TO INFEROTEMPORAL NEURONS

THE WHOLE IS NOT THE SUM OF THE PARTS

by

Erin Crowder Hare

BS, College of William & Mary, 2007

Submitted to the Graduate Faculty of

The Dietrich School of Arts & Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2017

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS & SCIENCES

This dissertation was presented

by

Erin Crowder Hare

It was defended on

April 6, 2017

and approved by

Dissertation Advisor: Marlene Behrmann, Professor, Psychology

Tai Sing Lee, Professor, Computer Science

Marlene Cohen, Assistant Professor, Neuroscience

Matthew Smith, Assistant Professor, Ophthalmology

Carl Olson, Professor, Center for the Neural Basis of Cognition

TO INFEROTEMPORAL NEURONS

THE WHOLE IS NOT THE SUM OF THE PARTS

Erin Crowder Hare, PhD

University of Pittsburgh, 2017

Vision seems to occur effortlessly and without mistakes. As a result, it is easy to lose sight of the complex representational mechanisms going on under the hood. In macaque monkeys, the brain region thought to be the ultimate mediator of object recognition is the inferotemporal cortex (IT). The purpose of this dissertation was to investigate how IT neurons respond to parts of a display. We used two different paradigms that disrupt the perception of object parts to query how different parts of a visual scene interact.

The first project was concerned with the behavioral phenomenon known as crowding, in which clutter causes peripheral objects to devolve into an unintelligible jumble. We are the first to develop a task conducive to concurrent behavior and neuronal recordings in monkeys. To demonstrate the relevance of our task, we turned to a hallmark of crowding: that what matters is the eccentricity and spacing between objects, not object size. Having demonstrated this, we were set to proceed to neuronal recordings.

Our primary question was whether crowding quantitatively reduced the strength of IT neuronal selectivity or alternatively whether crowding induced a qualitative change to the neuronal code. Our results support the latter hypothesis. We then asked additional follow-up questions regarding size-sensitivity and adjacency of part-part interactions. Overall, our results were incompatible with a pooling model of crowding and consistent with models based on attention, texture, or source confusion.

The final experiment was concerned with whether certain parts of compound objects were preferentially represented over others. To do this we recorded IT spiking activity while monkeys viewed composite shapes made up of overlapping outlines, as well as all the possible constituent closed parts created by the overlap. Humans tend to only perceive the simpler shapes originally used to create the composite, but the same was not true of IT neurons. Instead, they represented the composite more like its external contour than any other part, especially in the initial phase of the response.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Carl Olson, who is hands down the smartest person I have ever met. You taught me how to think critically, Carl, and I will always be grateful. I would also like to thank my committee members: Marlene Behrman, Tai Sing Lee, Marlene Cohen, and Matthew Smith. I always knew you were rooting for me. Finally, I would like to thank my outside examiner, David Sheinberg, for graciously sacrificing time and energy to participate in this final milestone of my doctorate.

This work would not have been possible without the animal husbandry assistance of Karen McCracken. Somehow you do the work of four people and yet still manage to keep us lab rats in line. Balancing the competing desires of the regulatory bodies and the researchers is no easy task, but you handled it with grace so that we could focus on the science.

Thank you also to my peers in the Olson, Colby, and Cohen labs who kept me company, kept me motivated, and rarely asked me to clean their monkeys. Especially, thank you to Marvin Leathers, Becca Gerth, Doug Ruff, Josh Alberts, Travis Meyer, Suchitra Ramachandran, Nat Williams, and Michael Shteyn.

Finally, this dissertation is dedicated to Frederick, Jim, and Russell. Thanks for putting up with me while I was working on this. I love you guys more than you’ll ever know.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 GENERAL INTRODUCTION

We spend every waking moment perceiving objects in the world around us. This process occurs so effortlessly that it is easy to forget what a challenging task object recognition poses for the brain. A central problem of visual perception is how the brain interprets two-dimensional patterns of light as belonging to three-dimensional objects. In order to interpret these shapes appropriately, they must first be grouped properly. But how does the brain know which features belong to which objects?

The goal of this introductory chapter is to briefly review the current state of our knowledge of the neuronal mechanisms underlying object recognition. The focus will primarily be on the visual system of the macaque monkey, with the aim of highlighting some critical gaps in our current understanding. This chapter will conclude by identifying a set of specific experimental aims that address these gaps. Chapter 2 will lay out the results of a series of experiments designed to investigate whether macaque monkeys experience the phenomenon of crowding in peripheral vision. Chapter 3 will include the results of a set of experiments designed to investigate the neuronal basis of crowding. Chapter 4 will present the results of an experiment designed to investigate how overlapping shapes are segregated and interpreted by the brain. Finally, chapter 5 will conclude by discussing how the results of experiments have furthered our understanding of the neuronal mechanisms of object recognition.

1.1 AGNOSIA AND GESTALT

Patients with apperceptive agnosia are not blind, but they lack the ability to make sense of the global organization of an object. Acuity, color vision, depth perception, and motion perception are all perfectly normal (Farah, 2004). However, when they are asked to reproduce drawings of complex objects the process is laborious and the result is an unintelligible jumble.

In the 1970s, a 73-year-old artist was rendered agnosic after a stroke. The patient was unable to recognize faces and about 75% of the objects presented to him. When drawing these unnamable objects, he tended to focus on the details instead of the global organization, and he often lost his place. As a result, his drawings contained some of the same features as the original image, but they were disjointed and disproportional (Fig. 1). Interestingly, despite his stroke, his drawings retained the same style as before. He maintained his techniques of perspective, shadowing, and texture, suggesting that all he lost was the capacity to meld object parts into a cohesive whole (Wapner et al., 1978).

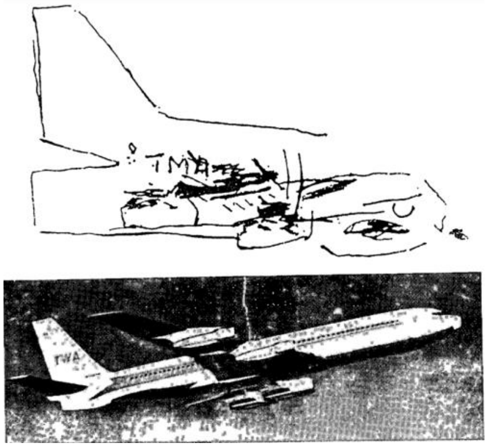

Figure 1. Airplane drawing by an artist with apperceptive agnosia attempting to copy from the photograph below. (Reprinted with permission from Cortex, Wapner et al. (1978), Fig. 6).

This notion of a cohesive whole visual percept dates all the way back to the Gestalt psychologists, based in the Berlin School of Experimental Psychology around the turn of the 20th century. Among this clan was Kurt Koffka, who famously said, “The whole is other than the sum of the parts,” (Koffka, 1935). What he meant by this statement is that the whole, or the gestalt, of a visual object has an existence in the perceptual system independent of its parts. The simple presence of the parts does not define the whole.

Rather, those parts must cohere in a specific way to create a meaningful whole.

Driven largely by introspection, the Gestaltists worked out a set of laws of perceptual organization. At the heart of these laws is the Law of Prägnanz, which is a German word meaning goodness of form (Metzger, 1936). This law commands that the environment is perceived in the simplest way possible. For instance, if two lines cross it is much more parsimonious to view them as both continuing on their original trajectories after the point of intersection rather than suddenly jibing to take up the other's course. This particular example is also known as the Law of Continuity.

Although simple observation lends credence to the Gestalt laws, evidence for an underlying mechanism is scant (Lee and Nguyen, 2001; Sáry et al., 2007). One of the aims of this project is to look for a neuronal basis for the Law of Prägnanz by recording from neurons in the area of the macaque monkey brain thought to encode subjective visual experience (Sheinberg and Logothetis, 1997).

1.2 BRAIN STRUCTURES FOR OBJECT REPRESENTATION

Wapner's agnosic artist acquired his injury from a blockage in his posterior cerebral artery (1978). This artery supplies blood to the occipital cortex and the ventral portion of the temporal lobe (Gray, 1918). Brain imaging of another famous apperceptive agnosic, patient DF, revealed that her lesion bilaterally affected a brain structure called lateral occipital cortex (LOC; James et al., 2003). Area LOC (Fig. 2A) is functionally defined in healthy subjects as being more active when viewing whole objects compared to viewing scrambled versions of those same objects (Grill-Spector et al., 2001; James et al., 2003). Thus, LOC is defined as a region involved in holistic form perception more so than the perception of local features. Just as in Wapner's stroke patient, the neuronal basis of perceiving parts and wholes appears to be dissociable (1978).

A homologous structure to human LOC in the macaque monkey is inferotemporal cortex (IT; Tsao et al., 2003). Just like LOC, IT (Fig. 2B) is more active when viewing whole objects versus viewing the scrambled parts of those objects (Desimone et al., 1984; Tsao et al., 2003; Vogels, 1999). IT lesions impair discrimination of complex visual objects (Cowey and Gross, 1970; Ungerleider and Mishkin, 1982), and neurons in this region fire vigorously in response to complex visual stimuli (Desimone et al., 1984; Gross et al., 1972; Tanaka et al., 1991). These neurons are broadly tuned, responding in a graded fashion to a large number of stimuli (Rolls et al., 1994) that share features (Brincat and Connor, 2004) or fall into the same category (Freedman et al., 2003). This selectivity to object structure and identity is accompanied by the amazing invariance across identity-preserving manipulations (Dicarlo et al., 2012), such as size, position (Ito et al., 1995), and 3D viewpoint (Ratan Murty and Arun, 2015).

IT receives most of its input from feed-forward projections along a massively convergent hierarchy of visual cortical areas (Felleman and Van Essen, 1991). Shape information is progressively refined as it travels up the ventral "what" stream (Fukushima, 1980; Riesenhuber

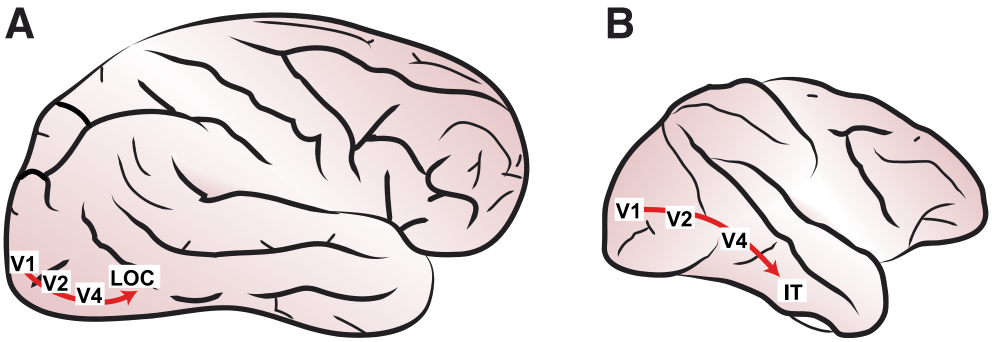

Figure 2. Visual object representation in the brain. **A**, Lateral Occipital Cortex (LOC) in the human brain. **B**, Inferior Temporal Cortex (IT) in the macaque monkey brain.

and Poggio, 1999). The ideal stimulus at each successive stage becomes increasingly complex (Guclu and van Gerven, 2015).

Starting at the very beginning of the process, light information travels from the retina to the lateral geniculate nucleus of the thalamus. The thalamus then sends projections to primary visual cortex (V1), which is the earliest cortical stage of shape processing in the ventral stream (Ungerleider and Mishkin, 1982). Neurons in V1 respond to luminance contrast at varying spatial frequency and orientation (Hubel and Wiesel, 1965). The primary purpose of this stage is to create a veridical representation of lines and edges. Lesions to this area produce a scotoma (Heinen and Skavenski, 1991; Koerner and Teuber, 1973; Miller et al., 1980).

V1 then projects to area V2, which has slightly larger receptive fields (Kobatake and Tanaka, 1994), and responds to higher level features such as illusory contours (Lee and Nguyen, 2001), border ownership (Zhou et al., 2000), and texture (Freeman et al., 2013). Thus, in V2 we already see a higher-order visual representation taking shape, rather than a compressed replication of the scene. Whereas V1 is necessary to detect low-level features such as line orientation, V2 is necessary for detecting lines defined by grouping and collinearity or texture (Merigan et al., 1993).

The step from V2 to V4 results in even further abstraction and cognitive influence emerges. Receptive fields get larger still (Kobatake and Tanaka, 1994), and the effects of border ownership become both stronger and more common (Zhou et al., 2000). V4 is also where visual attention begins to exert strong effects (McAdams and J. Maunsell, 1999; Moran and Desimone, 1985; Motter, 1994). This area is necessary for focusing spatial attention to isolate one stimulus from distractors (De Weerd et al., 1999).

1.3 INFEROTEMPORAL CORTEX AS A WINDOW TO PERCEPTION

One of the defining characteristics of IT neurons is that they have large receptive fields, averaging about 10° in diameter, always including the fovea, and centered in the contralateral hemifield (Gross et al., 1969; Op De Beeck and Vogels, 2000). When multiple objects are present in one of these expansive receptive fields, the neuron fires at a rate equivalent to the mean of the rate evoked by each object individually (Zoccolan et al., 2005). This phenomenon is called divisive normalization. Fortunately, attention rescues individual objects from divisive normalization by restoring the firing rate of the attended object (Chelazzi et al., 1998; Moran and Desimone, 1985; Zhang et al., 2011). This process may form the neuronal basis by which primates can make sense of cluttered scenes. In this way, IT neurons encode what the subject perceives, not merely what is presented to the eye.

Another line of evidence linking IT to perceptual experience comes from the field of binocular rivalry. Binocular rivalry is an experimental paradigm in which each eye is presented with a different image. Under these conditions, the eyes compete and the winner takes all such that subjects report perceiving only one of the images at a time. Traveling up the neuronal hierarchy from primary visual cortex to IT, spiking activity becomes progressively more correlated with the subject’s perceptual report (Leopold and Logothetis, 1996; Sheinberg and Logothetis, 1997). In a more naturalistic setting in which monkeys freely searched for target objects in natural scenes, IT neurons only responded to their preferred stimulus if it was actually noticed (Sheinberg and Logothetis, 2001). When playing tricks on the monkey by changing the target image mid-trial, IT neuronal firing predicted whether the animal would choose the initial target or the new target as well as the timing of this decision (Mruczek and Sheinberg, 2007).

1.4 CROWDING

What happens in IT when the perception of objects goes awry? To approach this question, we turned to a well-studied phenomenon, known as crowding, that in healthy humans looks a lot like agnosia (Strappini et al., 2017). In peripheral vision individual features are detectable and discriminable, but their locations in space seem to become unglued, creating a jumbled percept (Lettvin, 1976).

The first investigations into crowding involved showing subjects arrays of letters at varying eccentricities and with varying space between them (Bouma, 1970). As long as letters were above the acuity level for the subject they were perfectly legible when presented alone (Fig. 3). When placed in close proximity, however, the letters interfered with each other such that subjects could no longer read any of them individually. Thus, crowding is not a failure of acuity, but a failure to segregate objects. Bouma also found that the distance between letters that gave rise to crowding was proportional to the eccentricity of the whole array (1970), which became known as Bouma's Law (Pelli and Tillman, 2008). Although Bouma discovered this law using letters, it has since been demonstrated to generalize to other types of feature classes, including orientation, hue, saturation, and size (van den Berg et al., 2007), and even the features within a face (Pelli and Tillman, 2008). As long as features are similar (Kooi et al., 1994), they can crowd one another. Unfortunately, that's where the data stops being neat and tidy.

Despite great effort across a great many labs, the ensuing decades proved crowding to be "an enigma wrapped in a paradox and shrouded in a conundrum" (Levi, 2008). It does not appear to arise from surround suppression (Petrov et al., 2007). Crowding may (He et al., 1996) or may not (Freeman and E. P. Simoncelli, 2011) represent a limit on the peripheral resolution of spatial

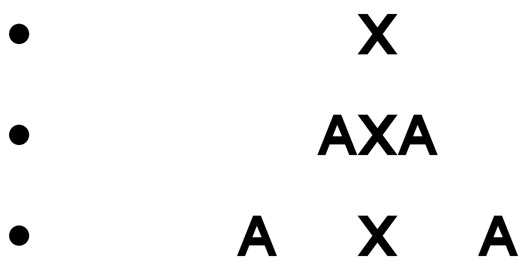

Figure 3. Crowding demonstration. On each line, fixate the black dot and try to read the letter X. Since the X is legible in isolation or with far away flankers, you clearly possess the necessary acuity for letters of this size and eccentricity. When flankers are nearby the X dissolves into a hodgepodge of features. However, If you fixate the X directly, it is clearly discriminable in all cases, demonstrating that crowding is specific to peripheral vision.

attention. Crowding appears to match a representation based on pooled summary statistics (Balas et al., 2009), which correspond to the information encoded in mid-level visual areas (Freeman et al., 2013) pooled over those neurons' receptive fields (Freeman and E. P. Simoncelli, 2011). Yet this explanation cannot capture the striking anisotropy of the fields over which peripheral features are pooled (Toet and Levi, 1992). First and foremost, it's not clear how the neuronal code gets corrupted to give rise to crowding. Answering this question requires recording from visual neurons while subjects view crowded and uncrowded displays, and that approach forms the basis for the first major thrust of this dissertation.

1.5 DO MONKEYS BEHAVE AS IF THEIR VISION IS CROWDED?

In the decades of crowding research, humans have been used exclusively. The macaque monkey provides an attractive model for studying the neuronal mechanisms behind crowding, but without having first established that they experience visual crowding it is not clear whether any insights from the monkey would generalize across species. The aim of the experiments described in chapter 2 will be to explicitly test whether rhesus macaques exhibit the behavioral hallmarks of crowded peripheral vision. A positive result establishes a monkey-friendly task that can be used to investigate how crowding arises in the brain at the level of single neurons.

1.6 HOW DO IT NEURONS ENCODE CROWDED DISPLAYS?

It has been well established that IT neurons exhibit divisive normalization when multiple objects fall within their receptive field (Carandini and Heeger, 2011; Chelazzi et al., 1998; Sripati and Olson, 2010a; Zoccolan et al., 2005). Divisive normalization weakens the representation of simultaneously-presented stimuli by averaging together the responses to individual stimuli. When those stimuli are far apart, however, divisive normalization can be overcome by spatial attention (Chelazzi et al., 1998; J. Lee and J. H. Maunsell, 2009; Reynolds and Heeger, 2009). Critically, under a divisive normalization/attention framework, only the strength of the signal is altered, not its nature.

Many psychologists who study crowding think that the underlying mechanism has to do with averaging (Greenwood et al., 2009; Harrison and Bex, 2015; Parkes et al., 2001), or that the spotlight of attention has minimum size limits such that nearby clutter may not be able to be excluded (Cavanagh et al., 1999). These hypotheses are not mutually exclusive. No prior studies have systematically investigated the connection between divisive normalization and crowding, nor have they had the temporal resolution to measure how attention relates to crowding. These are two of the hypothesis we plan to test in chapter 3.

