response error are no longer predicted well by the model.

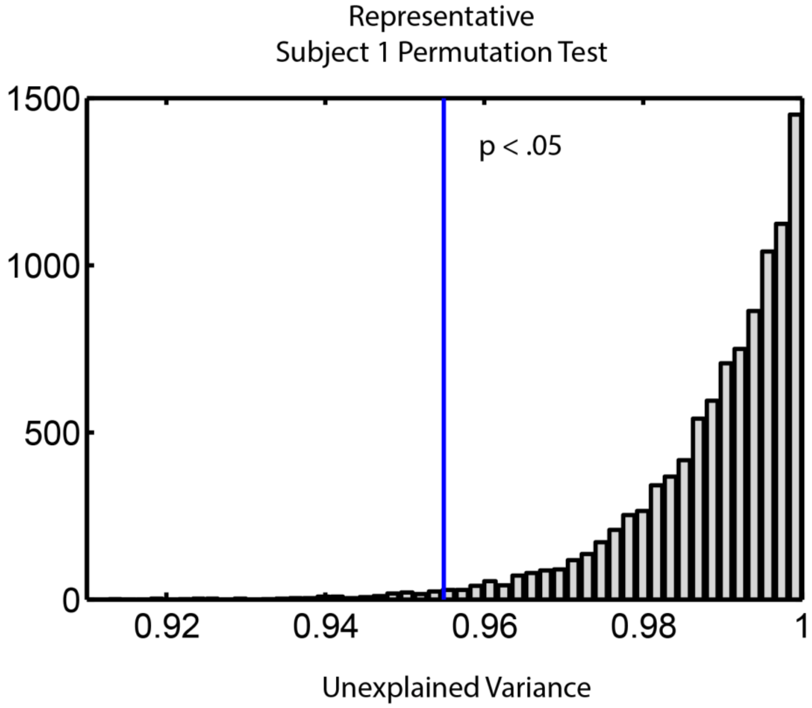

Figure 4.3: Results of the goodness-of-fit analysis for Experiment 1. The histogram shows the unexplained variance after fitting the derivative of a Gaussian to the data from a single subject with a shuffled trial sequence with 10,000 repetitions. All subjects except one subject with a non-

significant negative amplitude of serial dependence are fit significantly better when the correct trial order is used to code the data, showing direct evidence for the serial dependence effect.

Experiment 2 - Serial Dependence in Walker Emotion

In Experiment 1 we showed that serial dependence can occur between biological motion stimuli, specifically heading direction of point light walkers. In Experiment 2, we extended our results from Experiment 1 to a new dimension of biological motion stimuli: emotion.

Observers:

Five subjects participated in the experiment (2 male, 3 female, age range 19-31). One was an author of the study. The other participants were unaware of the purpose of the study. All had normal or corrected to normal acuity.

Stimuli:

The stimulus in Experiment 2 is nearly identical to that of Experiment 1 except that instead of varying the direction of heading of the target walker, we presented a single walker that varied along the dimension of emotion from sad to happy. We used 100 morphs of point-light walkers, ranging from saddest (0) to happiest (100). Again, each walker consisted of 60 image frames shown for 1 second at 60Hz, starting at a randomized frame. This way each walker was shown for one complete gait cycle. The walker was followed by a white noise mask at the same visual field location for 500ms. After 750ms, a response walker appeared at the same location with a randomly assigned emotion from sad to happy. The walkers were shown at an eccentricity of 6° of visual angle, and the walkers subtended approximately 3°in height and 2°in width (see Experiment 1 Stimulus for details).

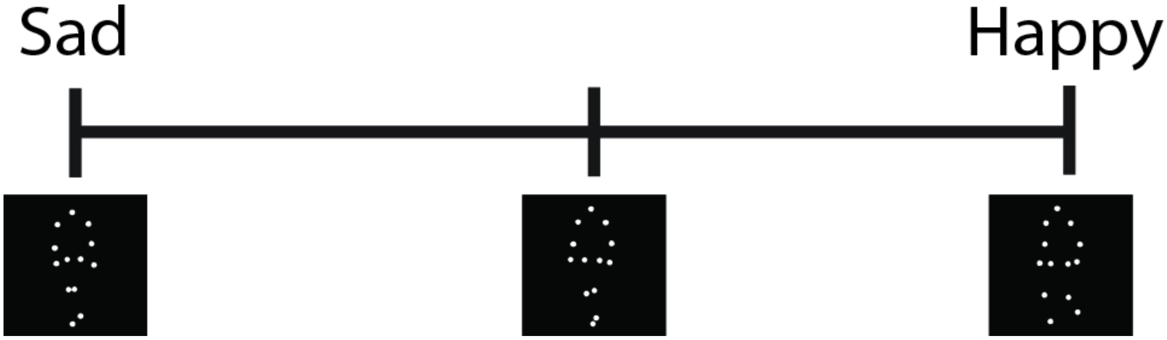

Figure 4.4: The stimuli for Experiment 2. As in Experiment 1, stimuli consisted of a single animated walker (approximately 3° tall and 2° width) presented at 6° visual angle in the right hemifield. Each walker took one of 100 possible values representing a continuum from sad to happy. The target walker is presented at 60Hz for 1 second, performing one full gait cycle (one step with each foot), beginning on a random frame in the 60 frame gait sequence each trial.

Task and Procedure:

The task and procedure for Experiment 2 were similar to Experiment 1 except that subjects were asked to match the emotion of the target walker. Subjects were instructed to maintain central fixation on a fixation dot for the duration of the experiment. After the target walker, mask, delay, and appearance of the response walker, subjects were asked to use the left and right arrow keys to adjust the perceived emotion of the response walker to match that of the target. When subjects were satisfied that the emotion they had just seen had been matched, they pressed the spacebar to enter their response and move on to the next trial.

Results

We conducted the same analyses for Experiment 2 as in Experiment 1. First, we looked at subjects' responses as a function of stimulus value and noted a similar pattern to Experiment 1. There is clear compression of responses near the most extreme stimulus values and the data in the middle of the stimulus range is well modeled by the one-to-one line. We note that the magnitude of errors tends to be larger in this experiment relative to the size of the stimulus space (stimulus space here is size 100, in the previous experiment it is 180, and yet the standard deviations of the response errors are 15.6 and 14.13 respectively).

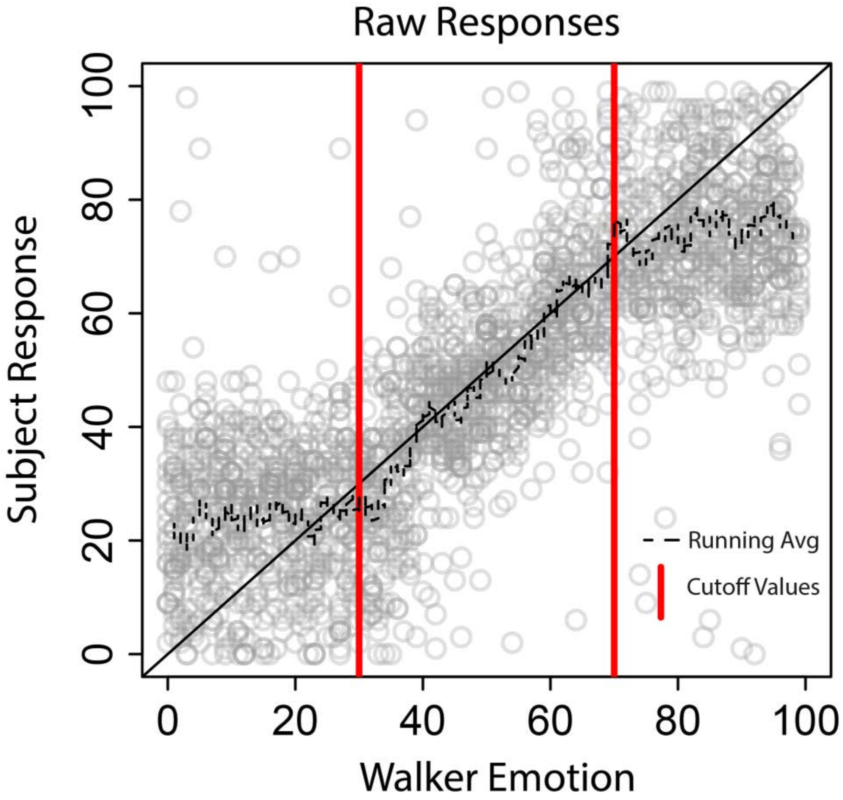

Figure 4.5: Raw responses for the group of subjects as a function of the emotion of the presented target walker. The black dotted line represents the running average of responses. The black solid line is the one-to-one, or correct response, line. Significant deviation from this line occurs at about 2 standard deviations of response error away from the edges. We rejected trials with stimuli outside the red cutoff values.

Our primary analysis of the data excludes the outer 30 values of the data, as in Experiment 1, to eliminate the possible influence of edge effects on our fits. After rejecting these trials, we conduct the same analysis as in Experiment 1, fitting a derivative of a Gaussian to the response errors as a function of the difference between the previous and current stimulus. We generate bootstrap confidence intervals for each subject by resampling the data points with replacement. The results for one representative subject are shown in Figure 4.6A. The positive value of the fit for positive x-values near x=0 and negative value for negative x values near x=0 indicate the serial dependence effect. The dotted line represents the k-Nearest Neighbor mean with k=20, and the model fit follows the running average. The results for all of the individual subjects can be seen in Figure 4.6B. Three of the five subjects individually showed significant positive serial dependence (p <.0001, p<.0913, p<.0012,p<.8173, p<.0093), with one subject showing a negative amplitude. We also test for significance at the group level by averaging the bootstrap amplitude estimates across subjects on each iteration. The mean amplitude is 2.395 units with a bootstrap 95% confidence interval of [0.9473 ,4.4152] and a p-value of .0013. These results overall indicate the presence of serial dependence for emotional walkers after controlling for edge effects, despite somewhat noisier performance on the emotional walker task.

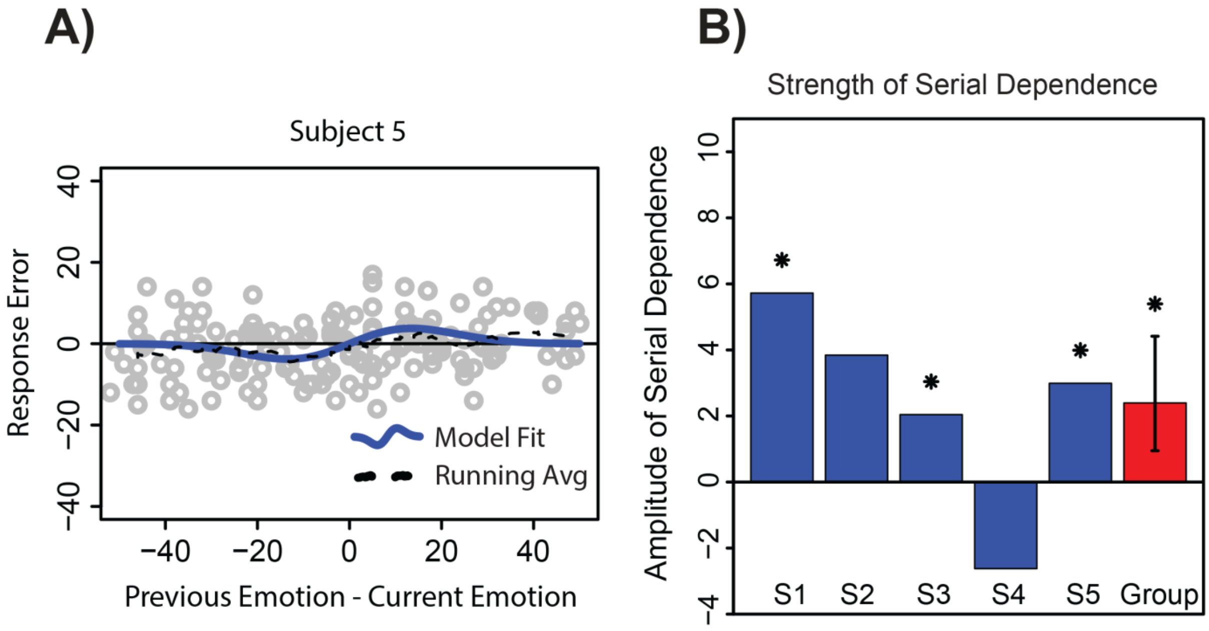

Figure 4.6: Results of derivative of a Gaussian fit for Experiment 2. A) Results for a representative subject. As in Experiment 1, we plot response error as a function of the difference in stimulus value from previous trial to current trial, this time using emotion of the walker. We fit a derivative of a Gaussian model to the data (blue line) and measure the amplitude as a measure of serial dependence. B) Estimated amplitudes for each individual subject and the average for the group. We test for significance using a 95% bootstrap confidence interval.