An alternative hypothesis about the neuronal basis of crowding states that rather than weakening object representations there is a qualitative change in how crowded objects are encoded (Chastain, 1982; Freeman et al., 2012; Strasburger and Malania, 2013). In these models of crowding, the qualitative change tends to take the form of substitution of distracter features for those of the target (Chastain, 1982; Krumhansl and Thomas, 1977; Wolford, 1975), or substituting a whole distractor for the target (Strasburger et al., 1991), or a mixture of the two (Freeman et al., 2012). In any case, this account of crowding dictates that the neuronal code is not merely weakened but qualitatively altered. Attention may be involved in this model as well (Strasburger, 2005), although there is no neuronal evidence to support this claim at the present time. In chapter 3 we will also examine whether a qualitative change in the IT neuronal code occurs as inter-object spacing decreases. If so we will attempt to characterize the nature of that change.

1.7 ARE COMPOUND OBJECT PARTS SIMPLY SUMMED?

While the previous section was concerned with the interferences between different shapes in peripheral vision, next we explore how the parts of foveal objects cohere into a meaningful whole. Are wholes represented as merely the sum of their parts? Or were the Gestaltists right in asserting that the whole is something altogether different (Koffka, 1935)?

For IT neurons, when parts were spatially segregated to opposite poles of a baton, the whole was no more than the sum of its parts (Sripati and Olson, 2010a). However, when two overlapping shape outlines were presented, IT neurons exhibited response suppression compared to when the outline in the forefront was presented alone (Missal et al., 1999). An important caveat to this experiment is that since the researchers didn’t show the background shape alone it is not clear whether the reduction they observed was the result of divisive normalization (Zoccolan et al., 2005) or some other process. They also did not investigate whether the new features created by the overlapping outlines played a role in shaping neuronal responses to the whole.

The aim of the experiment described in chapter 4 will be to test whether IT neurons follow the Gestalt law of simplicity by decomposing overlapping shape outlines into the parts that seem most natural, as humans tend to do (Metzger, 1936; Pomerantz et al., 1977). An alternative hypothesis is that IT neurons represent compound images as the sum (or average) of any complete set of parts.

2.0 MACAQUE MONKEYS EXPERIENCE VISUAL CROWDING

In peripheral vision, objects easily discriminated on their own become less discriminable in the presence of surrounding clutter. This phenomenon is known as crowding. The neuronal mechanisms underlying crowding are not well understood. Better insight might come from single-neuron recording in nonhuman primates, provided they exhibit crowding. However previous demonstrations of crowding have been confined to humans. In the present study, we set out to determine whether crowding occurs in rhesus macaque monkeys. We found that animals trained to identify a target letter among flankers displayed three hallmarks of crowding as established in humans. First, at a given eccentricity, increasing the spacing between the target and the flankers improved recognition accuracy. Second, the critical spacing, defined as the minimal spacing at which target discrimination was reliable, was proportional to eccentricity. Third, the critical spacing was largely unaffected by object size. We conclude that monkeys, like humans, experience crowding. These findings open the door to studies of crowding at the neuronal level in the monkey visual system.

2.1 INTRODUCTION

In comparison to foveal vision, our view of the periphery is impoverished. This is due in part to the fact that there are fewer cones and fewer retinal ganglion cells dedicated to peripheral than to foveal locations (Wässle et al., 1989). Foveal overrepresentation persists as visual information flows through the thalamus to primary visual cortex where the amount of tissue devoted to a given eccentricity is directly proportional to acuity. Thus acuity is believed to decrease in the periphery as a direct result of coarser sampling by cones (Cowey and Rolls, 1974).

Peripheral vision suffers not only from reduced acuity but also from information loss due to crowding. The essence of crowding is that a peripheral item recognizable on its own becomes illegible when surrounded by other nearby items. Crowding is usually quantified in terms of critical spacing, the maximum distance at which surrounding clutter interferes with object recognition. Critical spacing, like acuity, scales with eccentricity (Bouma, 1970). However, unlike acuity, critical spacing possesses no well-established neuronal explanation. Mechanisms that appear to have been ruled out include surround suppression (Petrov et al., 2007) and impaired feature detection (Levi et al., 2002a; Parkes et al., 2001; Pelli et al., 2004). Pooling of feature information within neuronal receptive fields remains, however, a plausible explanation (Flom et al., 1963b). The rate at which critical spacing scales with eccentricity in humans is explicable by pooling within windows roughly the size of neuronal receptive fields in area V2 of the monkey (Freeman and E. P. Simoncelli, 2011). Yet, pooling within V2 receptive fields cannot be the full explanation because these receptive fields lack the anisotropic structure required to account for radial-tangential differences in critical spacing (Toet and Levi, 1992). Elliptical zones of integration might conceivably arise from a top-down selection process (He et al., 1996) tied to the saccadic system, in which an anisotropy for precision mirrors the anisotropy for critical spacing (Harrison et al., 2013; Nandy and Tjan, 2012).

To draw firm conclusions concerning the neuronal processes that underlie crowding will require studying the phenomenon by means of invasive techniques such as are typically employed in nonhuman primates. However, nonhuman primates will be appropriate for study only if they exhibit crowding. The aim of the present study was to determine whether they do. The universal hallmark of crowding is Bouma’s law, which in its most general form states that the critical spacing at which an object becomes unidentifiable among similar flankers depends solely on eccentricity, regardless of the nature of the object (Pelli and Tillman, 2008). It follows that critical spacing is independent of object size (Levi, 2008; Pelli et al., 2004). To determine whether monkeys exhibit crowding, we trained two macaques to perform a visual discrimination task in which we could vary the spacing, eccentricity and size of the target and the flanking distractors. We found that psychometric functions relating accuracy to target-flanker distance resembled those of humans, that critical spacing was proportional to eccentricity, and that critical spacing was largely unaffected by object size. We conclude that monkeys experience crowding. This observation paves the way for investigations into the neuronal mechanisms underlying crowding in awake, behaving monkeys.

2.2 METHODS

2.2.1 Animals and Equipment

Two adult male rhesus macaque monkeys (Macaca mulatta) were used in these experiments (monkey 1 and monkey 2). Experimental procedures were approved by the Carnegie Mellon University Institutional Animal Care and Use Committee and were in compliance with the United States Public Health Service Guide for the Care and Use of Laboratory Animals.

For behavioral testing, each monkey was seated in a primate chair with the head stabilized by a surgically implanted post. Events during each trial were controlled by Cortex software (NIMH). Visual stimuli were presented on a 17” LCD screen with 1024 x 768 pixels of resolution positioned 18” from the animal’s eyes. Eye position was tracked by an infrared system (ISCAN). The system was calibrated by requiring the monkey, at the beginning of each block of trials, to fixate a small target presented successively at four locations corresponding to the corners of a 14° x 14° square centered on the screen. Offline, the readings on each trial were converted to degrees of visual angle by performing a linear transformation based on the stored calibration voltages.

2.2.2 Task Design

On each trial, the monkey responded to presentation of a target in the right visual field by making a saccade directly above or below fixation (Fig. 4A). The targets were Sloan letters A, F, H, U, and Z (courtesy of Denis Pelli) and counterparts obtained by rotating them 90°. Each letter had an aspect ratio of one. A letter and its rotated counterpart were associated with saccades in opposite directions according to rules counterbalanced across animals. The target-saccade mapping shown in Fig. 4B was used for monkey 1 and reversed for monkey 2. Flankers, when present, consisted of Sloan letters K, P, T, and Y. Their arrangement varied from trial to trial (Fig. 4C). They were always of the same size as the target. We chose letters as stimuli because their use is common in human studies of crowding. The monkeys' prior experience with letters was fundamentally different from the experience of human subjects. To allay concern that this might affect crowding, we collected data from two human subjects using identical displays as described in a later section.

Within each block of trials, the size and eccentricity of the target were fixed. Size and eccentricity were manipulated across three experiments: Experiment 1 (size 1° at eccentricity 6°), Experiment 2 (size 0.5° at eccentricity 3°), and Experiment 3 (size 0.5° at eccentricity 6°). To characterize the effect of size and eccentricity on choice accuracy required cross-block

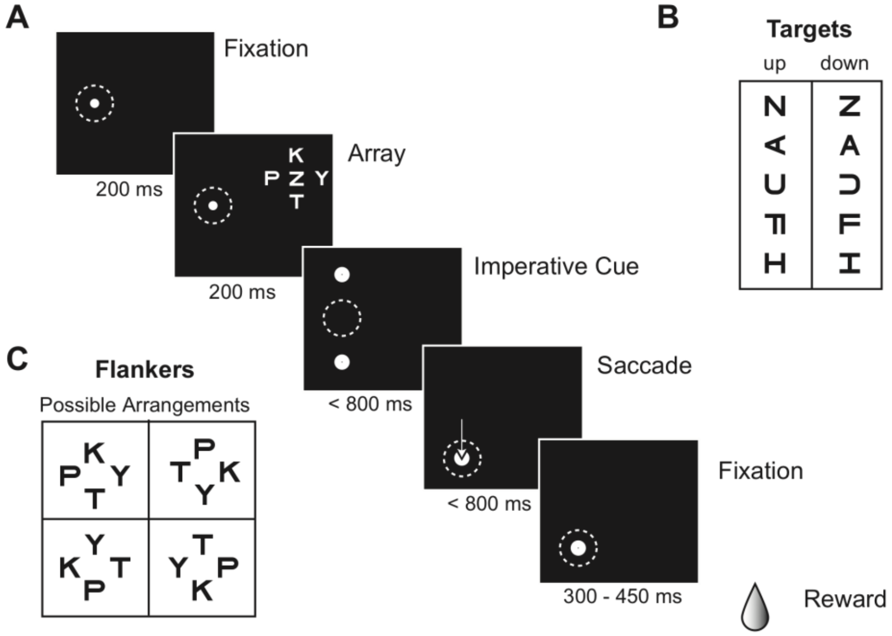

Figure 4. Task Description. **A,** Sequence of events during a typical trial. Dashed circle indicates the location of the animal's gaze during each epoch. **B,** The full set of targets with their associated responses. The association of targets with "up" and "down" responses in monkey 1, as indicated here, was reversed for monkey 2. **C,** The flankers surrounding the target could appear in any of four configurations.

comparison. To minimize the influence of random fluctuations from block to block, each monkey completed twelve blocks of trials for each experiment. Each block required performing a discrimination under 192 different conditions as described next.

Within each block, the variable of key interest was the center-to-center spacing between the target and the flankers. On a given trial, this could assume any of six values with equal likelihood. In a block involving the presentation of targets at an eccentricity of 6°, the possible spacings were 1.1°, 1.45°, 1.8°, 2.15°, 2.5° and infinity (target alone). In a block involving the presentation of targets at an eccentricity of 3°, the possible spacings were 0.6°, 0.8°, 1.0°, 1.2°, 1.4° and infinity (target alone). Other incidental factors varying within a block were fully counterbalanced against spacing. These factors included target identity, saccade direction, placement of the target in the upper or lower visual field and flanker configuration. In each block, we employed as targets two letters and their rotated counterparts. The four targets appeared with equal frequency. Saccades in upward and downward directions were demanded with equal frequency because targets associated with the two directions were equally common. In each block, the target appeared equally often above and below the zero-degree horizontal meridian. In blocks involving the presentation of targets at 3° and 6° eccentricity, the vertical displacement from the horizontal meridian was 0.5° and 1°, respectively. The flankers could appear in any of four configurations (Fig. 4C).

Full counterbalancing required assessing behavior under 192 conditions corresponding to all possible combinations of six spacings, four targets, two vertical locations and four flanker configurations. The conditions were imposed in random order with the sole exception that each combination of target, spacing and flanker occurred once in the first half of the block (when the display was centered at one vertical location) and again in the second half of the block (when the display was centered at the other vertical location). The sequence of vertical locations was upper-then-lower for half of the target-spacing-flanker combinations and lower-then-upper for the other half.

During a single block, the monkey had to complete a trial successfully under each of the 192 conditions. A trial was considered successful if the monkey made a saccade in the correct direction. This culminated in juice reward followed by an immediate advance to the next trial. A trial was aborted if the monkey's gaze deviated by more than 2° horizontally or 3° vertically from the central fixation point. In practice, the gaze rarely deviated more than 1° horizontally or 2° vertically (Fig. 5). Breaking fixation or making an erroneous response resulted in withholding of reward and a time-out of several seconds. The condition was returned to the pool from which future trials would be drawn. We based behavioral analysis exclusively on

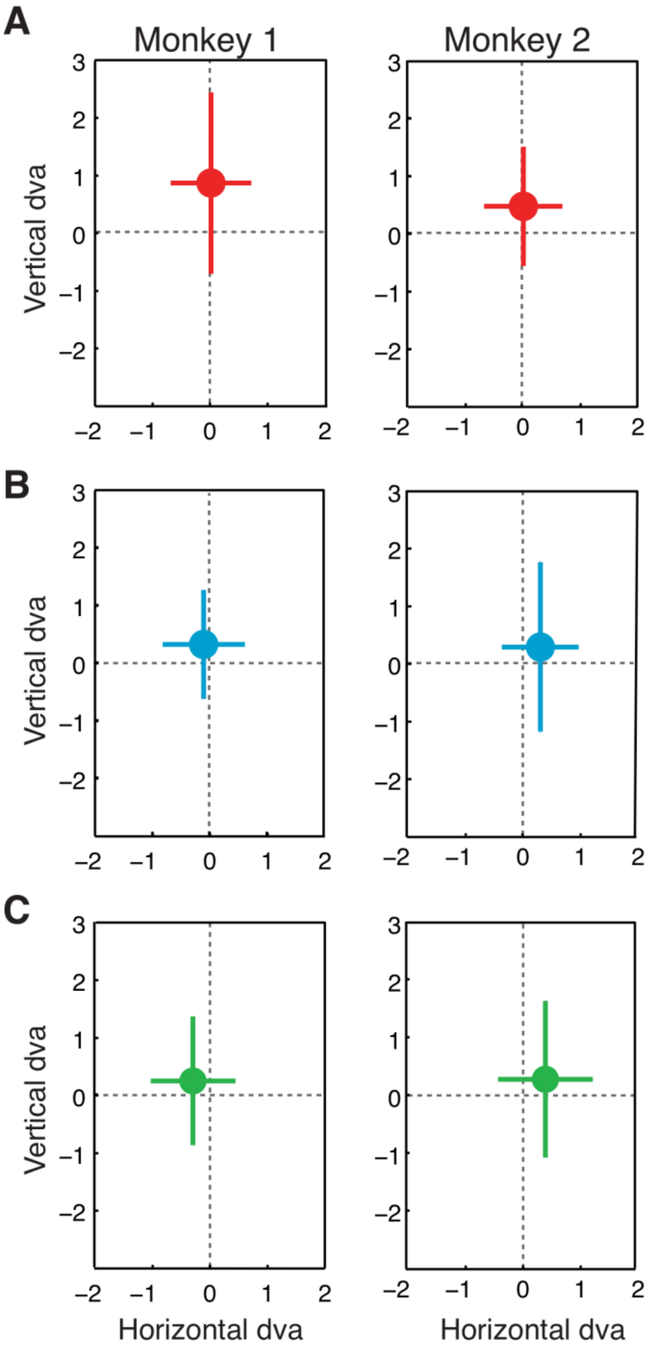

Figure 5. Gaze Angle. For each monkey, during each trial, we measured the mean eye position during the period in which the letter array was on the screen. Each panel shows the grand mean and the horizontal and vertical standard deviations of the values obtained from one monkey in one experiment. Positive values on the vertical axis indicate displacement of gaze above the fixation point. Positive values on the horizontal axis indicate displacement of gaze toward the target. **A**, Experiment 1. **B**, Experiment 2. **C**, Experiment 3.

those 192 trials in which the monkey made the first saccadic decision under a given condition without regard to whether the decision was correct or incorrect.

2.2.3 Schedule of Training and Testing

After training on basic skills such as maintaining gaze on a central fixation point and making a saccade to a suddenly appearing peripheral circle, the monkeys were introduced to a visual discrimination task in which a 1° target appearing at fixation instructed an upward or downward saccade. This phase took one month in monkey 1 and three months in monkey 2. Next, the monkeys were eased into performing the same discrimination on 1° targets presented at an eccentricity of 6°. This phase took one month in monkey 1 and three months in monkey 2. Next, they were habituated to performing in the presence of flanking distractors at various spacings by presenting the distractors at very low contrast initially and gradually increasing their contrast. This phase took one month in monkey 1 and four months in monkey 2. We continued to train the monkeys with flankers fully visible until their performance stabilized. This took four months in monkey 1 and two weeks in monkey 2. We then introduced them to task variants with 0.5° targets centered at an eccentricity of 3° or 6°. To achieve stable behavior under multiple interleaved conditions took two weeks in monkey 1 and six weeks in monkey 2. Finally, we collected behavioral data over the course of one month in each animal, interleaving blocks of trials with 0.5° objects at an eccentricity of 3°, 0.5° objects at an eccentricity of 6°, and 1° objects at an eccentricity of 6°.

2.2.4 Data Analysis

The universal measure of crowding is critical spacing, the maximal spacing at which flankers seriously interfere with target discrimination (Bouma, 1970; Pelli and Tillman, 2008). Human studies typically adopt a definition of critical spacing based on a fixed threshold halfway between chance and perfect accuracy, taking the critical spacing to be that spacing at which a psychometric function fitted to the data intersects the threshold (Chung, 2007; Kooi et al., 1994; Toet and Levi, 1992). The use of a predefined threshold would be problematic in monkeys because their performance is more erratic than the performance of humans. Even under undemanding conditions, overall accuracy rarely approaches 100%. Furthermore, overall accuracy can be affected by minor changes in a task, including, in the present instance, alterations of target size and eccentricity and the addition of flankers. It is impossible, in such cases, to distinguish between a bottom-up cause (such as poor acuity) and a top-down cause (such as poor motivation or confusion in the face of difficulty). To circumvent this difficulty, we adopted the following approach.

We defined threshold as the inflection point of a sigmoidal function fitted to points representing accuracy as a function of flanker spacing:

$$\boldsymbol{P(s) = \beta_1 + \frac{\beta_0 - \beta_1}{1 + \left(\frac{s}{s_C}\right)^{\beta_2}}} \qquad \textbf{Eq. 1}$$

where P(s) is the probability of a correct response at a given spacing (s), β_0 represents the lower asymptote, β_1 is the upper asymptote, β_2 determines the slope at the inflection point, and s_C is the inflection point. Model parameters were fitted using nonlinear least-squares (provided in the MATLAB Curve-Fitting Toolbox). We operationally defined critical spacing as the model coefficient s_C. This approach conforms in spirit to the practice in human studies of selecting a threshold midway between chance and perfect performance.

Human studies often include data from trials in which flankers were absent in the set of data to which the psychometric curve is fitted (Levi et al., 2002b; Pelli et al., 2004). The inclusion of singleton data would be problematic in monkeys due to reasons noted above. Accordingly, we based our estimate of critical spacing exclusively on trials in which flankers were present.