In a separate analysis, as in Experiment 1, we also control for regression-to-the-mean or central tendency bias in the data by including only trials for which the previous trial was more extreme than the current trial value. We combined the data across subjects into one super subject and used the derivative of Gaussian fitting procedure to measure the amplitude of serial dependence. Unlike Experiment 1, we did not get a reliable measure of the amplitude of serial dependence for the emotional walkers with this aggressively conservative analysis. The best fit of the derivative of Gaussian is essentially a positive sloped line that is not tuned near 0. There are a few possible reasons for this difference. First, subjects' responses are significantly noisier for emotional walkers than for direction of heading with standard deviations of 15.6 and 14.13 on stimulus spaces that consist of 100 and 180 morphs respectively. This means that on average errors are twice as large as in Experiment 1. Second, the width of the derivative of Gaussian fits is substantially wider relative to the full size of the stimulus space in Experiment 2. This means that most trials remaining after the conservative rejection procedure lay on the interval -20 to 20 morph units, where we expect to see serial dependence. Since a positively sloped line well characterizes the middle portion of the derivative of Gaussian with positive amplitude, it is sensible to see a linear best fit. Furthermore, this linear trend cannot be explained by artifacts like edge effects or regression to the mean due to our rejection procedure.

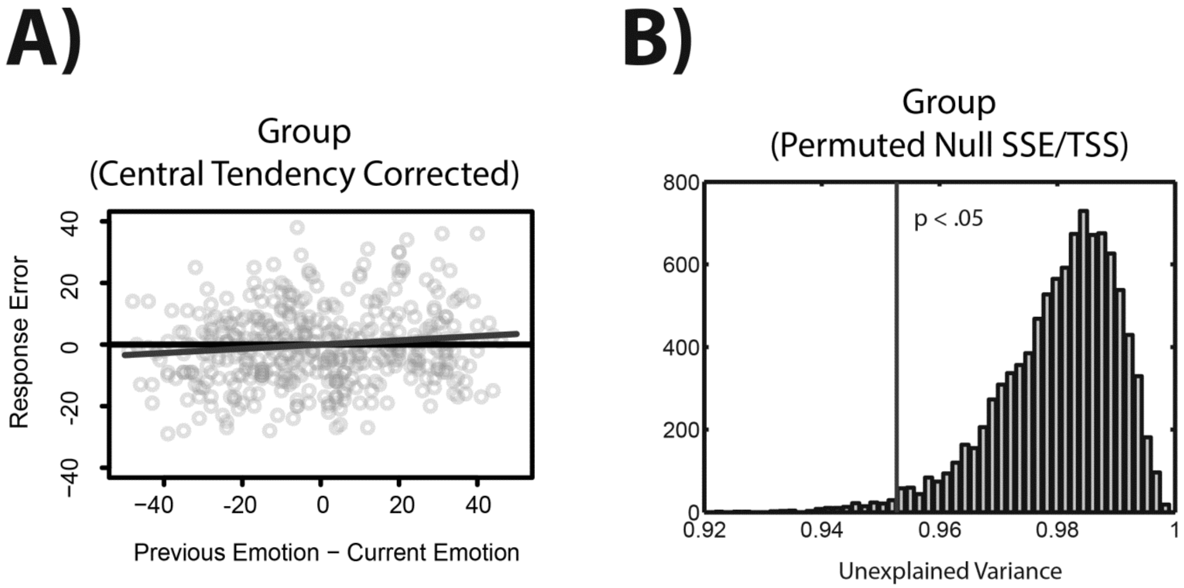

Figure 4.7: Two control analyses for Experiment 2. A) Central tendency correction. We removed all trials except where the current trial is more extreme than the previous trial. This controlled for central tendency in subjects' responses. The data was collapsed across subjects. Because of the narrow range of stimuli, we saw a positive trend rather than tuned serial dependence. B) The results of the goodness-of-fit test at the group level. On each permutation, we shuffled the order of the trials while keeping the stimulus value / response value pairs intact. We measured the unexplained variance and find that the model explained the true data significantly better than the reshuffled data (p < .0166).

As in the analysis for Experiment 1, we also conducted a goodness-of-fit test for our model on the emotional walker data. We compared the unexplained variance in the responses when the model is fit to the true data to the unexplained variance in the responses when the model is fit to data with a shuffled presentation order, keeping the pairs of stimulus values and response errors intact. Because of the increased level of noise in the emotional walker responses compared to the direction of heading task, 60% of trials were eliminated by the edge effect control (as opposed to 33% in Experiment 1). Therefore, the goodness-of-fit test was calculated for the group average rather than for individual subjects. We compared the average unexplained variance across subjects for the model fit to the true data to average unexplained variance across subjects for the model fit to the shuffled data. We found that on average the model performs significantly better for the true data than the shuffled data (p < .0166).

Our finding of serial dependence for the emotion of perceived biological motion further strengthens the results of Experiment 1 and it shows that serial dependence occurs across more than one dimension of a stimulus, even for complex dynamic representations such as biological motion. It also shows that serial dependence is robust across stimulus dimensions that may have different autocorrelation structure in the natural world; the timescales over which direction of motion and perceived emotion of walkers change may be quite different but serial dependence occurs for both dimensions.

General Discussion

The visual world is very noisy, with motion, clutter, changes in lighting, head movements, eye movements, and internal system noise, yet we perceive the world to be stable and comprehensible. One of the ways this is achieved by the visual system is serial dependence: what we see is biased toward what we have seen in the recent past. We have demonstrated that this serial dependence affects the perception of biological motion in two different stimulus domains: walking direction and emotion. Our results show that the brain uses serial dependence to achieve this appearance of stability, even when the world is actually in motion. We have also offered some important analysis considerations for investigating serial dependence in these types of stimulus domains.

Evidence of serial dependence has been shown in many types of visual stimuli, from orientation (Fischer & Whitney, 2014; Fritsche et al., 2017), to size, to face identity, emotion, (Liberman et al., 2014; Taubert, Alais, & Burr, 2016), and attractiveness (Kondo, Takahashi, & Watanabe, 2012; Taubert, Van der Burg, & Alais, 2016; Xia, Leib, & Whitney, 2016), with our perception of the stimulus systematically biased toward the recent past. It has been hypothesized that this serial dependence effect is the result of a specialized spatio-temporal operator, the continuity field, which integrates the current stimulus with congruent recently seen stimuli to promote stability of perception. Static scenes are perfectly autocorrelated, so it is beneficial to maintain stability of perception both to represent the actual characteristics of the world and to save computation.

However, dynamic scenes also contain structure that is highly autocorrelated, and our results show that there is serial dependence for relatively stable aspects of dynamic scenes such as direction of heading or emotion of a walker. Humans can quickly, accurately, and robustly perceive biological motion from a set of impoverished cues, such as a point-light walker or even a degraded point-light walker (Johansson, 1973). We can also infer a wide variety of information from point-light walkers, including identity (Troje, Westhoff, & Lavrov, 2005), direction of heading, gender (Pollick, Kay, Heim, & Stringer, 2005), and emotion (Atkinson, Dittrich, Gemmell, & Young, 2004; Dittrich, Troscianko, Lea, & Morgan, 1996). Motion integration is important for the recognition of biological motion and it has been shown that perception of biological motion is more accurate under moving rather than static conditions. However, motion information alone is not always sufficient and other studies have demonstrated the important role of form information in biological motion with inversion effects (Sumi, 1984) and classification images (Lu & Liu, 2006). Overall, biological motion perception likely involves some integration of both local motion information and configural form. This complex type of representation requires integration over both space and time to accurately compute and extract meaningful attributes, such as emotion or walking direction.

This is the first time that serial dependence has been demonstrated for a dynamic high level representation which requires both structural inference and integration over time. Other work has recently shown that serial dependence can occur for the perception of motion (Alais, Leung, & Van der Burg, 2017), but our results show that serial dependence not only occurs for visual motion itself, but for representations which require visual motion in a particular structural arrangement.

Our results are also important because they addressed a few potential analysis issues when measuring serial dependence. Some of the earliest work on serial dependence utilized circular stimulus spaces to measure serial dependence, such as orientation and even emotion or identity of a face. This avoids potential confounds due to edge artifacts. However, in many domains of interest, such as size or facial attractiveness, such a circular stimulus space does not exist and special precautions must be taken to avoid contamination by edge effects. In our case, the walking direction of point light walkers was ambiguous when 360° of walking directions are utilized, and so we were limited to using only the frontal 180°. Because of this, edge effects were introduced which can cause the appearance of serial dependence despite no real perceptual dependence on the previous trial. If subjects' responses had a central tendency in the stimulus space, this could also produce what looks like serial dependence, when in fact it is an artifact.

We developed several novel analysis methods (see Analysis section for details) to control for these effects and remove possible biases. First, we mitigated edge effects by rejecting trials where the stimulus value was within 2 standard deviations of subjects' response errors of the minimum or maximum possible value. Second, we controlled for central tendency biases by performing a separate analysis looking at only trials where the current stimulus was more extreme than the previous. This is a conservative analysis that includes trials where a central tendency bias would weaken serial dependence. We also developed a goodness-of-fit measure for our model that directly tests serial dependence by examining the effect of trial order.

Taken together, our results show that there is serial dependence for the perception of apparent biological motion, providing a mechanism for the perception of stable characteristics in dynamic representations. We also argue that some care must be taken when investigating serial effects in linear stimulus spaces and developed some new analysis techniques.

Chapter 5: Conclusion

Object recognition is one of the central tasks for the human visual system. We have argued that a complete account of object recognition should be seen not simply as a sequential process operating on a single object, but rather as a process that allows for the interaction of multiple objects through spatial and temporal context at many different levels of the visual system. Specifically, we have shown that both visual crowding and serial dependence are interactions between a target object and its context that influence the veridicality of object perception. We have also seen that any complete account of object recognition must allow for top-down influence in the processing of low level features, allowing lighting interpretation to influence interaction amongst oriented Gabors, and for the perception and extraction of high level features from impoverished point-light displays.

Visual crowding cannot be explained by information loss at each successive level of information processing, as it posited by many current models of crowding. This is at odds with our ability to recognize complex objects in the periphery, and with many findings in the crowding literature including the configural effects of grouping on crowding, and as we have developed here, the influence of lighting interpretation within a scene on crowding. Instead, crowding is better explained by the Hierarchical Sparse Selection Model presented here which argues that crowding is actually a product of insufficient sampling read-out from a population code that represents with high fidelity the information present at each level of processing.

In both our crowding and serial dependence findings, we extend the space of representations that can be selectively involved in contextual interactions to 3D representations that are derived from the 2D stimulus based on prior assumptions about the structure of the world. The 3D representation that causes a selective release from crowding in Chapter 3 is due to a basic assumption of a single light source coming from above. In Chapter 4, we see that serial dependence influences the perception of a stimulus that requires configural priors and integration over time.

Taken together, our results point to an influence of context in visual object recognition in both space and time that operates at many different levels of visual processing.

References

Aghdaee, S. M. (2005). Adaptation to spiral motion in crowding condition. *Perception-London 34*(2): 155-162.

Alais, D., Leung, J., & Van der Burg, E. (2017). Linear summation of repulsive and attractive serial dependencies: orientation and motion dependencies sum in motion perception. *The Journal of Neuroscience*. http://www.jneurosci.org/content/early/2017/03/22/JNEUROSCI.4601-15.2017.abstract

Amano, K., B. A. Wandell, et al. (2009). Visual field maps, population receptive field sizes, and visual field coverage in the human MT+ complex. *Journal of Neurophysiology 102*(5): 2704-2718.

Anderson, E. J., Dakin, S. C., Schwarzkopf, D. S., Rees, G., & Greenwood, J. A. (2012). The neural correlates of crowding-induced changes in appearance. *Current Biology*, *22*(13), 1199–1206. http://doi.org/10.1016/j.cub.2012.04.063

Andriessen, J. and H. Bouma (1976). Eccentric vision: Adverse interactions between line segments. *Vision Research 16*(1): 71-78.