To be sure that the results obtained from monkeys were not an artifact of these choices with regard to the how critical spacing was measured, we repeated all analyses using two alternative models: a model in which β_1 was fixed at the performance level when no flankers were present and a model in which β_2 was fixed at the average slope across experiments. The essential findings were the same (Fig. 10). We also applied to human data the measurement procedure customized for use in monkeys. The essential findings were the same (Fig. 11).

2.3 RESULTS

Both animals were able to discriminate the target at a rate well above chance when the flankers were sufficiently far away. Both experienced a falloff in accuracy as the flankers moved closer to the target. To determine whether the pattern of falloff was consistent with expectations based on crowding, we assessed performance as a function of the eccentricity of the display and size of the letters within it.

2.3.1 Experiment 1

We first assessed performance with displays consisting of 1° letters with the target at 6° eccentricity (Fig. 6A). These parameters are within the range commonly used to demonstrate crowding in humans (Chung, 2007; Levi et al., 2002b; Tripathy and Cavanagh, 2002). As in humans (Kooi et al., 1994; Tripathy and Cavanagh, 2002; Yeshurun and Rashal, 2010) accuracy Increased as a function of target-flanker spacing in a pattern well fit by a sigmoidal function (Fig. 6C-D; goodness of fit: R^2 = 1.0 in monkey 1 and 0.99 in monkey 2). That the fit was good is not surprising inasmuch as there were five data points and the model had four free parameters. The

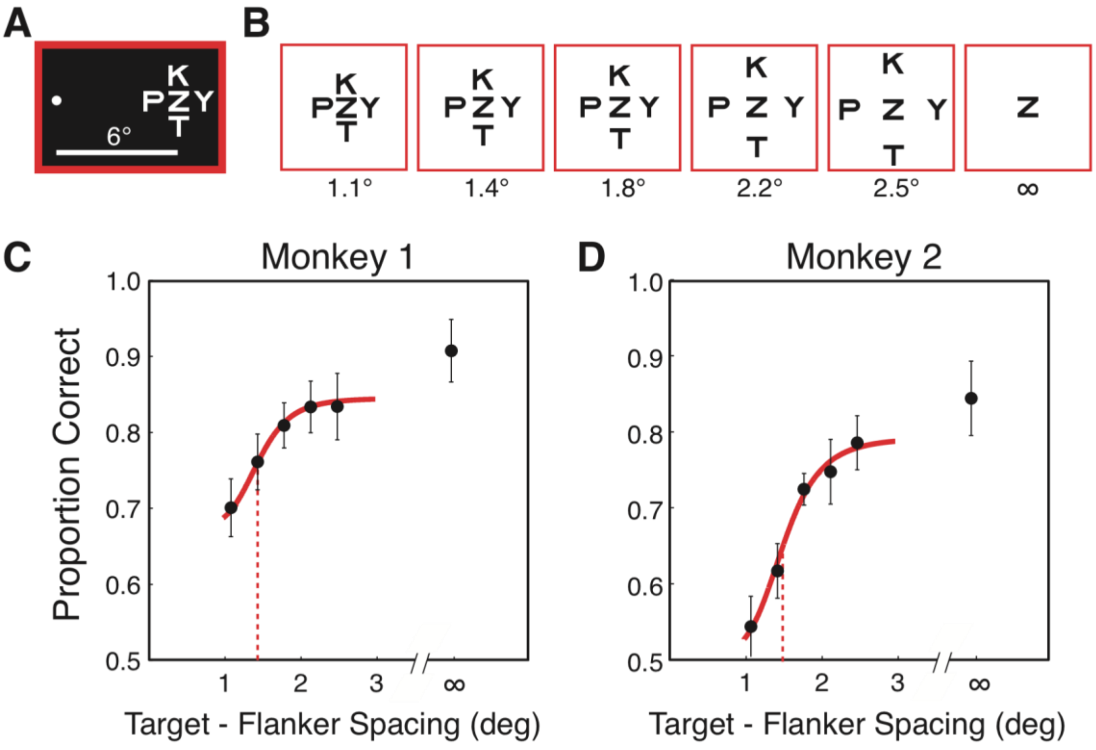

Figure 6. Experiment 1. **A,** The target was placed at an eccentricity of 6° in the right visual field. Each target and flanker subtended 1°. **B,** Flankers were spaced at five center-to-center distances from the target. In a sixth condition, flankers were absent. **C,** Accuracy as a function of spacing in monkey 1. Each data point reflects the mean over all blocks. Error bars indicate the SEM across blocks. The red curve is fit to five points representing performance when flankers were present. The dashed red line indicates the critical spacing defined as the inflection point of the fitted curve. **D,** Equivalent psychometric data for monkey 2.

purpose of curve-fitting was to allow us to establish the inflection point of the best-fit curve, which serves as an operational measure of critical spacing (Methods).

This measure possesses the virtue of being insensitive to asymptotic accuracy, which typically varies from monkey to monkey. Monkey 1 (Fig. 6C) was superior to monkey 2 (Fig. 6D) in asymptotic accuracy. Nevertheless, the measured critical spacing was virtually identical in the two animals: 1.45° in monkey 1 and 1.47° in monkey 2.

2.3.2 Experiment 2

If the critical spacing, as measured above, genuinely arose from crowding, then, according to Bouma's law (Bouma, 1970; Pelli and Tillman, 2008), it should decline with a reduction in eccentricity. To test this prediction, we scaled the display down by a factor of 0.5, reducing target eccentricity to 3° and letter size to 0.5° and contracting the range of target-flanker spacings proportionately. As in the first experiment, the data were well fit by a sigmoidal function (Fig. 7C-D; goodness of fit: $R^2 = 0.99$ in monkey 1 and 0.90 in monkey 2). The measured critical spacing was 0.82° in monkey 1 (diminished from experiment 1 by a factor of 0.57) and 0.90° in monkey 2 (diminished from experiment 1 by a factor of 0.61). These values were close to the value of 0.5 predicted from Bouma's law.

2.3.3 Experiment 3

The reduction in critical spacing from experiment 1 to experiment 2 might in principal have arisen either from scaling down the eccentricity of the display or from scaling down the size of the letters. However, classic accounts of crowding predict that critical spacing should be largely

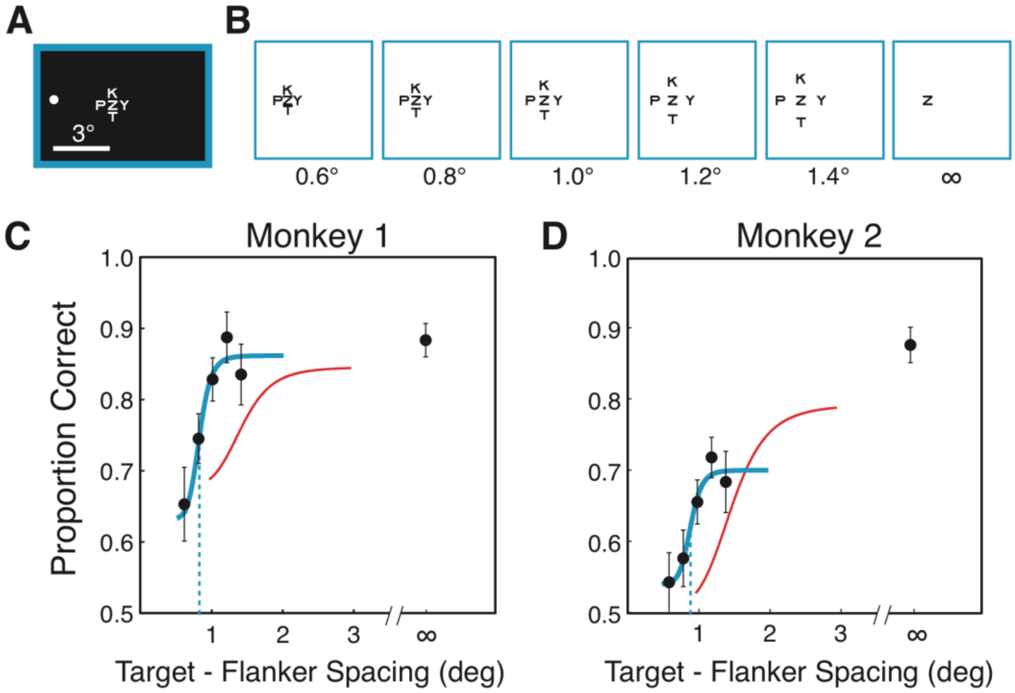

Figure 7. Experiment 2. **A,** The target was placed at an eccentricity of 3° in the right visual field. Each target and flanker subtended 0.5°. **B,** Flankers were spaced at five center-to-center distances from the target. In a sixth condition, flankers were absent. **C,** Accuracy as a function of spacing in monkey 1. Each data point reflects the mean over all blocks. Error bars indicate the SEM across blocks. The blue curve is fit to five points representing performance when flankers were present. The dashed blue line indicates the critical spacing defined as the inflection point of the fitted curve. The red curve is carried over from experiment 1 for comparison. **D,** Equivalent psychometric data for monkey 2.

independent of letter size (Levi et al., 2002b; Pelli et al., 2004; Tripathy and Cavanagh, 2002) unlike lateral masking, which should scale with size (Levi et al., 2002a; Pelli et al., 2004). To test this prediction, we presented the display at the original eccentricity of 6° while employing small letters (0.5°). The data were well fit by a sigmoidal function (Fig. 8C-D; goodness of fit: $R^2 = 0.93$ in monkey 1 and 0.90 in monkey 2). The critical spacing derived from the fitted curves was 1.42° in monkey 1 and 1.43° in monkey 2. These values were very close to values measured in experiment 1 with letters twice as large. This outcome is compatible with observations in humans viewing crowded displays. It is incompatible with an explanation based solely on lateral masking.

2.3.4 Comparison

The results of experiments 1-3 are summarized in Fig. 9. From this figure, it is clear that the critical spacing depended primarily on eccentricity (6° in experiments 1 and 3 as compared to 3° in experiment 2) and not on letter size (as indicated by the dashed white lines superimposed on the bars). As a basis for statistical comparison among the outcomes of the three experiments, we computed critical spacing for each of the twelve blocks of trials completed by each monkey in each experiment.

To determine whether eccentricity influenced the critical spacing with size held constant, we carried out an ANOVA with monkey (1 or 2) and eccentricity (3° in experiment 2 or 6° in experiment 3) as factors. In accordance with Bouma's law, there was a significant main effect of eccentricity ($p < 0.01$). The interaction between monkey and eccentricity was not significant ($p = 0.12$).

To determine whether letter size influenced critical spacing with eccentricity held constant, we carried out an ANOVA with monkey (1 or 2) and size (1° in experiment 1 or 0.5° in experiment 3) as factors. In accordance with Bouma's law, size had no significant main effect on critical spacing ($p = 0.21$). However, the interaction between monkey and size did approach significance ($p = 0.065$). Post hoc analysis revealed that this effect arose from a tendency in monkey 1 for the critical spacing to increase in conjunction with letter size (two-tailed t-test, $p = 0.01$). In monkey 1, a 100% increase in letter size produced a 25% increase in critical spacing. In monkey 2, it produced a 4% decrease. Even the effect observed in monkey 1 was far too small to support an explanation based solely on lateral masking.

Each of the aforementioned ANOVAs revealed a marginally significant main effect of monkey ($p = 0.03$ when eccentricity was a factor and $p = 0.07$ when size was a factor). This arose from a tendency for the critical spacing to be smaller in monkey 1 than in monkey 2. It is not surprising that there should have been a difference between the monkeys. Humans also show inter-individual differences in critical spacing (Toet and Levi, 1992).

In each block of trials, the monkey was required to discriminate between two pairs of targets out of the five that were available for testing (Fig. 4B). To be sure that the results generalized across targets, we sorted the data from all of the blocks by target-pair. We found that overall accuracy varied with target-identity (ANOVA with target-pair as factor, *p* < 0.01 for both M1 and M2), with the pattern of dependence differing between monkeys as if each had learned some target-pairs better than others. To test whether the dependence of critical spacing on

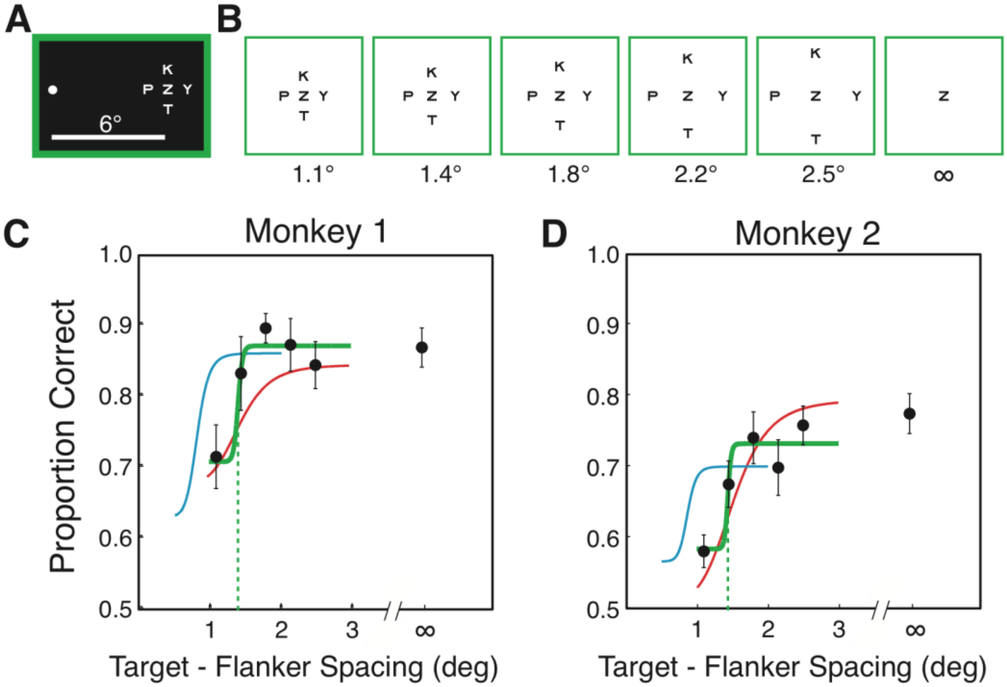

Figure 8. Experiment 3. **A,** The target was placed at an eccentricity of 6° in the right visual field. Each target and flanker subtended 0.5°. **B,** Flankers were spaced at five center-to-center distances from the target. In a sixth condition, flankers were absent. **C,** Accuracy as a function of spacing in monkey 1. Each data point reflects the mean over all blocks. Error bars indicate the SEM across blocks. The green curve is fit to five points representing performance when flankers were present. The dashed green line indicates the critical spacing defined as the inflection point of the fitted curve. The red and blue curves are carried over from experiments 1 and 2 for comparison. **D,** Equivalent psychometric data for monkey 2.

eccentricity in experiments 2 and 3 was a function of target-identity, we carried out an ANOVA on data from each monkey with target-pair and eccentricity as factors. This revealed no significant main effect of target-pair (*p* = 0.25 and 0.19 in M1 and M2), a significant main effect of eccentricity (*p* < 0.01 for M1 and M2) and, critically, no significant interaction (*p* = 0.35 and 0.94 in M1 and M2). To test whether the lack of dependence of critical spacing on size in experiments 1 and 3 was a function of target-identity, we carried out an ANOVA on data from each monkey with target-pair and size as factors. This revealed no significant main effect of target-pair (*p* = 0.27 and 0.61 in M1 and M2), no significant main effect of size (*p* = 0.08 and 0.81 in M1 and M2) and, critically, no significant interaction (*p* = 0.72 and 0.85 in M1 and M2). We conclude that the key results of the study did not depend on the identity of the targets.

Finally, we asked whether an identical approach would yield comparable results in humans. This step was motivated by two considerations. First, prior exposure to letters was different in the monkeys than it typically is in human subjects. The monkeys were intensively trained on arbitrary letter-saccade associations whereas human subjects are literate. Second, the method by which we computed the critical distance in monkeys was adapted to their particular pattern of performance. In particular, we fitted a curve to data derived exclusively from flanker-present conditions. Using stimuli and methods of analysis identical to those employed in the monkey study, we assessed

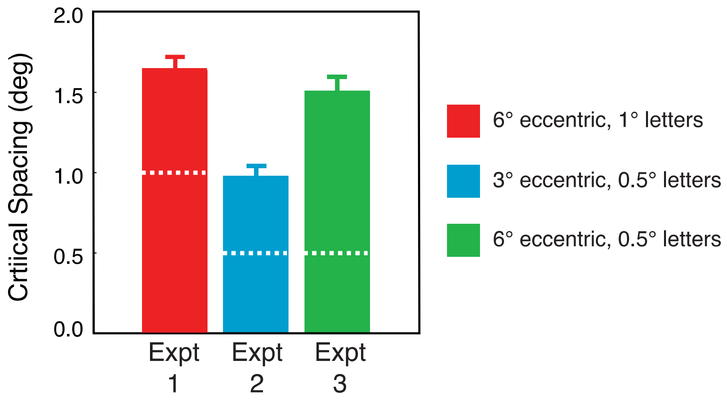

Figure 9. Comparison of critical spacing across experiments 1-3. For statistical comparison, critical spacing was estimated separately for each block of trials. The height of each bar indicates the mean of critical spacing across all 24 blocks in a given experiment. Each error bar indicates the SEM across the blocks. Critical spacing, as indicated by bar height, was almost entirely unrelated to object size, which is indicated by the dashed line superimposed on each bar.

crowding in two human participants (see Supplementary Materials: Human Experiments Paralleling the Monkey Experiments). The results were closely similar to those obtained in monkeys (Fig. 11). An ANOVA on critical spacing values with species (monkey or human) and experiment (1, 2 or 3) as factors revealed a marginally significant effect of species (p = 0.035), a highly significant effect of experiment ($p = 3.5 \times 10^{-14}$) and no interaction effect (p = 0.21). The absence of a significant interaction effect indicates that the pattern of variation across experimental conditions (the signature of crowding) did not differ between species. In one additional human experiment, we showed that expanding the set of test conditions to include smaller target-flanker spacings exerted no systematic effect on measured critical spacing (see Supplementary Materials: Human Experiment with Narrow Spacing and Fig. 12). We conclude that the observations obtained in monkeys are not an artifact either of their specialized experience with letters or of the methods of analysis necessary for characterizing their behavior.

2.4 DISCUSSION

We have carried out tests in macaque monkeys to determine whether they exhibit visual crowding. The key findings are the following. First, the ability of the monkeys to identify a peripheral target declined with decreasing distance between the target and the surrounding flankers. Second, the critical spacing scaled with eccentricity. Third, the critical spacing did not scale with object size. Together these results meet the standard diagnostic criteria for crowding (Levi, 2008; Pelli et al., 2004; Pelli and Tillman, 2008). This is the first demonstration that nonhuman primates exhibit crowding. The finding that crowding occurs in monkeys means that it will be possible in the future to investigate the neuronal underpinnings with invasive methods not generally applicable in human studies.