Astle, A.T., Mcgovern, D.P., A., & Mcgraw, P. V. (2014). Characterizing the role of disparity information in alleviating visual crowding. *Journal of Vision*, *14*(6), 1–14. http://doi.org/10.1167/14.6.8

Atas, A., Faivre, N., Timmermans, B., Cleeremans, A., & Kouider, S. (2014). Nonconscious learning from crowded sequences. *Psychological Science*, *25*(1), 113–9. http://doi.org/10.1177/0956797613499591

Atkinson, A. P., Dittrich, W. H., Gemmell, A. J., & Young, A. W. (2004). Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception*, *33*(6), 717–746. https://doi.org/10.1068/p5096

Atkinson, A. P., Dittrich, W. H., Gemmell, A. J., & Young, A. W. (2004). Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception*, *33*(6), 717–746. https://doi.org/10.1068/p5096

Balas, B., Nakano, L., & Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision*, *9*(12), 13.1-18. http://doi.org/10.1167/9.12.13

Baldi, P. and W. Heiligenberg (1988). How sensory maps could enhance resolution through ordered arrangements of broadly tuned receivers. *Biological Cybernetics 59*(4-5): 313-318.

Banno, H. and J. Saiki (2012). Calculation of the mean circle size does not circumvent the bottleneck of crowding. *Journal of Vision 12*(11).

Bex, P. J., Dakin, S. C., & Simmers, A. J. (2003). The shape and size of crowding for moving targets. *Vision Research*, *43*(27), 2895–2904. http://doi.org/10.1016/S0042-6989(03)00460-

7

Bi, T., Cai, P., Zhou, T., and Fang, F. (2009). The effect of crowding on orientation-selective adaptation in human early visual cortex. *Journal of Vision*, *9*(11), 13,1-10. http://doi.org/10.1167/9.11.13

Bouma, H. (1970). Interaction Effects in Parafoveal Letter Recognition. *Nature*, *226*(5241), 177–178. http://doi.org/10.1038/226177a0

Bouma, H. (1973). Visual interference in the parafoveal recognition of initial and final letters of words. *Vision Research, 13*(4), 767–782. http://doi.org/10.1016/0042-6989(73)90041-2

Bulakowski, P. F., R. B. Post, et al. (2009). Visuomotor crowding: the resolution of grasping in cluttered scenes. *Frontiers in Behavioral Neuroscience 3*.

Buschman, T. J. and E. K. Miller (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science 315*(5820): 1860-1862.

Cavanagh, P. (2001). Seeing the forest but not the trees. *Nature Neuroscience 4*(7): 673-673.

Chakravarthi, R., & Pelli, D. G. (2011). The same binding in contour integration and crowding. *Journal of Vision*, *11*(2011), 1–12. http://doi.org/10.1167/11.8.10

Chaney, W., Fischer, J., & Whitney, D. (2014). The hierarchical sparse selection model of visual crowding. *Frontiers in Integrative Neuroscience*, *8*(September), 73. http://doi.org/10.3389/fnint.2014.00073

Chastain, G. (1981). Inhibition of feature extraction with multiple instances of the target feature in different orientations. *Psychological Research 43*(1): 45-56.

Chastain, G. (1982). Feature mislocalizations and misjudgments of intercharacter distance. *Psychological Research 44*(1): 51-65.

Chung, S. T. (2007). Learning to identify crowded letters: Does it improve reading speed? Vision *Research 47*(25): 3150-3159.

Chung, S. T., D. M. Levi, et al. (2001). Spatial-frequency and contrast properties of crowding. *Vision Research 41*(14): 1833-1850.

Cicchini, G. M., Anobile, G., & Burr, D. C. (2014). Compressive mapping of number to space reflects dynamic encoding mechanisms, not static logarithmic transform. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(21), 7867–72. https://doi.org/10.1073/pnas.1402785111

Corbett, J. E., Fischer, J., & Whitney, D. (2011). Facilitating stable representations: Serial dependence in vision. *PLoS ONE*, *6*(1). https://doi.org/10.1371/journal.pone.0016701

Corbetta, M., F. M. Miezin, et al. (1993). A PET study of visuospatial attention. The Journal of *Neuroscience 13*(3): 1202-1226.

Curcio, C. and J. Harting (1978). Organization of pulvinar afferents to area 18 in the squirrel monkey: evidence for stripes. *Brain Research 143*(1): 155-161.

Cutting, J., & Kozlowski, L. (1977). Recognizing Friends by Their Walk: Gait Perception Without Familiarity Cues. *Bulletin of the Psychonomic Society*, *9*(5), 353–356. https://doi.org/10.3758/BF03337021

Dakin, S. C., J. Cass, et al. (2010). Probabilistic, positional averaging predicts object-level crowding effects with letter-like stimuli. *Journal of Vision 10*(10).

Dittrich, W. H., Troscianko, T., Lea, S. E. G., & Morgan, D. (1996). Perception of emotion from dynamic point-light displays represented in dance. *Perception*, *25*(6), 727–738. https://doi.org/10.1068/p250727

Dittrich, W. H., Troscianko, T., Lea, S. E. G., & Morgan, D. (1996). Perception of emotion from dynamic point-light displays represented in dance. *Perception*, *25*(6), 727–738. https://doi.org/10.1068/p250727

Efron, B., & Tibshirani, R. (1986). Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Statistical Science*, *1*(1), 54–75. http://doi.org/10.1214/ss/1177013815

Engel, S., Glover, G., Wandell, B., (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cerebral Cortex* **7**(2): 181–192.

Eurich, C. W. and H. Schwegler (1997). Coarse coding: calculation of the resolution achieved by a population of large receptive field neurons. *Biological Cybernetics 76*(5): 357-363.

Faivre, N., & Kouider, S. (2011). Multi-feature objects elicit nonconscious priming despite crowding. *Journal of Vision*, *11*(3), 1–10. http://doi.org/10.1167/11.3.2

Faivre, N., Berthet, V., & Kouider, S. (2012). Nonconscious influences from emotional faces: A comparison of visual crowding, masking, and continuous flash suppression. *Frontiers in Psychology*, *3*(MAY), 129. http://doi.org/10.3389/fpsyg.2012.00129

Farzin, F., Rivera, S. M., & Whitney, D. (2009). Holistic crowding of Mooney faces. *Journal of Vision*, *9*(6), 1–15. http://doi.org/10.1167/9.6.18

Fischer, J. and D. Whitney (2011). Object-level visual information gets through the bottleneck of crowding. *Journal of Neurophysiology 106*(3): 1389-1398.

Fischer, J. and D. Whitney (2012). Attention gates visual coding in the human pulvinar. Nature *Communications* **3**: 1051.

Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Nature Neuroscience*, *17*(5), 738–743. https://doi.org/10.1016/j.cub.2014.09.025

Fischer, J., N. Spotswood, et al. (2011). The emergence of perceived position in the visual system. *Journal of Cognitive Neuroscience 23*(1): 119-136.

Flom, M., Weymouth, F., & Kahneman, D. (1963). Visual Resolution and Contour Integration.

Journal of the Optical Society of America, 53(9), 1026–32. http://doi.org/10.1364/JOSA.53.001026

Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nature Neuroscience, 14*(9), 1195–1201. http://doi.org/10.1038/nn.2889

Fritsche, M., Mostert, P., Lange, F. P. De, Fritsche, M., Mostert, P., & Lange, F. P. De. (2017). Opposite Effects of Recent History on Perception and Decision Report Opposite Effects of Recent History on Perception and Decision. *Current Biology, 27*(4), 590–595. https://doi.org/10.1016/j.cub.2017.01.006

Georgieva, S. S., Todd, J. T., Peeters, R., & Orban, G. A. (2008). The extraction of 3D shape from texture and shading in the human brain. *Cerebral Cortex*, *18*(10), 2416–2438. http://doi.org/10.1093/cercor/bhn002

Greenwood, J. A, Bex, P. J., & Dakin, S. C. (2009). Positional averaging explains crowding with letter-like stimuli. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(31), 13130–5. http://doi.org/10.1073/pnas.0901352106

Greenwood, J. A., Bex, P. J., & Dakin, S. C. (2010). Crowding Changes Appearance. *Current Biology, 20*(6), 496–501. http://doi.org/10.1016/j.cub.2010.01.023

Harp, T. D., D. W. Bressler, et al. (2007). Position shifts following crowded second-order motion adaptation reveal processing of local and global motion without awareness. *Journal of Vision* **7**(2).

Harvey, B. and Dumoulin, S. (2011). The relationship between cortical magnification factor and population receptive field size in human visual cortex: constancies in cortical architecture. *The Journal of Neuroscience 31*(38): 13604-13612.

He, S., Cavanagh, P., & Intriligator, J. (1996). Attentional resolution and the locus of visual awareness. *Nature, 383*(6598), 334–337. http://doi.org/10.1038/383334a0

Herzog, M. H., & Manassi, M. (2015). Uncorking the bottleneck of crowding: A fresh look at object recognition. *Current Opinion in Behavioral Sciences, 1*. http://doi.org/10.1016/j.cobeha.2014.10.006

Herzog, M. H., Sayim, B., Chicherov, V., & Manassi, M. (2015). Crowding, grouping, and object recognition: A matter of appearance. *Journal of Vision*, *15*(6). http://doi.org/10.1167/15.6.5

Huckauf, A., Heller, D., & Nazir, T. A. (1999). Lateral masking: limitations of the feature interaction account. *Perception & Psychophysics, 61*(1), 177–189. http://doi.org/10.3758/BF03211958

Hussain, Z., B. S. Webb, et al. (2012). Perceptual learning reduces crowding in amblyopia and in the normal periphery. *The Journal of Neuroscience 32*(2): 474-480.

Ikeda, H., & Watanabe, K. (2016). Action Congruency Influences Crowding When Discriminating Biological Motion Direction. *Perception*, *45*(9), 1046–1059.

http://doi.org/10.1177/0301006616651952

Ikeda, H., Watanabe, K., & Cavanagh, P. (2013). Crowding of biological motion stimuli. *Journal of Vision*, *13*(4), 1–6. http://doi.org/10.1167/13.4.20.doi

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception Psychophysics*, *14*(2), 201–211. https://doi.org/10.3758/BF03212378

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception Psychophysics*, *14*(2), 201–211. https://doi.org/10.3758/BF03212378

Kaas, J. H. and D. C. Lyon (2007). Pulvinar contributions to the dorsal and ventral streams of visual processing in primates. *Brain Research Reviews 55*(2): 285-296.

Kleffner, D. a, & Ramachandran, V. S. (1992). On the perception of shape from shading. *Perception & Psychophysics*, *52*(1), 18–36. https://doi.org/10.1038/331163a0

Kondo, A., Takahashi, K., & Watanabe, K. (2012). Sequential effects in face-attractiveness judgment. *Perception*, *41*(1), 43–49. https://doi.org/10.1068/p7116

Kooi, F. L., Toet, A., Tripathy, S. P., & Levi, D. M. (1994). The effect of similarity and duration on spatial interaction in peripheral vision. *Spatial Vision*, *8*(2), 255–279. http://doi.org/10.1017/CBO9781107415324.004

Kouider, S., Berthet, V., & Faivre, N. (2011). Preference is biased by crowded facial expressions. *Psychological Science*, *22*(2), 184–9. http://doi.org/10.1177/0956797610396226

Krumhansl, C. L. and E. A. Thomas (1977). Effect of level of confusability on reporting letters from briefly presented visual displays. *Perception & Psychophysics 21*(3): 269-279.

Lange, J., & Lappe, M. (2006). 2006 A Model of Biological Motion Perception from Configural Form Cues, *26*(11), 2894–2906. https://doi.org/10.1523/JNEUROSCI.4915-05.2006

Lau, J. S. F. and S.-H. Cheung (2012). Illusory contour formation survives crowding. *Journal of Vision 12*(6).

Levi, D. M. (2008). Crowding-An essential bottleneck for object recognition: A mini-review. *Vision Research*, *48*(5), 635–654. http://doi.org/10.1016/j.visres.2007.12.009

Levi, D. M., Klein, S. A., & Aitsebaomo, A. P. (1985). Vernier acuity, crowding and cortical magnification. *Vision Research, 25*(7), 963–977. http://doi.org/10.1016/0042-6989(85)90207-X

Liberman, A., Fischer, J., & Whitney, D. (2014). Serial dependence in the perception of faces. *Current Biology*, *24*(21), 2569–2574. https://doi.org/10.1016/j.cub.2014.09.025

Liberman, A., Zhang, K., & Whitney, D. (2016). Serial dependence promotes object stability during occlusion. *Journal of Vision*, *16*(15), 1–10.