2.4.1 Comparison with Human Crowding Literature

The average critical spacing for monkeys in this study was 0.26 φ where φ denotes target eccentricity. For humans tested under identical conditions, the average critical spacing was 0.25 φ. These values are lower than the value of 0.5 φ reported in the classic study of Bouma (1970). Many other reports on crowding also describe values less than 0.5 φ. Pelli and colleagues (2004) note, with regard to results from a particular series of experiments, that "Bouma was right to say 'roughly' 0.5. For some of our data, this value drops as low as 0.3." Similarly, Chung et al. (2001) list, in their Table 1, prior studies yielding critical spacings as low as 0.1 φ and as high as 0.5 φ. Variability in measurements of critical spacing can arise from many sources. These include the arrangement of the elements in the display (Toet and Levi, 1992), the degree of similarity between the targets and distractors (Kooi et al., 1994), the duration of the display (Tripathy and Cavanagh, 2002), the predictability of the display's location (Yeshurun and Rashal, 2010) and the amount of prior training of the observers (Chung et al., 2007). The outcome also is dependent on the method for computing critical spacing. At present, no single method can be taken as representing a gold standard. The clipped line fit (Pelli et al., 2004) gives comparatively large readings because it yields a critical distance that lies close to the shoulder of the performance-versus-distance function. The approach of fitting a continuous curve to the data and noting the point at which it intersects a criterial performance level (Tripathy and Cavanagh, 2002) yields comparatively small readings because, with commonly used criteria, the intersection occurs on the slope rather than at the shoulder of the performance-versus-distance function. Our approach falls into the latter category.

To determine how our measurements of critical spacing compare to results obtained in previous studies, we carried out detailed analyses of data from two studies employing methods similar to ours and reporting results independently for multiple individuals. Details are provided in Supplementary Materials. Toet and Levi (1992) graphically depict, in Figure 6, the critical spacings of six individuals. Most relevant for comparison are horizontal critical spacings for displays centered on the horizontal meridian at eccentricities of 2.5°, 5° and 10°. We found, by taking measurements directly from the figure, that the critical spacing, as measured across all individuals and eccentricities, had a mean of 0.27 φ. Table 1 of Chung (2007) contains values from which the critical spacings of eight observers may be derived. We found that the critical spacings of individuals studied before intensive training, and thus comparable to our human observers, had a mean of 0.20 φ. The studies just discussed examined crowding induced by two rather than four flankers. There is little difference between critical spacings measured under two-flanker and four-flanker conditions (Pelli et al., 2004). Nevertheless, we thought it worthwhile to compare our results to those of a study that also employed four flankers. The "same colour" data points in Figure 2 of Põder (2007) represent the average performance of seven observers required to identify a 0.5° letter flanked by four other letters at an eccentricity of 3.3° - a close match to the geometry of displays employed in experiment 2 of our study. Applying our estimation procedure to these data points yielded a critical spacing of 0.23 φ (see Supplementary Materials: Analysis of Data from Prior Studies and Fig. 13). We conclude that our measurements of critical spacing are compatible with results reported previously.

In our study, both monkey and human observers exhibited some degree of flanker cost: performance even at the largest spacing tested was worse than when targets were presented in isolation as singletons. Among the monkeys, the mean percent correct for the largest spacing was

80% as compared to 86% under the singleton condition, giving a flanker cost of 6%. Among the humans, the mean percent correct for the largest spacing was 93% as compared to 98% under the singleton condition, giving a flanker cost of 5%. Flanker cost could arise from any of several sources. It might reflect genuine albeit weak crowding arising when flankers encroach on the penumbra of the crowding field. It might arise from a reduction of attention to the target induced by the presence of flankers. It might arise from basing report on a flanker rather than on the target. Although we cannot identify the origin of the flanker cost in our study, we can ask whether it is comparable to the cost observed previously under parametrically equivalent conditions. In the monkeys of our study, the largest spacing was, on average across all experiments, 1.71 times the critical spacing. In our human observers, the ratio was 1.74. On the assumption that the flanker cost arises from genuine but weak crowding occurring when the flankers lie at the edge of the crowding field, we selected for comparison a prior study in which the ratio was approximately the same (Chung, 2007). Measurements taken from Figure 4 and parameters taken from Table 1 of the cited paper indicate that the largest tested spacing was, on average across all eight observers, 1.76 times the critical spacing. Among observers in the cited study, the mean percent correct score for the largest spacing was 82% as compared to 99% under the singleton condition, giving a flanker cost of 16% with rounding error. Details are provided in Supplementary Materials. We conclude that the flanker costs observed among monkeys and humans in our study did not exceed flanker costs observed in previous studies of crowding.

2.4.2 Crowding versus Masking and Attention

Crowding differs from ordinary masking in several ways preeminent among which is the pattern of dependence on target eccentricity and size (Levi et al., 2002b; Pelli et al., 2004; Tripathy and Cavanagh, 2002). In crowding, the critical spacing scales with eccentricity independently of size, whereas, in masking, the reverse is true. To a very close approximation, the results that we obtained in both monkeys and humans conformed to the pattern expected from crowding. In monkeys there was a slight deviation from the ideal insofar as, when letters were presented at an eccentricity of 6°, the critical spacing was slightly greater if the target subtended 1° than if it subtended 0.5° (Fig. 9). This effect was not significant in the combined data but did achieve significance in a post hoc test on data from monkey 1. We do not believe that this argues against interpreting our results as due to crowding for the following reasons. First, the results for both monkeys conformed much more closely to the pattern expected from crowding than to that expected from masking. Second, the dependence on target size was weak and inconsistent. Third, it has been observed in some studies of crowding in humans that the critical spacing increases slightly as target size increases (Levi et al., 2002b; Tripathy and Cavanagh, 2002).

Distinguishing the effects of crowding from the effects of attention is a difficult challenge because the two processes may be closely related. It has been hypothesized that the critical spacing arises from a limit on the spatial resolution of visual attention (Cavanagh et al., 1999; He et al., 1996; Intriligator and Cavanagh, 2001). We know at present only that crowding and attention interact. For example, precueing the hemifield in which the display will appear (Yeshurun and Rashal, 2010) or planning an eye movement to the target (Harrison et al., 2013) shrinks the critical spacing. The effects of attention and crowding are dissociable under special circumstances, for example when subjects judge average orientation across a group of Gabor patches (Dakin et al., 2009). However, the design of our study did not allow a firm dissociation.

2.4.3 Limitations of the Present Study

The tasks presented in this paper were designed to be performed by nonhuman primates in conjunction with neurophysiological recording. The design therefore differs in some regards from methods established in psychophysical studies on humans. We tested only two subjects of each species, we presented a limited library of target and flanker letters, and we required subjects to discriminate only between two orientations of each letter. Possibly as a result of these experimental choices, our results exhibited some patterns not typically present in psychophysical studies of crowding in humans.

First, flanker cost was highly variable across tasks and across subjects. Most strikingly, monkey 2 exhibited a flanker cost of 19% during experiment 2 (Fig. 7D), but only 6% for experiment 1 (Fig. 6D) and 1% for experiment 3 (Fig. 8D). In contrast, monkey 1 exhibited a sizeable flanker cost for experiment 1 (8%, Fig. 6C) compared to smaller flanker costs for experiment 2 (2%, Fig. 7C) and experiment 3 (3%, Fig. 8C). Flanker cost varied across subjects and experiments in the human studies as well, but to a lesser degree (Fig. 11). It is possible that, with more practice, greater motivation, or better sampling of target-distractor distances, the mean and variance of flanker cost would have decreased. To minimize and stabilize flanker cost would be a desirable goal for future experiments in nonhuman primates.

We also observed considerable variability in the lower asymptote of the psychometric function and in its slope. To accommodate this variability, we adopted a curve fitting procedure that allowed these parameters to vary independently. Variability across individuals might have been less had we tailored letter size to the acuity limit of each subject (Chung, 2007). Additionally, we might have used a larger set of targets so as to prevent performance based on idiosyncratically selected diagnostic features. Steps such as these, aimed at equating to the greatest extent possible the performance of different individuals, would be a desirable feature of future experiments in nonhuman primates.

Finally, monkey 1 exhibited a significant reduction in critical spacing when letter size was decreased with all other factors held constant. This effect suggests that crowding was confounded with masking when the larger letters closely abutted. It would be desirable, in future experiments, to avoid this problem through the use of relatively small targets.

2.4.4 Questions for Future Physiological Inquiry

We know from human studies that crowding is a cortical phenomenon (Flom et al., 1963a) increasing in magnitude along the ventral stream (Anderson et al., 2012; Bi et al., 2009). However, much remains to be learned about the underlying neuronal mechanisms. Having established that monkeys experience crowding, we are now in a position to attack these questions in experiments based on neuronal recording. There are at least two fundamental outstanding questions about the neuronal correlates of crowding that could be addressed in such experiments.

2.4.4.1 Reduction in Strength

We take inferotemporal cortex as an example in terms of which to consider this question. This region is a logical candidate for the study of crowding because it is necessary for the efficient discrimination of letter-like images (Cowey and Gross, 1970) and because it contains neurons selective for letter-like stimuli (Sripati and Olson, 2010a). We speculate that crowding is manifest among inferotemporal neurons in the form of a reduction in the strength of the signal encoding the identity of the target. The neuronal representation of an image is reduced by the simultaneous presence of other images even when they are separated by distances greater than those at which crowding occurs (Zhang et al., 2011; Zoccolan et al., 2005). The reduction takes a form well described in terms of divisive normalization (Carandini and Heeger, 2011). When images are far apart, neuronal activity representing the identity of a given item can be restored to almost normal strength by allocating top-down attention to it (Chelazzi et al., 1998; Moran and Desimone, 1985; Zhang et al., 2011). When images are close together, this mechanism might fail either because bottom-up pooling of information about the target and flankers has rendered the target unavailable for independent selection (Parkes et al., 2001) or because the spatial resolution of top-down attention is limited (He et al., 1996). In either event, one would expect the neuronal signal representing the identity of the target to become progressively weaker, and attentional selection to become progressively less effective, as the elements of the display move closer to each other over the range of distances at which crowding operates. To test this prediction would require no more than recording target-discriminating neuronal activity during performance of the crowding task.

2.4.4.2 Qualitative Change

In the simplest form of the bottom-up pooling model (Parkes et al., 2001) and in any model based on the limited spatial resolution of top-down attention (He et al., 1996), one would expect inferotemporal cortex neurons to fire at a rate representing the average of rates elicited by the separate elements of the display considered individually. However, there are other models in which the weakening of the representation of the target results from a more profoundly nonlinear interaction with the flankers that cannot be explained by divisive normalization. It has been proposed that the degradation of the target representation arises from the computation of textural statistics (Balas et al., 2009), from the illusory conjunction of features (Greenwood et al., 2010; Pelli et al., 2004; Põder and Wagemans, 2007), from substitution of a flanker for a target (Freeman et al., 2012), and from a combination of these processes (Hanus and Vul, 2013). Each of these scenarios gives rise to specific predictions, potentially testable in a neuronal recording experiment, concerning the effect of the presence of the flankers on neuronal selectivity for the target.

2.5 SUPPLEMENTARY MATERIAL

2.5.1 Human Experiments Paralleling the Monkey Experiments

The results of human experiments paralleling the monkey experiments are presented in the main text and depicted in Fig. 11. We describe in this section the methodology of the human experiments.

Two right-handed adults, both female, completed tests conducted under a protocol approved by the Institutional Review Board of Carnegie Mellon University. Subject 1 is an author (EC). The other subject was unaware of the specific purpose of the experiment. All spatial conditions, including screen distance and the configuration, size and eccentricity of the stimuli,

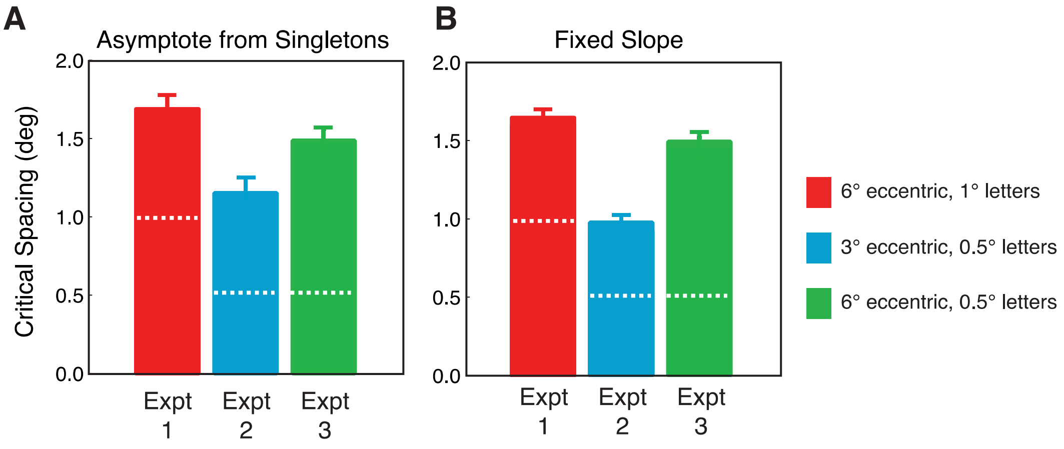

Figure 10. Varying the curve-fitting procedure had only a minimal effect on estimates of critical spacing. **A**, Critical spacing estimated using a model in which the asymptote of the accuracy curve was fixed at the accuracy measured for singleton targets. **B**, Critical spacing estimated using a model in which the slope was fixed at the average value measured across all three experiments. Conventions as for Fig. 9.

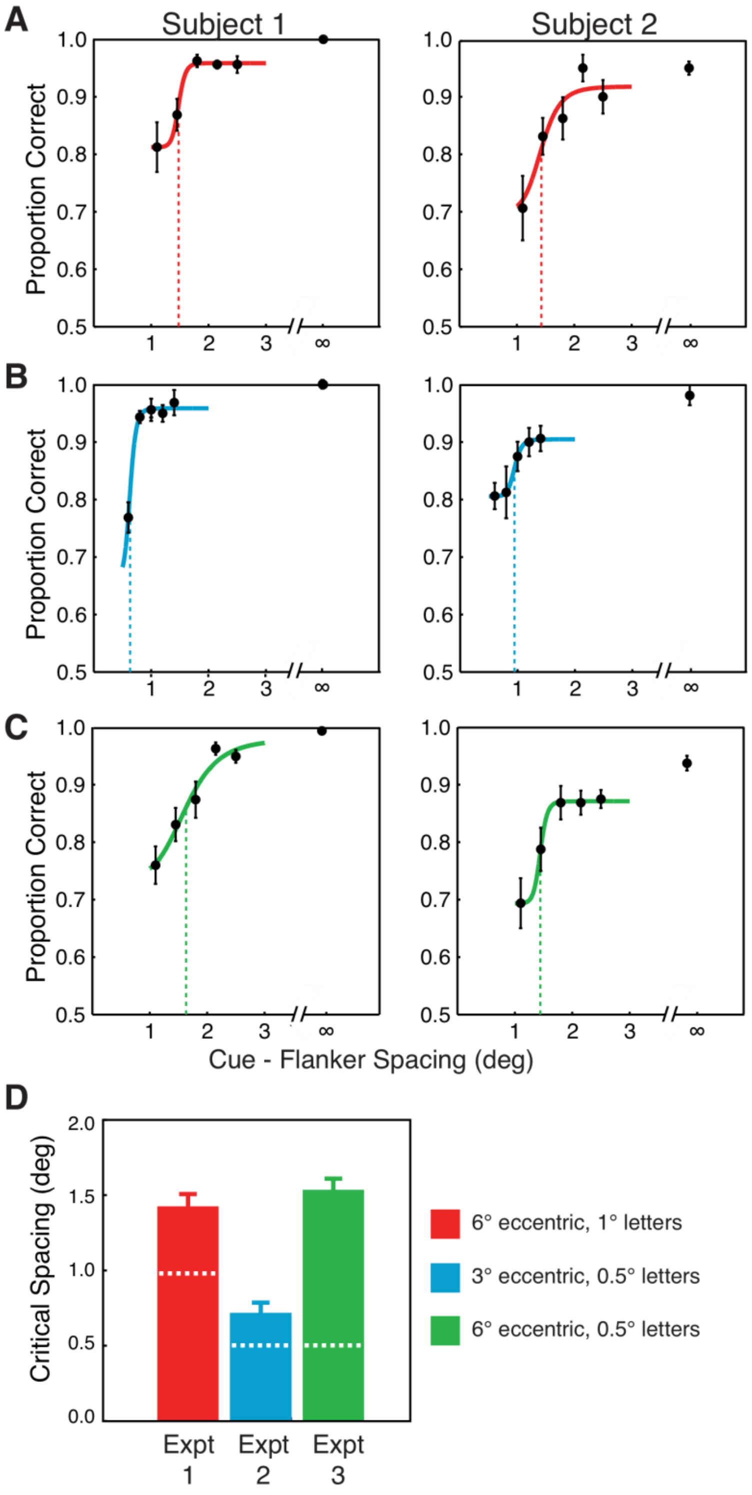

Figure 11. Results from human subjects performing the same tasks as the monkeys. **A,** Experiment 1. Conventions as in Fig. 6. **B,** Experiment 2. Conventions as in Fig. 7. **C,** Experiment 3. Conventions as in Fig. 8. **D,** Comparison across experiments 1-3. Conventions as in Fig. 9.

were identical to those imposed during the monkey experiment. Procedures for data analysis were likewise identical to those employed in the monkey experiment. We used a chin rest to enforce viewing distance and stabilize the head.

For each experiment, the subject completed five blocks. A block consisted of 192 trials conforming to the same conditions as in the corresponding monkey experiment. Each block used two pairs of targets just as in the monkey experiment. Across the five blocks, each of the five pairs of targets was employed twice. At the beginning of each block, to solidify the target-response associations, the subject performed 16 practice trials with singleton targets. These were not included in the analytic dataset.

A trial began with onset of a fixation point in the center of the screen (Fig. 4A). When the subject had attained fixation and was ready to view the array, she pressed the spacebar. This triggered the immediate appearance of the display. The duration of the display was restricted to 100 ms so as to negate any contribution from reflexive saccades. The subject reported the identity of the target by pressing a key on a keyboard. Responses on the up and down arrow keys were mapped onto the targets according to the same rules that governed upward and downward saccades in monkey 1 (Fig. 4B).   Feedback was given on each trial in the form of a click if the response was correct and silence otherwise. A trial in which the subject hit neither key was repeated later in the block. Trials in which the response was incorrect were not repeated.

2.5.2 Human Experiment with Narrow Spacing

We carried out an additional set of tests in subject 1 to determine whether measurements of critical spacing would be significantly altered by expanding the set of test conditions to include smaller target-flanker spacings.

Experiment 2 was repeated with center-to-center spacings that included, in addition to the six used previously, a narrower spacing of 0.4°. We reduced the letter size to 0.3° so that at even the narrowest spacing the targets and flankers would not touch. The principles of blocking and trial structure were the same as in the original version of the experiment. A run encompassed 32 trials at each of seven flanker spacings for a total trial count of 224. The subject completed five runs just as in the original version of the experiment. Critical spacing for subject 1 in the original experiment was 0.60° ± 0.02°. Critical spacing in the modified experiment was 0.64° ± 0.04° (Fig. 12A,C). The difference was not signficant (two-sample t-test, $t(8) = 0.79$, $p = 0.45$).