Livne, T., & Sagi, D. (2007). Configuration influence on crowding. *Journal of Vision*, *7*(2), 4.1-

12. http://doi.org/10.1167/7.2.4

Livne, T., & Sagi, D. (2010). How do flankers' relations affect crowding? *Journal of Vision*, *10*(2010), 1–14. http://doi.org/10.1167/10.3.1.

Louie, E. G., Bressler, D. W., & Whitney, D. (2007). Holistic crowding: selective interference between configural representations of faces in crowded scenes. *Journal of Vision*, *7*(2), 24.1-11. http://doi.org/10.1167/7.2.24

Lu, H., & Liu, Z. (2006). Computing dynamic classification images from correlation maps. *Journal of Vision*, *6*(4), 475–483. https://doi.org/10.1167/6.4.12

Manassi, M., Hermens, F., Francis, G., & Herzog, M. H. (2015). Release of crowding by pattern completion. *Journal of Vision*, *15*(8). http://doi.org/10.1167/15.8.16

Manassi, M., Herzog, M. H., Sayim, B., & Herzog, M. H. (2012). Grouping , pooling , and when bigger is better in visual crowding. *Journal of Vision*, *12*(10), 1–14. http://doi.org/10.1167/12.10.13.Introduction

Manassi, M., Liberman, A., Chaney, W., & Whitney, D. (2016). The serial dependence of perception in crowds. *Journal of Vision*, *16*(12), 238.

Manassi, M., Lonchampt, S., Clarke, A., & Herzog, M. H. (2016). What crowding can tell us about object representations. *Journal of Vision*, *16*(3). http://doi.org/10.1167/16.3.35

Martelli, M., N. J. Majaj, et al. (2005). Are faces processed like words? A diagnostic test for recognition by parts. *Journal of Vision* **5**(1).

Maus, G. W., J. Fischer, et al. (2013). Motion-Dependent Representation of Space in Area MT+. *Neuron 78*(3): 554-562.

McKone, E. (2004). Isolating the special component of face recognition: peripheral identification and a Mooney face. *Journal of Experimental Psychology: Learning, Memory, and Cognition 30*(1): 181.

Millin, R., Arman, A. C., Chung, S. T. L., & Tjan, B. S. (2014). Visual crowding in V1. *Cerebral Cortex*, *24*(12), 3107–3115. http://doi.org/10.1093/cercor/bht159

Nandy, A. S. and B. S. Tjan (2012). Saccade-confounded image statistics explain visual crowding. *Nature Neuroscience 15*(3): 463-469.

Nazir, T. A. (1992). Effects of lateral masking and spatial precueing on gap-resolution in central and peripheral vision. *Vision Research 32*(4): 771-777.

Oliva, A. (2005). Gist of the scene. *Neurobiology of Attention 696*: 64.

Op De Beeck, H. and R. Vogels (2000). Spatial sensitivity of macaque inferior temporal neurons. *Journal of Comparative Neurology 426*(4): 505-518.

Op de Beeck, H. P., J. Haushofer, et al. (2008). Interpreting fMRI data: maps, modules and dimensions. *Nature Reviews Neuroscience 9*(2): 123-135.

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, *4*(7), 739–744. http://doi.org/10.1038/89532

Pelli, D. G. (2008). Crowding: A cortical constraint on object recognition. *Current Opinion in Neurobiology* ***18***(4): 445.

Pelli, D. G., & Tillman, K. a. (2008). The uncrowded window of object recognition. *Nature Neuroscience*, *11*(10), 1129–1135. http://doi.org/10.1038/nn1208-1463b

Pelli, D. G., Palomares, M., & Majaj, N. J. (2004). Crowding is unlike ordinary masking: Distinguishing feature integration from detection. *Journal of Vision*, *4*(12), 12. http://doi.org/10.1167/4.12.12

Petersen, S. E., D. L. Robinson, et al. (1987). Contributions of the pulvinar to visual spatial attention. *Neuropsychologia 25*(1): 97-105.

Petrov, Y. and A. V. Popple (2007). Crowding is directed to the fovea and preserves only feature contrast. *Journal of Vision* **7**(2).

Petrov, Y., & Meleshkevich, O. (2011). Locus of spatial attention determines inward–outward anisotropy in crowding. *Journal of Vision*, *11*(4), 1. http://doi.org/10.1167/11.4.1.Introduction

Petrov, Y., Popple, A. V, & McKee, S. P. (2007). Crowding and surround suppression: not to be confused. *Journal of Vision*, *7*(2007), 12.1-9. http://doi.org/10.1167/7.2.12

Põder, E. (2007). Effect of colour pop-out on the recognition of letters in crowding conditions. *Psychological Research 71*(6): 641-645.

Pollick, F. E., Kay, J. W., Heim, K., & Stringer, R. (2005). Gender recognition from point-light walkers. *Journal of Experimental Psychology. Human Perception and Performance*, *31*(6), 1247–65. https://doi.org/10.1037/0096-1523.31.6.1247

Pollick, F. E., Kay, J. W., Heim, K., & Stringer, R. (2005). Gender recognition from point-light walkers. *Journal of Experimental Psychology. Human Perception and Performance*, *31*(6), 1247–65. https://doi.org/10.1037/0096-1523.31.6.1247

Pouget, A., P. Dayan, et al. (2000). Information processing with population codes. Nature *Reviews Neuroscience 1*(2): 125-132.

Purushothaman, G. and D. C. Bradley (2004). Neural population code for fine perceptual decisions in area MT. *Nature Neuroscience 8*(1): 99-106.

Qiu, A., Rosenau, B., et al. (2006). Estimating linear cortical magnification in human primary visual cortex via dynamic programming. *NeuroImage 31*(1): 125-138.

Raiguel, S., M. Hulle, et al. (1995). Shape and spatial distribution of receptive fields and antagonistic motion surrounds in the middle temporal area (V5) of the macaque. *European Journal of Neuroscience* **7**(10): 2064-2082.