Experiment 3 was repeated with center-to-center spacings that included, in addition to the six used previously, a narrower spacing of 0.75°. The principles of blocking and trial structure were the same as in the original version of the experiment. A run encompassed 32 trials at each of seven flanker spacings for a total trial count of 224. The subject completed five runs just as in the original version of the experiment. Critical spacing for subject 1 in the original experiment was 1.55° ± 0.12°. Critical spacing for the same subject in the modified experiment was 1.19° ± 0.21° (Fig. 12B,C). The difference, although not significant (two-sample t-test, $t(8) = 1.43$, $p = 0.19$), was in the direction expected from previous demonstrations that training can ameliorate crowding within limits (Chung, 2007). To determine whether the critical spacing of subject 1 had achieved stability at this point, we repeated the experiment modified to include seven flanker spacings. The resulting measure of critical spacing (1.30° ± 0.20°) was not significantly different from the measures obtained in the original experiment (two-sample t-test, $t(8) = 1.12$, $p = 0.30$) and the previous run of the modified experiment (two-sample t-test, $t(8) = 0.45$, $p = 0.66$).

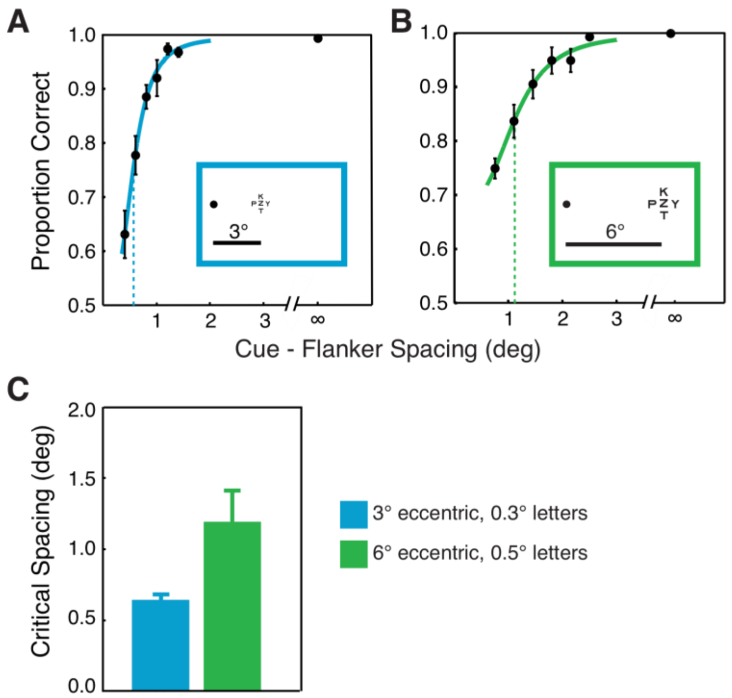

Figure 12. Results from human subject 1 collected under conditions in which the set of center-to-center spacings had been expanded to include an especially small spacing. Details in Supplementary Materials: Human Experiment with Narrow Spacing. **A**, Modified experiment 2. **B**, Modified experiment 3. **C**, Critical spacings measured in the two experiments. Conventions as in Fig. 11.

To summarize: including in the test set a condition with especially narrow spacing led to an increase in measured critical spacing in one case and a decrease in the other case, with neither change statistically significant. We conclude that including a condition with narrower spacing in the test set did not produce a systematic change in critical spacing.

2.5.3 Analysis of Data from Prior Studies

2.5.3.1 Measuring Critical Spacing in Toet and Levi (1992)

We based our analysis on Figure 6 of the cited paper. For each subject, we measured the distance from the central point to the circle at 10° eccentricity along the horizontal axis. We computed the average of the six measured lengths to get X, the distance in the figure corresponding to 10°. For each subject, we measured the horizontal extent of the interaction polygon for a target placed at 10° horizontal eccentricity (L10) and did likewise for targets placed at 5° (L5) horizontal eccentricity and 2.5° (L2.5) horizontal eccentricity. From these widths, we computed critical spacing (c) as a fraction of eccentricity (φ) using the formula given below. The term of 0.5 in the numerator adjusts for the fact that each horizontal line encompassed two center-to-center distances.

$$c\varphi = \frac{0.5L_{\varphi}}{\frac{\varphi}{10}X} \qquad \text{Eq. 2}$$

The mean across all subjects and eccentricities was 0.27 φ with a standard deviation of 0.13 φ. The results for the individual subjects are given in Table 1.

Table 1. Critical spacing relative to eccentricity in Toet and Levi (1992).

Subject	cϕ, ϕ = 2.5°	cϕ, ϕ = 5°	cϕ, ϕ = 10°
AT	0.48	0.46	0.35
JT	0.24	0.25	0.49
MS	0.08	0.12	0.20
JE	---	0.17	0.39
JW	0.27	0.18	0.18
PB	0.32	0.20	0.17

2.5.3.2 Measuring Critical Spacing in Chung (2007)

The aim of this analysis was to compute in units of eccentricity (φ) the pre-test spatial extent of crowding provided for each subject in Table 1 of the cited paper. The eccentricity of the target was always 10°. The spatial extent of crowding (S) is given in units of letter size. Letter size for each subject was 1.4 times the critical print size. The critical print size (P, in degrees) is provided for each subject in Table 1. We used the following conversion formula:

$$c\varphi = \frac{1.4SP}{10} \qquad \text{Eq. 3}$$

The mean across all subjects was 0.20 φ with a standard deviation of 0.05 φ. The values for the individual subjects are provided in Table 2.

Table 2. Critical spacing relative to eccentricity in Chung (2007).

Subject	S	P	cϕ
AS	0.97	0.95	0.13
LG	1.08	1.38	0.21
MM	1.08	1.64	0.25
NV	1.26	1.4	0.25
SA	1.2	1.26	0.21
SU	1.33	1.5	0.28
SW	1.12	0.97	0.15
TN	0.93	1.17	0.15

2.5.3.3 Measuring Critical Spacing in Põder (2007)

The aim of this analysis was to compute in units of eccentricity (φ) the spatial extent of crowding for "same colour" data presented in Figure 2 of the cited paper. We measured the height of each point on the plot and and linearly transformed each height into the appropriate units. Then we applied our inflection point method for calculating critical spacing (Eq. 1). The resulting critical spacing was 0.23 φ. The curve fit is depicted according to our conventions in Figure 13.

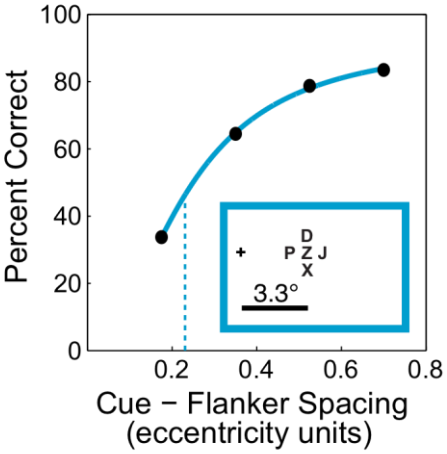

Figure 13. Application of our method for estimating critical spacing to data from Fig. 2 of Põder (2007). Inset indicates the geometry of the display in the task on which the figure is based. Conventions as in Fig. 12.

2.5.3.4 Measuring Distractor Cost in Chung (2007)

We quantified distractor cost by taking measurements from Figure 4 of the cited paper. In the plot for each subject, we inferred the percent correct in the absence of distractors (A) from the height of the horizontal dashed line by taking the ratio of this height to the height of the y-axis. Likewise, we inferred the percent correct in the presence of distractors at the largest spacing tested (D) from the height of the rightmost open symbol. The distractor cost was given by A-D. The mean of A-D across all subjects was 16.1% with a standard deviation of 7.0%. The values for the individual subjects are provided in Table 3.

Table 3. Widest spacing versus critical spacing in Chung (2007).

Subject	A	D	A-D
AS	99.8	90.2	9.6
LG	99.8	90.2	9.6
MM	99.5	76.0	23.5
NV	95.2	85.4	9.7
SA	99.8	80.4	19.4
SU	94.8	71.0	23.8
SW	99.5	76.0	23.5
TN	100.0	90.2	9.8

2.5.3.5 Widest Spacing versus Critical Spacing in Chung (2007)

The widest spacing tested (W, in units equal to 1.4 times the critical print size) was 2.0 in all subjects except AS, in whom it was 1.6. The critical spacing (S, in units equal to 1.4 times the critical print size) is given for each subject in column 1 of Chung's Table 1. To express the widest spacing tested as a ratio of the critical spacing, we computed W/S. The mean of W/S across all subjects was 1.76 with a standard deviation of 0.20. The values for the individual subjects are provided in Table 4.

Table 4. Comparing critical spacing to widest spacing in Chung (2007).

Subject	W	S	W/S
AS	1.6	0.97	1.65
LG	2.0	1.08	1.85
MM	2.0	1.08	1.85
NV	2.0	1.26	1.59
SA	2.0	1.2	1.67
SU	2.0	1.33	1.50
SW	2.0	1.12	1.79
TN	2.0	0.93	2.15

3.0 CROWDING CONFUSES THE NEURONAL CODE

Peripheral vision suffers from information loss, not only from acuity, but by the mysterious phenomenon known as crowding, in which the presence of clutter causes perfectly recognizable objects in isolation to become a jumbled hodgepodge. Although crowding has been studied extensively for the last half century in humans, methodological limitations have prevented the field from asking questions about the fundamental brain mechanisms behind crowding. Of particular interest is how crowding affects the neurons associated with object recognition. We recorded single neuron responses from macaque monkeys in a series of experiments, each designed to get at different aspects of crowding. First, we discovered that crowding qualitatively altered neuronal preferences, thus confusing the object code. Next, we observed that aspects of this code change persisted even when the entire display was scaled up to escape crowding. Finally, we showed that even non-adjacent parts of crowded objects interact.

3.1 INTRODUCTION

In primates, the visual representation of peripheral space is crowded, meaning that an item recognizable on its own becomes unintelligible when imbedded in clutter. Crowded objects don’t disappear or blur, but rather they devolve into a jumble of features. Because the vast majority of visual space is peripheral and objects rarely appear in isolation, crowding is ubiquitous. Crowding affects everything from reading speed (Pelli et al., 2007) to avoiding obstacles while driving (Whitney and Levi, 2011).

Fortunately, crowding provides a natural experiment in which feature detection and feature integration are decoupled (Pelli et al., 2004). These early stages of visual processing are critical for laying the foundations for object representation. Therefore, the study of crowding may provide new insights about how objects are represented in the brain.

The computational mechanism underlying crowding has been variously characterized as averaging of visual signals for nearby stimuli (Greenwood et al., 2009; Harrison and Bex, 2015; Parkes et al., 2001), confusion between target and distractor elements (Chastain, 1982; Freeman et al., 2012; Strasburger and Malania, 2013), and reducing objects to texture (Balas et al., 2009; Freeman and E. P. Simoncelli, 2011; Lettvin, 1976). At the implementation level, regions over which crowded stimuli interact could be set by receptive fields (Flom et al., 1963b; Freeman and E. P. Simoncelli, 2011; Keshvari and Rosenholtz, 2016), cortical distance (Mareschal et al., 2010; Pelli, 2008), or the spotlight of spatial attention (Chen et al., 2014; He et al., 1996). Even the motor system gets a piece of the action with the hypothesis that crowding results from limitations in the accuracy of saccadic eye movements (Harrison et al., 2013; Nandy and Tjan, 2012).

Largely, theories of crowding have been shaped by psychophysical experiments, which cannot get at neuronal mechanisms directly. To delve into the brain-based origins of crowding, several groups have recently turned to brain imaging (Anderson et al., 2012; Bi et al., 2009; Chen et al., 2014; Fang and He, 2008; Millin et al., 2013) and EEG (Chen et al., 2014; Ronconi et al., 2016). From these studies, we have learned that crowding appears to be a multi-stage process (Ronconi et al., 2016) that is evident as early as V1 (Millin et al., 2013) and increases along the ventral visual hierarchy (Anderson et al., 2012), peaking in LOC (Herzog et al., 2015). Despite this success with mapping where crowding is apparent in the brain, the spatial and temporal resolution of fMRI and EEG are not sufficient to capture how individual neurons within those brain regions encode crowded and uncrowded stimuli.

To draw firm conclusions concerning the neuronal underpinnings of crowding will require studying the phenomenon by means of invasive techniques such as are typically employed in nonhuman primates. Having shown previously that nonhuman primates exhibit the same behavioral hallmarks of crowding as humans during a task conducive to neurophysiology (Crowder and Olson, 2015), we are poised to be the first to directly explore the crowding phenomenon at the level of single neurons.

The aim of the present study was to determine how crowding affects the neuronal code underlying visual object representation in the pinnacle of the ventral stream, a region known as inferotemporal cortex (IT). We recorded individual neurons IT of two macaque monkeys while they either passively viewed or discriminated between peripheral letters under various degrees of crowding. Since IT is known to be important for object recognition (Ungerleider and Mishkin, 1982) and tracks perception (Sheinberg and Logothetis, 1997), the effect of crowding on the neuronal code for peripheral objects should be evident at the level of IT.

Specifically, we tested two specific hypotheses. First, crowding could impair object recognition by weakening neuronal selectivity, as predicted by averaging models (Greenwood et al., 2009; Harrison and Bex, 2015; Parkes et al., 2001). A second possibility is that crowding could qualitatively change the neuronal code to confuse downstream neurons, as predicted by feature substitution models (Chastain, 1982; Freeman et al., 2012; Strasburger, 2005).

What we discovered first was that crowding both qualitatively altered neuronal preferences and quantitatively weakened selectivity. This qualitative change could be decomposed into main effects and interaction effects. Only qualitative changes to interaction effects seemed to vary with absolute spacing – rather than relative spacing – the way that crowding does behaviorally. Finally, we showed that even non-adjacent parts of crowded objects interacted.

3.2 MATERIALS AND METHODS

3.2.1 Animals and Equipment

Two adult male rhesus macaque monkeys (Macaca mulatta) were used in these experiments (monkey 1 and monkey 2). Experimental procedures were approved by the Carnegie Mellon University Institutional Animal Care and Use Committee and were in compliance with the United States Public Health Service Guide for the Care and Use of Laboratory Animals. Before the recording period, each monkey was surgically fitted with a cranial implant and headpost (Crist Instrument). After initial training, a 2 cm-diameter vertically oriented cylindrical recording chamber (Crist) was implanted over the left hemisphere in both monkeys. In both animals MRI brain scans were used to position the chamber mediolaterally above the superior temporal sulcus and rostrocaudally above anterior medial temporal sulcus.

For behavioral testing, each monkey was seated in a primate chair with the head stabilized using the headpost. Events during each trial were controlled by Cortex software (NIMH). Visual stimuli were presented on a 17” LCD screen with 1024 x 768 pixels of resolution positioned 18” from the animal’s eyes. The precise time at which images appeared on the screen was recorded using a photodetector circuit (designed by NIMH) and built in-house. Eye position was tracked by an infrared system (ISCAN). The system was calibrated by requiring the monkey, at the beginning of each block of trials, to fixate a small target presented successively at four locations corresponding to the corners of a 14° x 14° square centered on the screen. Offline, the readings on each trial were converted to degrees of visual angle by performing a linear transformation based on the stored calibration voltages.

After the initial behavioral training was complete, Each day’s recording session would begin with the insertion of a varnish-coated tungsten microelectrode with an initial impedance of 1.0 M at 1 kHz (FHC) into the temporal lobe through a transdural guide tube advanced using a hydraulic microdrive (Narishige). When mapping a new track electrodes were lowered to a depth such that its tip was 10 mm above the superior temporal sulcus, as estimated from MRI images of each animal’s brain. Using a grid inside the chamber with 1mm spacing between holes (Crist) the electrode could be advanced reproducibly along the same tracks day to day. The action potentials of a single neuron were isolated online by means of a commercially available spike-sorting system (Plexon). All threshold-crossing waveforms were recorded during the experiments. The threshold was chosen such that some noise and multiunits were recorded along with the single unit isolations. Final spike sorting was performed manually offline.

Neurons were probed first with a set of 32 colorful photographs of objects to see whether they were visually-responsive. If so, they were further tested with relevant stimuli from the specific experiments, which were presented foveally and in isolation during this initial phase to ensure that neurons and stimuli were selected in a way that remained agnostic to experimental questions. Stimuli were chosen to maximize both the mean and the range of firing rates evoked by the stimulus set.

3.2.2 Tasks and training

Monkeys were trained to fixate and discriminate peripheral letters as described previously (Crowder and Olson, 2015). Each animal engaged in three distinct tasks, which are described separately in the Results. All experiments were run as separate blocks with trials presented in a pseudorandom order. Incomplete or incorrect trials were repeated at random later in the block.

Although monkey 2 performed adequately on the similar behavioral tasks employed in our prior study (Crowder and Olson, 2015), following chamber implantation surgery this animal could not be motivated to report perceived stimulus identity. As a result, he only engaged in passive viewing in the present study. In all cases where monkey 1 performed a behavioral task while monkey 2 merely fixated, we performed analyses separately for the two animals to ensure that this behavioral difference did not meaningfully affect the results.

3.2.3 Data Analysis

Neurons were only considered for analysis if for all trials combined they fired at a significantly higher rate in the period 70 to 270 ms after stimulus onset compared to the baseline period, -100 to 50 ms. Significance was assessed using a paired Student’s t-test with $\alpha = 0.05$.

Peristimulus time histograms (PSTHs) were computed by aligning spikes to the time the photodetector registered the onset of the stimulus, counting the spikes in each 1ms bin, and convolving with an alpha function designed with time constants measured for excitatory post-synaptic potentials (1ms for growth, 20ms for decay) (Hanes et al., 1995). This approach has two advantages over smoothing with a gaussian kernel. First, it is a causal filter so it does not smear the timing of spikes to earlier timepoints. Second, it avoids the arbitrariness of choosing the standard deviation of the gaussian kernel because the time constants used in the alpha function are derived from physiology.

Before combining the responses of individual neurons into population responses for any of the analyses used in this study, neuronal preferences were determined using a leave-one-out cross-validation procedure. All trials but one were used to determine the neuronal preference, and depending on the stimulus presented on the held-out trial it could be labeled as either "preferred" or "non-preferred." This procedure was iterated until all trials were labeled. The advantage of this method is that it makes full use of all trials without introducing bias that would cause pure noise to appear selective. Unless otherwise noted, spike counts were always taken from the period 70ms to 270ms after stimulus onset, which captures the typical latency and transient burst of inferotemporal cortex neurons.

We used correlation analysis to compare how consistently populations of neurons responded across different conditions. To put these correlations in context of a theoretical ceiling we also computed the correlation across neurons within the same condition. Naturally, this required a split halves approach. To correct for reducing the number of paired observations by half, we used the Spearman-Brown prediction formula (Brown, 1910; Spearman, 1910).

To understand how letters interfere with one another during crowding, we divided neuronal responses into pairwise main effects and interaction effects. Main effects measure the strength of firing rate selectivity between different stimulus identities, regardless of other factors. In contrast, interaction effects measure how the identity of one stimulus affects the spike rate evoked by another. The simplest way to explain pairwise main and interaction effects is to imagine a two-by-two matrix in which columns represent the levels of the top element and rows represent the levels of the bottom element. Each box contains a mean firing rate for that particular set of top and bottom elements.

The main effect of the top element is then simply computed by taking the difference between the column averages and the main effect of the bottom element is equal to the difference between the row averages. Then the top and bottom element main effects can be averaged together to get a single main effect measurement. For interaction effects, we first calculated the average firing rate along the two diagonals of the two-by-two matrix, and then took the difference. Main and interaction effects were computed separately for each neuron. To combine these measurements into population metrics, we again used a leave-one-out approach to find preferred and non-preferred stimuli. For each held out trial, we used the remaining trials to find the preferred and non-preferred top element, bottom element, and diagonal. Then once all the trials were labeled in this way, we computed the main and interaction effects for all the trials together and averaged these across neurons.

3.3 RESULTS

Having demonstrated previously that macaque monkeys experience the hallmarks of visual crowding (Crowder and Olson, 2015) we sought to explore how crowding affects the inferotemporal cortex neuronal code in these same animals. To do this, we employed three complementary experiments, each designed to get at a different aspect of how crowding might impair neuronal object representations. Each will be discussed separately in the following sections.

3.3.1 Experiment 1

Since we are the first to pursue crowding at the single neuron level, our first experiment was designed to be a slightly simplified version of one of the tasks we used to demonstrate the crowding effect behaviorally in these animals (Crowder and Olson, 2015). Having both single unit spiking data and behavior from the same conditions we can then ask how the decline in behavioral accuracy characteristic of crowding correlates with changes in the neuronal code for the target of visual discrimination. There are two alternative hypotheses we aim to test. First, it could be that crowding diminishes the neuronal selectivity for the target through the well-characterized process of divisive normalization (Zoccolan et al., 2005). When stimuli are far apart divisive normalization can be overcome by attention (Chelazzi et al., 1993), but it remains a mystery whether this push/pull between attention and divisive normalization comes into play for crowded displays. A second hypothesis states that crowding disrupts the neuronal code for the target by qualitatively altering neuronal tuning. Under this regime, behavior would be disrupted not by weak target signals, but rather by confusing signals. These hypotheses make distinct predictions regarding strength of selectivity and consistency of the neuronal code between crowded and uncrowded displays and these are precisely what we aim to test in experiment 1.

3.3.1.1 Task and Stimuli Design

To ensure that our results would be pertinent to the phenomenon of crowding we kept the task essentially the same as experiment 1 in our previous paper (Crowder and Olson, 2015), except that we simplified it slightly by reducing the number of possible spacings between targets and flankers from five to three. This served to make the number of trials more manageable to complete while holding a stable neuronal isolation.

On each trial, monkey 1 had to determine the identity of a target in the right visual field and indicate his choice by making a saccade directly above or below fixation (Fig. 14A). Monkey 2 viewed the same displays but was not required to make saccades. The targets were Sloan letters A, F, H, U, and Z (courtesy of Denis Pelli) and counterparts obtained by rotating them 90°. A letter and its rotated counterpart were associated with saccades in opposite directions. The target–saccade mapping is shown in Figure 14B. Flankers, when present, consisted of Sloan letters K, P, T, and Y. Their arrangement varied from trial to trial (Fig. 14C). Targets and flankers were 1° and had an aspect ratio of 1. The array was always centered at 6° eccentricity.

The variable of interest was the center-to-center spacing between the target and the flankers (Fig. 14D). On a given trial, this could assume any of four values with equal likelihood, including 1.1°, 1.8°, 2.5°, and infinity (target alone, hereafter referred to as “singleton”). Other incidental factors were fully counterbalanced against spacing. These factors included target identity, saccade direction, placement of the target in the upper or lower visual field, and flanker configuration.

In each session, we employed as targets two letters and their rotated counterparts. The four targets appeared with equal frequency. Saccades in upward and downward directions were demanded with equal frequency because targets associated with the two directions were equally common. Likewise, in each block, the target appeared equally often 1° above and below the 0° horizontal meridian. The flankers could appear in any of four configurations (Fig. 14C).

Full counterbalancing required 128 conditions, corresponding to all possible combinations of four spacings, four targets, two vertical locations, and four flanker configurations. The conditions were imposed in random order with the sole exception that each combination of target, spacing, and flanker occurred once in the first half of the block (when the display was centered at one vertical location) and again in the second half of the block (when the display was centered at the other vertical location). The sequence of vertical locations was upper-then-lower for half of the target–spacing–flanker combinations and lower-then-upper for the other half.

The monkey had to complete eight trials successfully under each of the 128 conditions. A trial was considered successful if the monkey made a saccade in the correct direction (monkey 1) or maintained central fixation (monkey 2). Correct trials culminated in a juice reward followed by

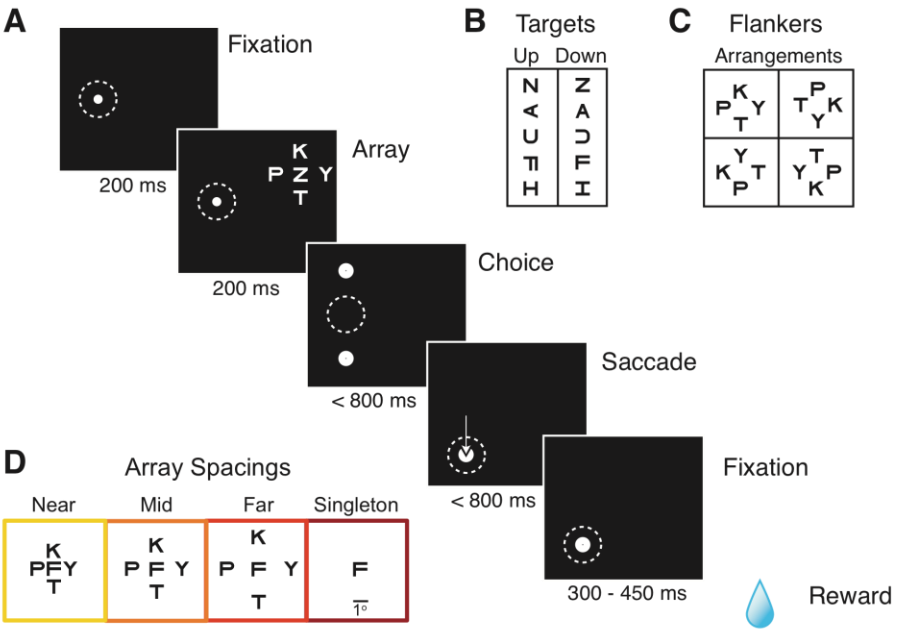

Figure 14. Experiment 1 task and stimulus design. **A,** The timing of task events for the discrimination task performed by monkey 1. The task for monkey 2 was the same except omitting the choice and saccade epochs. **B,** Half of the targets in the discrimination task corresponded to upward saccades while the other half were associated with downward saccades. **C,** The flankers surrounding the target always consisted of the same four Sloan letters, but they could be arranged in four different ways. **D,** The spacing between the target and flankers could be one of three values and targets were also presented as singletons.

an immediate advance to the next trial. A trial was aborted if the monkey's gaze deviated by more than 2° horizontally or 3° vertically from the central fixation point during fixation periods or if they made a saccade to the stimulus array thereafter. These types of errors were rare and generally the animals maintained fixation tightly on the central spot both horizontally (mean and standard deviation 0.05° ± 0.67° for monkey 1 and 0.08° ± 0.68° for monkey 2) and vertically (0.26° ± 0.88° for monkey 1 and 0.37° ± 0.97° for monkey 2). Breaking fixation or making an erroneous response resulted in withholding the reward and a time-out of several seconds. The failed condition was returned to the pool from which future trials would be drawn. We based neuronal analysis exclusively on the correct trials.

3.3.1.2 Crowding Reduces the Strength of the Singleton Code

As we showed previously (Crowder and Olson, 2015), target identification accuracy declined as a function of flanker spacing (Fig. 15A), which could be well fit with a two-parameter logit function (goodness-of-fit test, $\chi^2(29) = 0.004$, $p = 0.95$). Accuracy was significantly lower when flankers were nearby compared to mid spacing (Wilcoxon signed rank test, $z(29) = 3.84$, $p = 1.2 \times 10^{-4}$), far spacing flankers (Wilcoxon signed rank test, $z(29) = 4.54$, $p = 1.3 \times 10^{-6}$), or singletons (Wilcoxon signed rank test, $z(29) = 4.70$, p 2.6×10^{-6}). There were no other significant pairwise differences, so we consider the near spacing displays to be crowded whereas the mid and far spacing displays to be uncrowded.

The purpose for incorporating behavior is so that we can map our neuronal results onto the perceptual accuracy. Because only near flankers significantly deteriorated accuracy, a true neuronal correlate of the crowding phenomenon should severely impair neuronal target selectivity only at when flankers are near the target, while sparing other conditions. Selectivity impairments can take two primary forms: quantitative or qualitative. We will tackle the quantitative change first. Here we define neuronal preferences on the basis of singleton conditions, and operationalize selectivity as the mean firing rate difference between the most and least preferred targets.

We recorded 38 visually-responsive neurons from monkey 1 and 33 visually-responsive neurons from monkey 2. For each neuron, we designated the most preferred (Fig. 15B) and least preferred (Fig. 15C) target on the basis of the singleton trials using a leave-one-out procedure (see Methods), which enabled us to compare selectivity between singletons and all spacing conditions fairly. Time-varying selectivity was computed for each spacing independently (Fig. 15D). As letter spacing decreased, neuronal selectivity significantly decreased as well (linear regression, $F(71) = 29.84$, $p = 6.8 \times 10^{-7}$). Pairwise analysis revealed that selectivity was lower when flankers were near, compared to singleton (Wilcoxon signed rank test, $z(71) = 4.00$, $p = 3.2 \times 10^{-5}$), far ($z(71) = 3.18$, $p = 0.002$), or mid ($z(71) = 3.60$, $p = 1.6 \times 10^{-4}$) flanker spacings.

Because only monkey 1 was engaged in the discrimination task we wanted to be sure that these results were consistent across animals, so we repeated the previous analyses on each animal separately (Fig. 15E,F). Again, target selectivity significantly decreased with spacing for monkey 1 (linear regression, $F(38) = 7.1$, $p = 0.009$) as well as for monkey 2 (linear regression, $F(33) = 14.3$, $p = 3.2 \times 10^{-4}$). Likewise, the pairwise comparisons remained significant as well. For monkey 1, selectivity was still significantly lower for near flankers compared to singletons (Wilcoxon signed rank, $z(38) = 2.75$, $p = 0.003$) and mid spacing flankers ($z(38) = 2.15$, p = 0.02). For far flankers the difference was borderline significant ($z(38) = 1.51$, p = 0.06). For monkey 2, selectivity was still significantly lower for near flankers compared singletons (Wilcoxon signed rank test, $z(33) = 3.24$, $p = 6.0 \times 10^{-4}$), mid flankers ($z(33) = 2.18$, $p = 0.01$), and far flankers ($z(33) = 2.38$, $p = 0.008$).

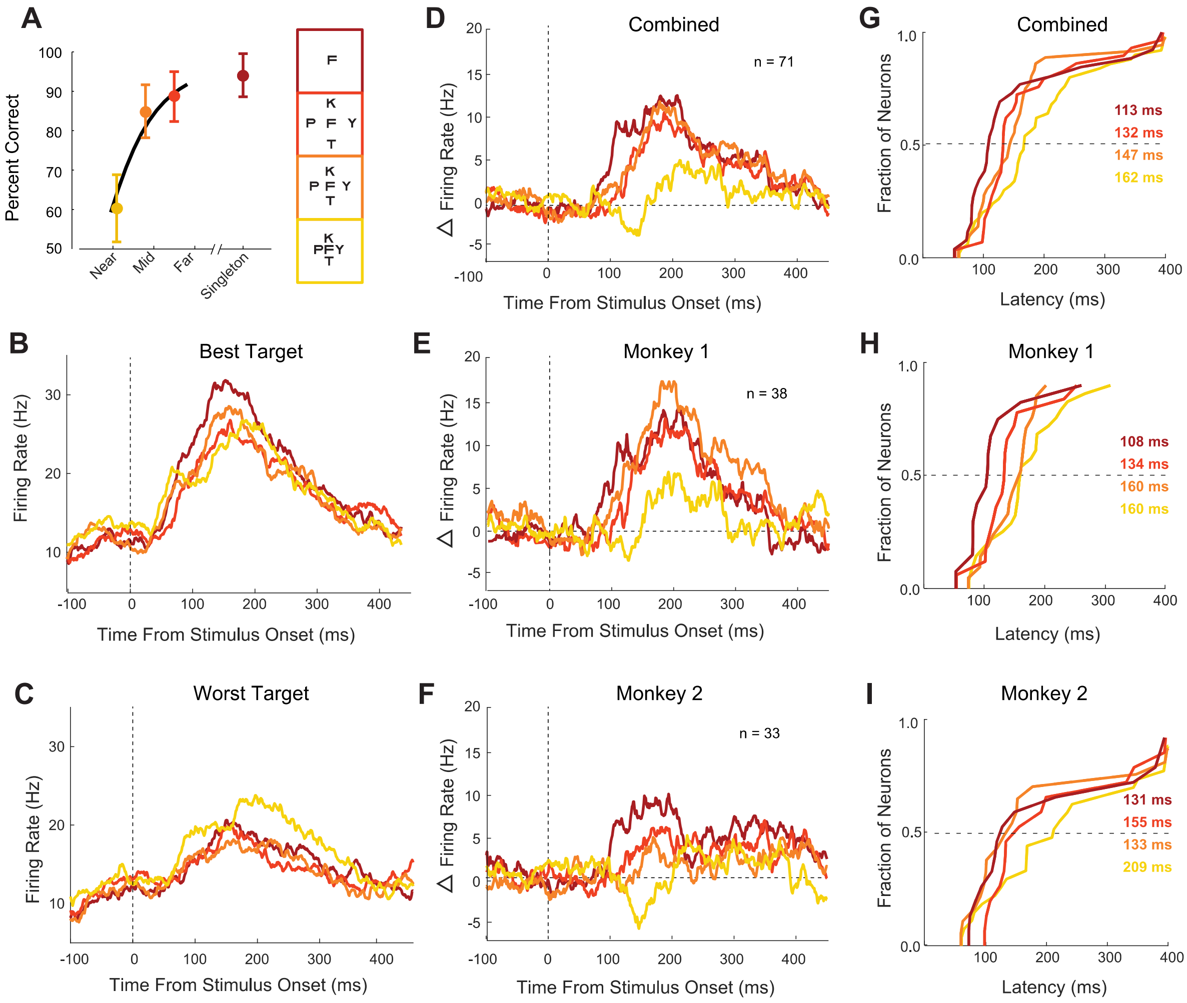

Figure 15. Experiment 1. Crowding reduces the strength of selectivity. **A**, Behavior on the discrimination task by monkey 1. **B**, Population firing rate responses to the preferred target averaged across both animals. **C**, Population responses to the non-preferred target averaged across both animals. **D**, Selectivity (difference in firing rate) between the best and worst targets averaged across both animals. **E**, The selectivity between best and worst targets for monkey 1. **F**, Selectivity between best and worst targets for monkey 2. **G**, Empirical cumulative distribution function for latency of target selection across the population of neurons, as defined by the time the delta function (shown as an average over the whole population in **D**) reached half-height. **H**, Same as **G** except for using data from monkey 1 only. **I**, Same as **G** except using data from monkey 2 only.

One of the advantages of neurophysiology is that not only do we get a glimpse of the fine-grain selectivity between specific stimuli, but we also have access to the millisecond-level timecourse of that signal. This timecourse can reveal insights about the dynamic processes that lead to neuronal responses. Of particular interest in this context is the relationship between top-down attention – the dynamic process of selecting a region of space to focus on – and crowding.

When analyzing the latency of target selectivity across different flanker spacings we found a striking relationship between latency and the degree of crowding in the display (Fig. 15G-I). To avoid the confound with signal strength we defined latency as the time the selectivity curve reached half-maximum height. Across the population of neurons recorded from both monkeys, latency significantly increased as spacing decreased (linear regression, $F(71) = 7.98$, $p = 0.005$). Pairwise comparisons revealed that latency was greater for near flankers compared to far flankers (Wilcoxon signed rank test, $z(71) = 2.80$, $p = 0.005$) or singletons ($z(71) = 4.18$, $p = 1.5 \times 10^{-5}$). There was a trend toward significance for mid spacing flankers ($z(71) = 1.34$, $p = 0.09$).

Again, we checked that the effect was present in both monkeys individually. For monkey 1 the target selectivity signal arose significantly later for near flankers versus singletons (Wilcoxon signed rank test, $z(38) = 2.63$, $p = 0.009$) or far flankers ($z(38) = 1.70$, $p = 0.04$). For monkey 2, there was a trend toward significance for near flankers versus singletons (Wilcoxon signed rank test, $z(33) = 1.46$, $p = 0.07$) and mid flankers ($z(33) = 1.65$, $p = 0.05$).

The finding that latency increases as flankers encroach on the target is consistent with an account of crowding that incorporates attention, but it does not rule out other interpretations. For instance, IT neuronal response latency is also increased by occlusion (Kosai et al., 2014) or the addition of noise (Emadi and Esteky, 2013), and latency in early visual cortex is increased by surround suppression (M. A. Smith et al., 2006).

3.3.1.3 Crowding Qualitatively Alters the Neuronal Code

In the previous section, we defined neuronal preferences on the basis of neuronal preferences for targets in isolation. While we found that the singleton code was diminished as flankers grew near, we also don't know at this point whether crowding qualitatively alters the neuronal code. Furthermore, although the previous section seemed to demonstrate crowding-induced quantitative weakening of selectivity, all we really know is that the selectivity for *singletons* was diminished. This does not preclude the situation in which absolute selectivity between arrays remains static, or even increases. Based on the results so far, both the quantitative weakening and the qualitative change hypotheses are still in the running.

To distinguish between these possibilities, we next defined neuronal preferences independently for each flanker spacing (Fig. 16) using a leave-one-out procedure to avoid bias (see Methods). Flanker proximity still had a significant effect on the magnitude of target selectivity in the combined data (linear regression, $F(71) = 16.9$, $p = 0.001$) as well as for both monkeys considered independently (linear regression, $F(38) = 18.82$, $p = 1.1 \times 10^{-4}$ for monkey 1, and $F(33) = 5.70$, $p = 0.02$ for monkey 2). Post hoc pairwise comparison on the full data set revealed that neurons were significantly less selective between targets with nearby flankers compared to flankers at intermediate spacing (Wilcoxon signed rank test, $z(71) = 1.78$, $p = 0.04$), far spacing (Wilcoxon signed rank test, $z(71) = 3.15$, $p = 8.2 \times 10^{-4}$), and singletons (Wilcoxon signed rank test, $z(71) = 3.17$, $p = 1.6 \times 10^{-4}$).