Ramachandran, V. S. (1988). Perception of shape from shading. *Nature*, *331*(6152), 163–166. http://doi.org/10.1038/331163a0

Rensink, R. A., & Cavanagh, P. (2004). The influence of cast shadows on visual search. *Perception*, *33*(11), 1339–1358. http://doi.org/10.1068/p5322

Roelfsema, P. R. (2006). Cortical Algorithms for Perceptual Grouping. *Annual Review of Neuroscience*, *29*(March), 203–227. http://doi.org/10.1146/annurev.neuro.29.051605.112939

Saarela, T. P., Sayim, B., Westheimer, G., & Herzog, M. H. (2009). Global stimulus configuration modulates crowding. *Journal of Vision*, *9*(2009), 1–11. http://doi.org/10.1167/9.2.5.

Sagi, D., & Julesz, B. (1984). Detection versus discrimination of visual orientation. *Perception*, *13*(5), 619–628. https://doi.org/10.1068/p130619

Sayim, B., Westheimer, G., & Herzog, M. H. (2010). Gestalt Factors Modulate Basic Spatial Vision. *Psychological Science, 21*(5), 641–644. http://doi.org/10.1177/0956797610368811

Schall, J. D., A. Morel, et al. (1995). Topography of visual cortex connections with frontal eye field in macaque: convergence and segregation of processing streams. *The Journal of Neuroscience 15*(6): 4464-4487.

Sereno, M., Dale, A., et al. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science 268*:889–893.

Smith, S. L. and M. Häusser (2010). Parallel processing of visual space by neighboring neurons in mouse visual cortex. *Nature Neuroscience 13*(9): 1144-1149.

Snippe, H. P. and J. J. Koenderink (1992). Discrimination thresholds for channel-coded systems. *Biological Cybernetics 66*(6): 543-551.

St John-Saaltink, E., Kok, P., Lau, H. C., & de Lange, F. P. (2016). Serial Dependence in Perceptual Decisions Is Reflected in Activity Patterns in Primary Visual Cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *36*(23), 6186–92. https://doi.org/10.1523/JNEUROSCI.4390-15.2016

Strasburger, H., & Wade, N. J. (2015). James Jurin (1684–1750): A pioneer of crowding research? *Journal of Vision*, *15*(1), 9, 1–7. http://doi.org/10.1167/15.1.9.doi

Strasburger, H., Rentschler, I., & Jüttner, M. (2011). Peripheral vision and pattern recognition: A review. *Journal of Vision*, *11*(5), 13. http://doi.org/10.1167/11.5.13.Contents

Sumi, S. (1984). Upside-down presentation of the Johansson moving light-spot pattern. *Perception*, *13*(3), 283–286. https://doi.org/10.1068/p130283

Sun, H.-M., & Balas, B. (2014). Face features and face configurations both contribute to visual crowding. *Attention, Perception & Psychophysics*, *77*, 508–519. http://doi.org/10.3758/s13414-014-0786-0

Taubert, J., Alais, D., & Burr, D. (2016). Different coding strategies for the perception of stable and changeable facial attributes. *Scientific Reports*, *6*, 2–8. https://doi.org/10.1038/srep32239

Taubert, J., Van der Burg, E., & Alais, D. (2016). Love at second sight: Sequential dependence of facial attractiveness in an on-line dating paradigm. *Scientific Reports*, *6*(March), 22740. https://doi.org/10.1038/srep22740

Toet, A., & Levi, D. M. (1992). The two-dimensional shape of spatial interaction zones in the parafovea. *Vision Research*, *32*(7), 1349–1357. http://doi.org/10.1016/0042-6989(92)90227-A

Torralba, A., A. Oliva, et al. (2006). Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychological Review 113*(4): 766.

Troje, N. F., Westhoff, C., & Lavrov, M. (2005). Person identification from biological motion: Effects of structural and kinematic cues. *Perception \& Psychophysics*, *67*, 667–675. https://doi.org/10.3758/bf03193523

Troje, N. F., Westhoff, C., & Lavrov, M. (2005). Person identification from biological motion: Effects of structural and kinematic cues. *Perception \& Psychophysics*, *67*, 667–675. https://doi.org/10.3758/bf03193523

Tyler, C. W., & Likova, L. T. (2007). Crowding: a neuroanalytic approach. *Journal of Vision*, *7*(2), 16.1-9. http://doi.org/10.1167/7.2.16

Wallace, J. M., & Tjan, B. S. (2011). Object crowding. *Journal of Vision*, *11*(6), 19. http://doi.org/10.1167/11.6.19

Westheimer, G., & Hauske, G. (1975). Temporal and spatial interference with vernier acuity. *Vision Research*, *15*(10), 1137–1141. http://doi.org/10.1016/0042-6989(75)90012-7

Whitney, D. (2005). Motion distorts perceived position without awareness of motion. *Current Biology 15*(9): R324-R326.

Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences*, *15*(4), 160–168. http://doi.org/10.1016/j.tics.2011.02.005

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: II. Bootstrap-based confidence intervals and sampling. *Perception & Psychophysics*, *63*(8), 1314–1329. http://doi.org/10.3758/BF03194545

Wilkinson, F., Wilson, H. R., & Ellemberg, D. (1997). Lateral interactions in peripherally viewed texture arrays. *Josa a*, *14*(9), 2057–2068. http://doi.org/10.1364/JOSAA.14.002057

Williams, M. A., S. Dang, et al. (2007). Only some spatial patterns of fMRI response are read out in task performance. *Nature neuroscience 10*(6): 685-686.

Wolford, G. (1975). Perturbation model for letter identification. *Psychological review 82*(3): 184.

Wolford, G., F. Marchak, et al. (1988). Practice effects in backward masking. *Journal of Experimental Psychology: Human Perception and Performance 14*(1): 101.

Xia, Y., Leib, A. Y., & Whitney, D. (2016). Serial dependence in the perception of attractiveness. *Journal of Vision*, *16*(15), 28.

Yeh, S.-L., He, S., & Cavanagh, P. (2012). Semantic Priming From Crowded Words. *Psychological Science*. http://doi.org/10.1177/0956797611434746

Yeshurun, Y., & Rashal, E. (2010). Precueing attention to the target location diminishes crowding and reduces the critical distance. *Journal of Vision*, *10*(10), 16. http://doi.org/10.1167/10.10.16