The same trend was present when considering the each monkey's data separately (Wilcoxon signed rank test, $z(38) = 1.70$, $p = 0.04$), far spacing ($z(38) = 2.51$, $p = 0.04$), and

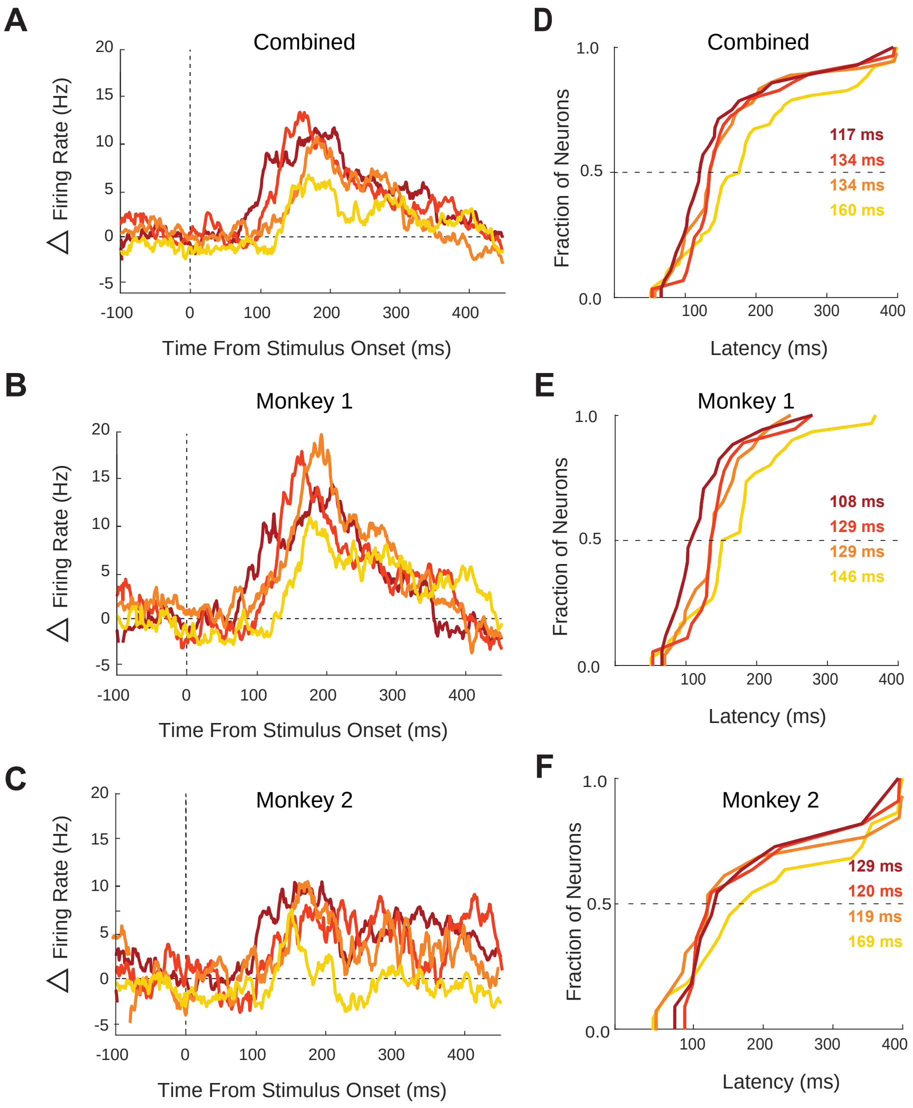

Figure 16. Experiment 1. Neuronal preferences determined separately within each spacing condition. **A,** combined across anomals. **B,** monkey 1. **C**, monkey 2. **D**, Latency distributions for combined data. **E**, monkey 1. **F**, monkey 2. Conventions the same as Fig. 15D-I.

singletons (*z*(38) = 2.80, *p* = 0.003). For monkey 2, post hoc pairwise comparisons revealed that selectivity was still significantly lower for targets surrounded by nearby flankers compared to far flankers (Wilcoxon signed rank test, *z*(33) = 1.72, *p* = 0.04) or singletons (*z*(33) = 2.00, *p* = 0.02). Overall, these findings provide continued support for the hypothesis that crowding quantitatively weakens the neuronal code.

However, quantitative and qualitative changes are not mutually exclusive. Despite the persistent quantitative effect, selectivity for the most crowded targets was strikingly increased when preferences were defined separately within each spacing condition (compare Fig. 15D-F to Fig. 16A-C). For easier direct comparison with the singleton-based analysis, we computed the average firing rate difference between most and least preferred target letters at each flanker spacing for each neuron under both methods (Fig. 17). Selectivity was significantly greater for the within-spacing-defined analysis for both the combined data (Wilcoxon signed rank test, *z*(71) = 3.39, *p* = 3.5 x 10^{-4}), as well as for each monkey considered separately (*z*(38) = 1.85, *p* = 0.03 for monkey 1 and *z*(33) = 2.48, *p* = 0.007 for monkey 2). The finding that using crowded trials to define neuronal preferences significantly improved neuronal selectivity to crowded targets supports the hypothesis that crowding causes a qualitative change in the neuronal code.

To be sure of the robustness of our latency results from the singleton-based analysis, we repeated that procedure on the within-spacing-based selectivity curves. Again, target selection occurred later when flankers were near the target compared to other spacings (Fig. 16D-F). This result was statistically significant for the combined data (Wilcoxon signed rank, *z*(71) = 2.37, *p* = 0.009 for singtons, *z*(71) = 2.06, *p* = 0.02 for far flankers, and *z*(71) = 2.32, *p* = 0.02 for mid flanker spacing).

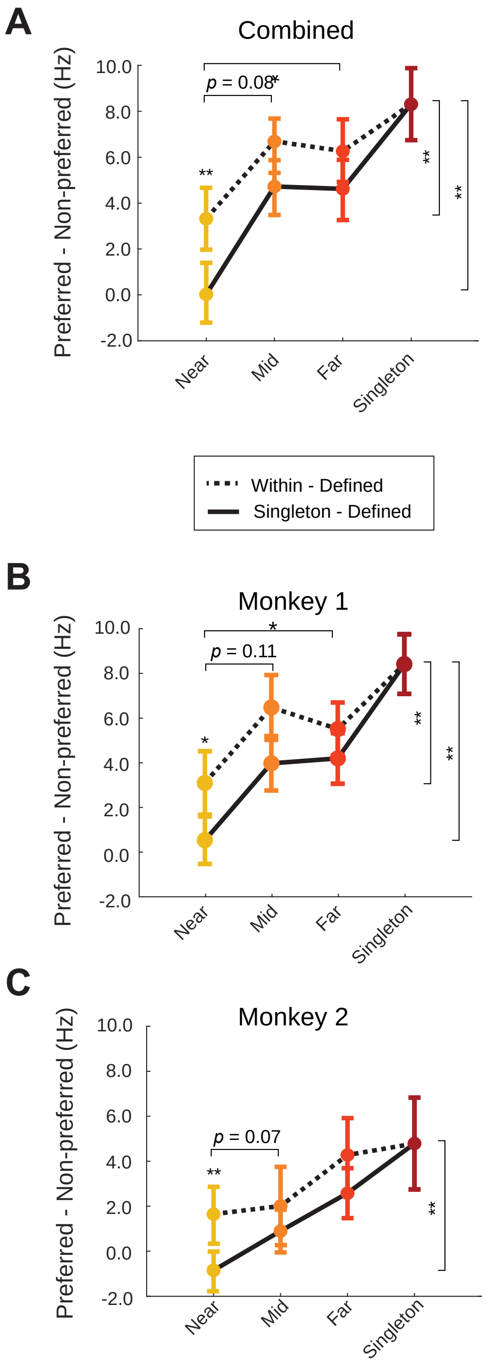

Figure 17. Experiment 1. Selectivity for target letter. **A**, Combined data across both monkeys. The solid line indicates mean firing rate difference between the most and least preferred targets when the preferences of each neuron were determined from singleton trials. Error bars indicate SEM. The dashed line indicates the mean selectivity when preference were defined separately within each spacing condition. Significant differences between the dashed and solid curves at each spacing are denoted by asterisks directly above each point. Significant differences between the singleton condition versus near-flanker selectivity are denoted by asterisks beside the vertical brackets, where the solid bracket corresponds to the singleton-defined preferences and the dashed bracket corresponds to the within-spacing-defined preferences. To examine the divergence between the two curves, we computed the difference between them and examined how it varied between spacings. In all cases, significant differences are denoted by * (Wilcoxon signed rank test, p < 0.05). In all cases, the statistical test was Wilcoxon signed rank with * denoting p < 0.05 and ** denoting p < 0.01. **B**, Monkey 1 data only. **C**, Monkey 2 data only.

For individual monkeys, the results were similar, but less statistically robust. For monkey 1, latency was significantly greater for near flanker trials versus singletons (Wilcoxon signed rank, $z(38) = 2.10$, $p = 0.02$) and trending toward significance for mid spacing ($z(38) = 1.54$, $p = 0.06$) and far spacing ($z(38) = 1.56$, $p = 0.06$) flankers. For monkey 2, latency was significantly greater when flankers were near compared to all other conditions (Wilcoxon signed rank, $z(33) = 2.11$, $p = 0.02$ for singletons, $z(33) = 1.78$, $p = 0.04$ for far spacing, and $z(33) = 1.92$, $p = 0.03$ for mid spacing).

Previously, all of our analyses have focused on the magnitude of neuronal selectivity, as measured by the difference between firing rates evoked by the most and least preferred stimuli, out of the possible four presented during each session. To better characterize the degree to which neurons were changing their target preferences as a function of crowding, we next considered how the responses to all four targets correlated across different flanker spacings. A relatively strong correlation between near-flanker (crowded) conditions and the other conditions would indicate conservation of the code and argue against a qualitative impact of crowding (Fig. 18). Importantly, we wanted to establish a ceiling against which to compare the correlations of interest, because what appears to be a "low" correlation across two spacings might actually be the largest correlation possible. To that end, we subdivided firing rates into even and odd trials so that we could calculate correlations within the same spacing (Fig. 18A-C, colored circles on main diagonal), which indicate the maximum possible correlation that any of the off-diagonal correlations could achieve. Since we were correlating across the population of neurons with naturally varying mean firing rates, we first z-scored the data so that spurious correlations wouldn't arise simply from some neurons firing more than others in general.

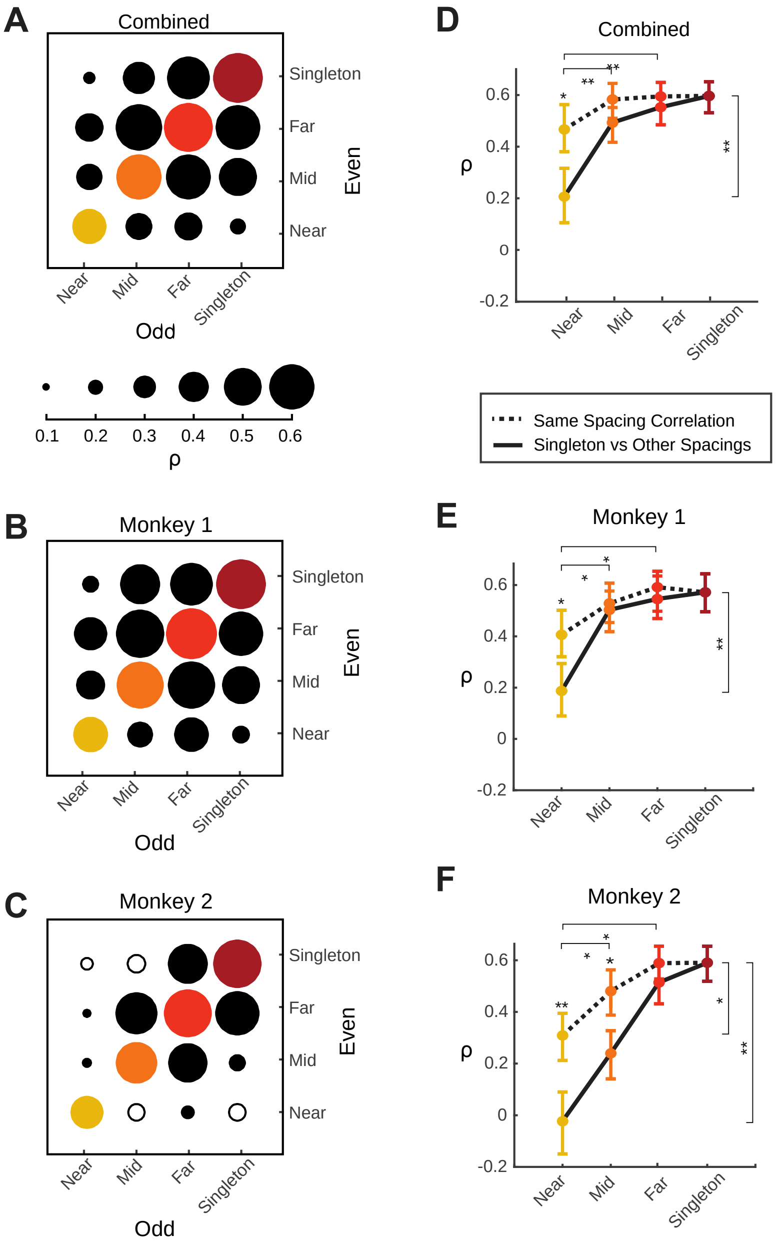

Figure 18. Experiment 1. Target preference correlations across spacing. **A**, Combined data across monkeys. The color code is retained from previous figures. Correlations are computed by comparing even and odd trials and axes are labeled accordingly. Correlation strength (Spearman-Brown corrected) is denoted by circle diameter. Negative correlations are denoted by open circles. **B**, Monkey 1. Conventions as in **A**. **C**, Monkey 2. **D**, Correlation strength as a function of spacing. Error bars indicate 95% confidence intervals. The dashed line indicates correlations between odd and even trials at the same spacing (i.e., the main diagonal in **A**). The solid line indicates the correlation between the singletons and the other spacings, averaged over the upper and lower triangles of the matrix. Significant differences between the correlation strength at a given spacing is denoted by * (Pearson correlation confidence intervals, $p < 0.05$) or ** ($p < 0.01$) directly above error bars. Significant differences between the singleton-singleton correlations and the near-singleton correlations are denoted by ** (Pearson correlation confidence intervals, $p < 0.01$) beside the vertical solid bracket. Significant differences between the singleton-singleton and near-near correlations are denoted by * ($p < 0.05$) beside the vertical dashed bracket. The difference between the curves is compared across spacings using a bootstrap analysis with 1,000 iterations. **E**, Monkey 1 data only. Same conventions as in **D**. **F**, Monkey 2 data only.

For both monkeys as well as in the combined data, the near-near correlation was significantly greater than the near-singleton correlation (95% confidence intervals; Fig. 18D-F, yellow points on dashed versus solid lines). This finding indicates that the neurons consistently signaled the identity of crowded targets, yet they used a different pattern of responses than when those same targets were presented in isolation.

To get a sense for whether the crowded code was particularly deviated from the singleton code, we compared the spread between the two curves (dashed versus solid lines) as a function of flanker spacing. What we found was that for the combined data the spread was greater for near compared to either mid or far flankers (1,000 iteration bootstrap, $p < 0.01$; Fig. 18C). Both monkeys showed the same trend, albeit weaker, in their individual data (1,000 iteration bootstrap, $p < 0.05$; Fig. 18D,E).

Overall, these results serve to bolster the notion that crowding qualitatively alters the pattern of neuronal responses across different targets. Together, these findings suggest that the detrimental effects of crowding on behavioral performance (Fig. 15A) were at least in part due to a confusion of the neuronal code and not solely due to divisive normalization. A firing rate pattern signifying one target in isolation could signify another as flankers crowd around. This is not only a novel task, but also a novel neuronal behavior.

3.3.2 Experiment 2

While the first experiment provided evidence that crowding qualitatively alters neuronal selectivity in IT, our experimental design was not equipped to determine exactly what that qualitative changes entailed. We designed experiment 2 explicitly for the purpose of decomposing selectivity into main effects from interaction effects. If the qualitative change was to manifest as a main effect, that would mean that the mere presence of another stimulus disrupts the neuronal code for the target. If, on the other hand, the qualitative change was to manifest as an interaction effect, that would mean that the identity of the flanker influences the nature of the code change. We also included an additional control in the form of stimulus size. Since crowding is driven by the absolute spacing between peripheral stimuli, independent of their size, we wanted to be sure that any neuronal mechanism we may ascribe to crowding possesses this same property.

3.3.2.1 Task and Stimuli Design

To better understand the mechanism by which crowding disrupts neuronal preferences, we orthogonalized three variables: stimulus spacing, stimulus size, and stimulus pairing. This design allowed us to calculate the main and interaction effects between the stimuli and observe how they changed with spacing, size, and the method of main and interaction effect determination.

Both animals were passively fixating while letter-like stimuli appeared in the contralateral visual hemifield (Fig. 19A). Each trial consisted of two alternating periods of fixation and passive viewing. The stimuli on the screen during the second passive viewing session were never the same as the first. The full set of all possible combinations of first and second stimuli were presented in random order throughout the session. These constraints aimed to minimize the effects of repetition suppression (McMahon and Olson, 2007) and perceptual learning (Miyashita, 1988) that could influence firing rates and introduce confounds.

Stimuli were a combination of Sloan letters and Sloan-like letters designed in the lab, which we refer to collectively as “elements.” Each session involved a set of four elements, chosen from one of two separate sets (Fig. 19A,B). At most two of these elements were on the screen at any given time, with one located vertically above the other. The vertical arrangement was chosen because IT receptive fields are usually shaped like a 2D Gaussian distribution, overlapping the fovea, and extending into the contralateral hemifield (Op De Beeck and Vogels, 2000). Vertically-arranged elements are less likely to exit the receptive field as spacing increases, unlike horizontally-arranged elements. The element pair was always presented at 6° eccentricity and vertically centered about the horizontal meridian. Two of the elements could appear on top and two could appear on bottom for a total of four possible pairings during a given session (Fig. 19A,B). Each stimulus was also presented by itself at 6° eccentricity, situated on the horizontal meridian. We had two such sets of letter-like stimuli at our disposal, and the choice between them was dictated by the preferences of the neurons we happened to be recording from.

The variables of interest were stimulus size, absolute center-to-center spacing between stimuli, and center-to-center spacing relative to size. We chose two letter sizes, 1° and 2°, as a compromise between the small letter sizes used in the crowding literature (as well as in our experiment 1) and the fact that the optimal stimulus size for IT neurons is 3.7° on average (Ito et al., 1995). The large element displays were identical to the small element displays except scaled up by a factor of 2 (Fig 19C). Inter-element spacings were chosen such that the smallest spacing for large elements was the same absolute center-to-center spacing as the largest spacing between small elements (2.2°).

Altogether there were 40 unique conditions in a session. These conditions included all possible combinations of the four stimulus pairs, two stimulus sizes, and four spacings relative to size, as well as the four large and four small stimuli shown by themselves. Each condition was presented eight times to get a good estimate of mean firing rate for each neuron. Since two conditions were presented in a given trial, the animals only had to complete 160 trials total during a recording session. Successful trials were those in which the animal maintained fixation on the fixation point at the center of the screen throughout the entire trial duration (Fig. 19D).

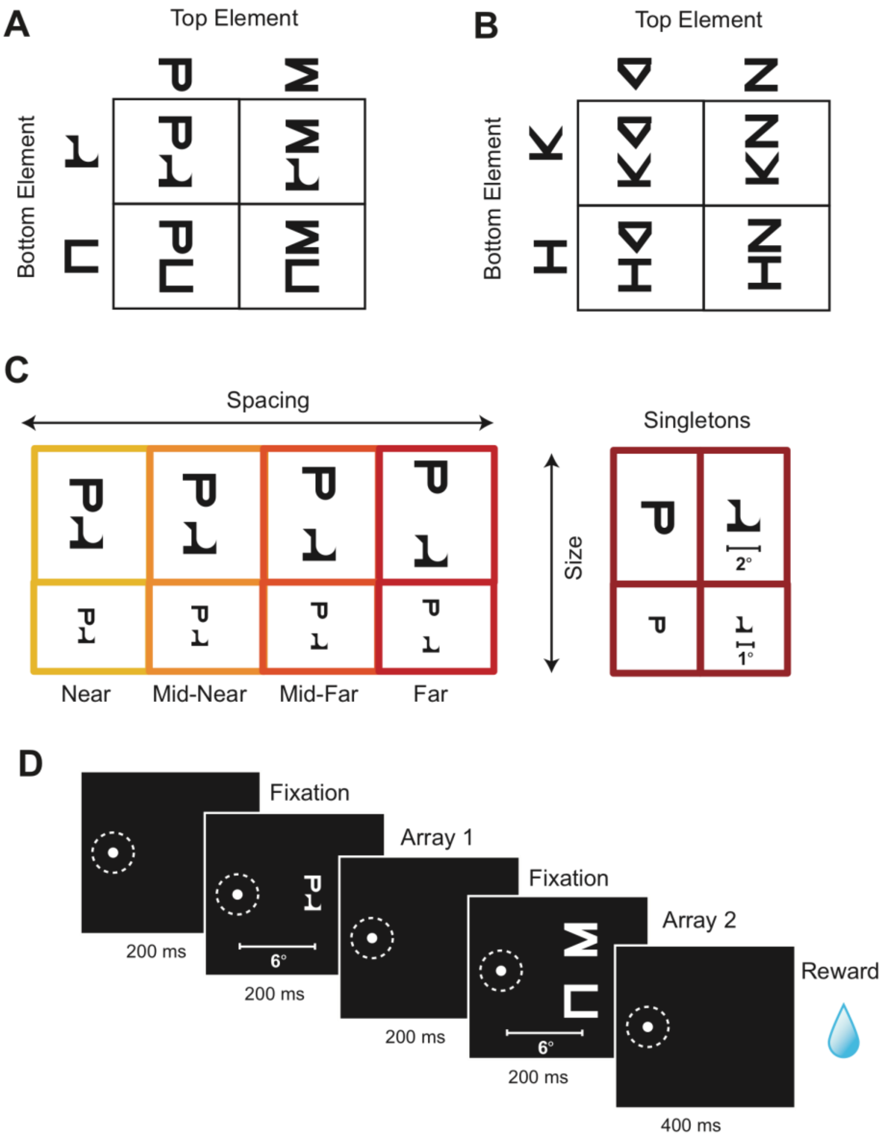

Figure 19. Experiment 2 task and stimulus design. **A**, The top and bottom elements and all possible combinations for stimulus set 1. **B**, The top and bottom elements as well as all combinations for stimulus set 2. **C**, There were four possible spacings, two sizes, as well as singletons. The large letters were twice the size of the small letters. Spacings were chosen such that the nearest relative spacing for the large letters was the same absolute spacing as the farthest possible spacing for the small letters. **D**, The sequence of events during each trial.

3.3.2.2 Main Effects Change with Size-Relative Spacing

We recorded 34 visually-responsive neurons from monkey 1 and 36 from monkey 2 while they performed the experiment 2 passive fixation task (Fig. 19). Because both monkeys performed the same task their data was combined in the figures, but we also repeated all tests separately for each monkey individually. The same trends generally persisted, with divergences noted where they appeared.

First we computed the main effect of target identity (see Methods) for each spacing and element size (Fig. 20A,B). Similar to experiment 1, preferences used to align main effects across neurons were defined on the basis of singleton conditions (solid line) and then again within each spacing independently (dashed line). Because the design of this experiment produced a main effect of both top and bottom elements we averaged those together to arrive at a single main effect measurement at each spacing and size.

We were interested in whether main effects underwent a qualitative change as spacing between elements decreased, and if so, whether this change was dependent upon relative or absolute spacing. So, for each spacing, size, and neuron we looked for differences in main effect strength across the two methods of preference designation (Fig. 20A,B, dashed versus solid line). For small elements, there was a significant difference between singleton-defined versus within-spacing-defined main effects for near-spacing (Wilcoxon signed rank test, $z(70) = 4.66$, $p = 3.2 \times 10^{-6}$), mid-near spacing ($z(70) = 3.07$, $p = 0.002$), and mid-far spacings ($z(70) = 2.24$, $p = 0.02$). For large elements, there was a significant difference for near-spacing ($z(70) = 2.82$, p = 0.005), mid-near-spacing ($z(70) = 7.10$, $p = 1.2 \times 10^{-12}$), and mid-far-spacings ($z(70) = 3.99$, $p = 6.7 \times 10^{-5}$). Main effects appear to change qualitatively with the relative spacing between elements.

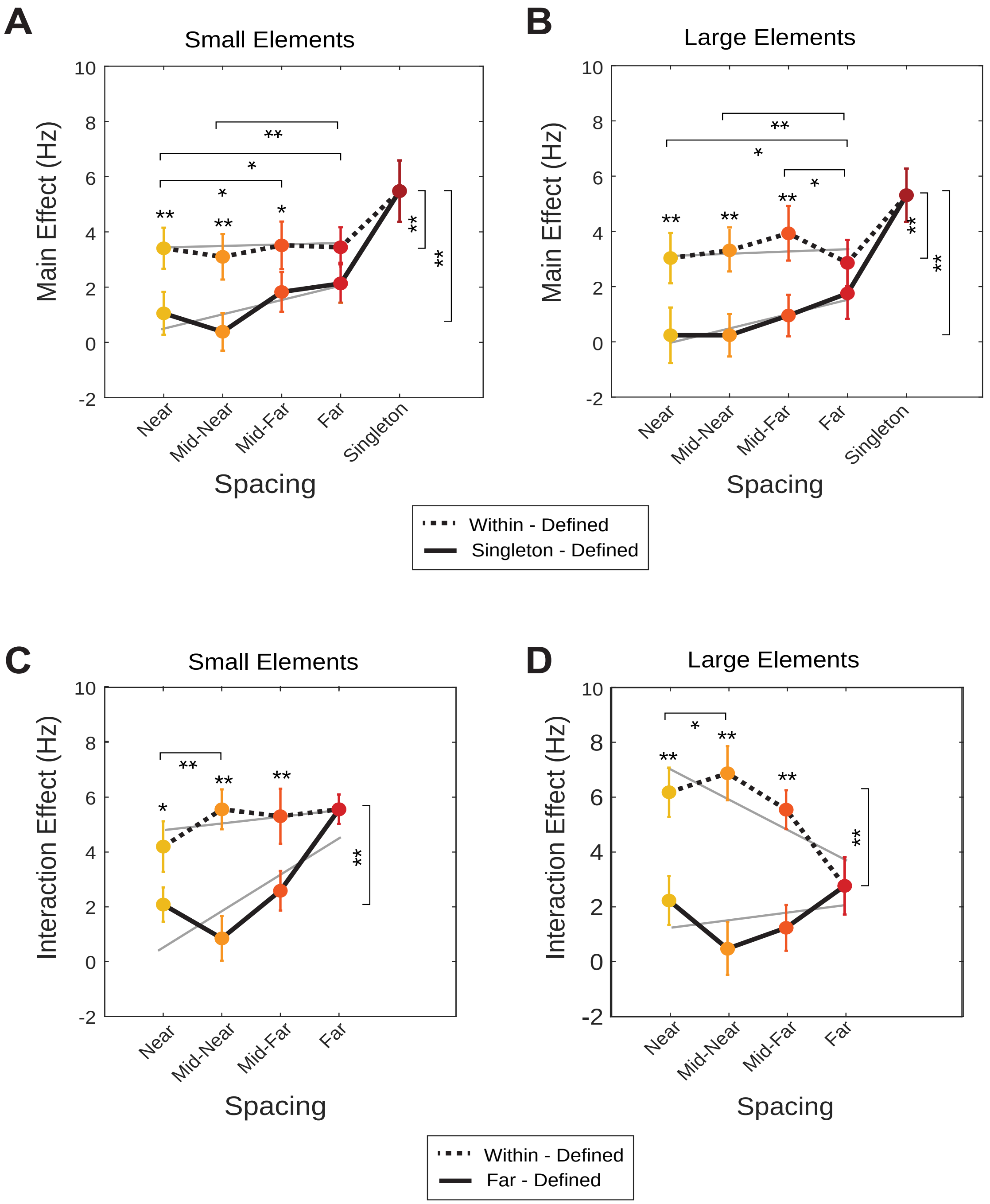

Figure 20. Experiment 2. Main and interaction effects change with spacing. **A,** Mean main effect averaged over the top and bottom elements for small elements as a function of spacing. Error bars indicate standard error of the mean. Solid line connects cases where singletons were used to define neuronal preferences. Dashed line connects cases where preferences were defined within each spacing separately. Gray lines indicate linear regression fit. **B,** Mean main effects for large elements. Conventions as in **A. C,** Interaction effect of small elements as a function of spacing. Solid line connects cases where preferences were defined from the far spacing trials. Dashed line connects the cases where preferences were defined separately within each spacing. **D,** Interaction effects for large elements. Conventions as in **C**. In all panels, significance is denoted with ** (*p* < 0.01) or * (*p* < 0.05). Significant difference between the dashed and solid curves within a given spacing are denoted by asterisks directly above the error bars. Significant differences between dashed and solid lines across spacings are denoted by horizontal brackets. Significant differences between far- and near-element spacings for the wintin-defined conditions are denoted by vertical brackets.

When the data from each monkey was considered separately, the overall trend remained the same but many comparisons between the singleton-defined and within-spacing-defined main effect strength failed to reach statistical significance. For monkey 1 and small elements, mid-far-spacing remained significant (Wilcoxon signed rank test, $z(34) = 2.25$, $p = 0.02$) and far spacing was trending toward significance ($z(34) = 1.63$, $p = 0.10$). For large elements the difference between the curves remained significant for near spacing ($z(34) = 2.48$, $p = 0.01$) and was trending toward significance for mid-near ($z(34) = 1.51$, $p = 0.13$) and mid-far spacings ($z(34) = 1.44$, $p = 0.15$). For monkey 2 and small elements, the singleton-defined and within-spacing defined main effect strength was still highly significantly different for near ($z(36) = 5.63$, p = 1.8×10^{-8}), mid-near ($z(36) = 5.90$, p = 3.5×10^{-9}), and mid-far ($z(36) = 4.14$, p = 3.5×10^{-5}) spacings. For large elements, the singleton-defined and within-spacing-defined main effects remained significantly different for mid-near ($z(36) = 6.60$, p = 4.0×10^{-11}) spacing, and was trending toward significance for mid-far spacing (Wilcoxon signed rank test, $z(36) = 1.74$, p = 0.08).

Having shown that main effects undergo a qualitative change as relative inter-element spacing decreases, we wanted to see whether the magnitude of this change intensified as the elements drew nearer, as we observed with selectivity in experiment 1. The purpose here is to determine whether the widening gap for selectivity – which we attributed to crowding – is still present when we've pulled out only main effects and controlled for relative versus absolute spacing. For small elements, the qualitative change in main effect was significantly larger for near compared to either mid-far (Wilcoxon signed rank test, $z(70) = 1.94$, $p = 0.03$) or far spacing ($z(70) = 1.72$, $p = 0.04$) conditions. Mid-near also underwent a significantly larger qualitative change compared to far spacing conditions ($z(70) = 2.43$, $p = 0.008$; Fig. 20A). Large elements exhibited a similar pattern. The separation between the two curves was significantly larger for mid-near (*z*(70) = 2.76, *p* = 0.003), mid-far (*z*(70) = 1.97, *p* = 0.02), and near-spacing (*z*(70) = 1.74, *p* = 0.04) conditions compared to far spacing conditions (Fig. 20B). The qualitative change in main effects continues to mirror what we saw in experiment 1 for selectivity. Furthermore, it is relative rather than absolute spacing that drives the qualitative change in main effects.

The same trend was present in both monkeys' individual data, albeit not as statistically robust as in the combined data. For monkey 1, there was a significant width difference when comparing large elements at mid-near versus mid-far spacings (Wilcoxon signed rank test, *z*(34) = 1.78, *p* = 0.04) or when comparing small elements at near spacing versus mid-far (*z*(34) = 1.61, *p* = 0.05) and far (*z*(34) = 2.05, *p* = 0.02) spacings. For monkey 2, the curves were significantly more separated for small elements at near spacing compared to mid-near (Wilcoxon signed rank test, *z*(36) = 1.75, *p* = 0.04), mid-far (*z*(36) = 2.14, *p* = 0.02), and far (*z*(36) = 2.46, *p* = 0.007) spacings. For large elements, the curves became significantly wider between mid-near (Wilcoxon signed rank test, *z*(36) = 5.05, $p = 4.5 \times 10^{-7}$) and far (*z*(36) = 2.80, *p* = 0.005) spacings compared with far spacing.

Because we saw such a pronounced impact of spacing on the latency of target selection during experiment 1, we were curious whether this effect persisted for the present experiment. If so, that would be an indication that the latency effect was driven by a bottom-up mechanism – because neither element in the experiment 2 pairings was more likely than the other to be a target of attention – and was not dependent on the number of items in the array. On the contrary, we did not find evidence for a connection between inter-element-spacing and main effect latency for experiment 2. When selectivity was determined separately for each spacing, the time to main effect half height across the population of neurons was not significantly affected by inter-element spacing for either small (linear regression, *F*(70) = 1.82, *p* = 0.18) or large (*F*(70) = 0.84, *p* = 0.36) elements. When main effect preferences were determined from the singleton conditions, there was a slight trend toward longer latency with narrower spacing for small elements ($F(70) = 3.25$, $p = 0.07$), but not for large elements ($F(70) = 1.37$, $p = 0.24$). There was also no significant latency effect for monkey 1 (singleton-defined: $F(34) = 0.96$, $p = 0.33$ for small elements and $F(34) = 1.78$, $p = 0.19$ for large elements; within-spacing-defined: $F(34) = 1.26$, $p = 0.27$ for small elements, $F(34) = 1.09$, $p = 0.30$ for large elements) or monkey 2 (singleton-defined: $F(36) = 0.47$, $p = 0.50$ for small elements and $F(36) = 0.09$, $p = 0.76$ for large elements; within-spacing-defined: $F(36) = 0.11$, $p = 0.74$ for small elements, $F(36) = 0.06$, $p = 0.80$ for large elements).

So far, the analyses we have presented for experiment 2 have considered large and small element conditions separately. Similar to experiment 1, we observed a qualitative change in main effects as inter-element spacing decreased in both cases. However, since the central question here is whether main effects constitute a neuronal correlate of the crowding phenomenon we must directly compare main effects at matched absolute and relative spacings. Since the degree of crowding is determined by eccentricity (which we kept constant here) and absolute center-to-center spacing between stimuli regardless of stimulus size (Levi, 2008; Pelli et al., 2004), we expected that if crowding were manifest in main effects then the pattern of main effects would vary systematically as a function of absolute, not relative, spacing.

First, we tested the effect of absolute spacing. To do this we called upon the pair of conditions in which absolute spacing was held constant while element size varied. When large elements were situated at the near spacing they had the same absolute spacing (2.2°) as when small elements were positioned at the far spacing (Fig. 19C). Comparing the within-spacing-defined main effects for these two conditions revealed no significant difference for either the combined data (Wilcoxon signed rank test, $z(70) = 0.10$, $p = 0.46$) or for either monkey individually ($z(34) =$

0.42, *p* = 0.74 for monkey 1 and *z*(36) = 1.22, *p* = 0.22 for monkey 2). When considering singleton-defined main effects, a borderline significant difference emerged for the combined data (*z*(70) = 1.47, *p* = 0.07) and monkey 2 (*z*(36) = 1.63, *p* = 0.10), but not for monkey 1 (*z*(34) = 0.56, *p* = 0.58). The most striking evidence against absolute spacing, however, comes from the difference between the curves (Fig. 20A,B, compare the spread for far spacing with small elements to near spacing with large elements). There was a significantly greater difference for large elements spaced near one another compared to small elements spaced far apart (Wilcoxon signed rank test, *z*(70) = 2.60, *p* = 0.005). This was also true of the individual monkey data (*z*(34) = 1.97, *p* = 0.05 for monkey 1, and *z*(36) = 2.45, *p* = 0.01 for monkey 2).

Since absolute spacing did not appear to be the primary driver of the qualitative main effect change, we next turned to size-relative spacing. For this step, we were comparing more than just a single pair of conditions, so we performed an ANCOVA analysis to determine how the pattern of main effects changed as a function of scale. Essentially, we were asking whether the slope of the fit line (gray line in Fig. 20A) for small elements was different from the slope of the fit line for the large elements (gray line in Fig. 20B). We did this separately for both singleton-defined and within-spacing-defined main effects. In all cases, size had no significant impact on the pattern of results (*F*(70) = 0.28, *p* = 0.60 for singleton-defined main effects, and *F*(70) = 0.08, *p* = 0.78 for within-spacing-defined main effects). This held true for both monkey 1 (*F*(34) = 1.43, *p* = 0.24 for singleton-defined, and *F*(34) = 0.73, *p* = 0.40 for within-spacing-defined main effects) and for monkey 2 (*F*(36) = 0.35, *p* = 0.55 for singleton-defined, and *F*(36) = 0.02, *p* = 0.89 for within-spacing-defined main effects). We take this as evidence that the qualitative change in main effects is a function of the relative spacing between elements, not absolute spacing. Therefore, we are hard-pressed to claim that main effects are at the heart of the crowding phenomenon. Next, we turn to interaction effects as a potential candidate for the neuronal correlate of crowding.

3.3.2.3 Interaction Effects Are Sensitive to Element Size

As with the main effects, we calculated interaction effects for each spacing and for each element size. Rather than defining preferences on the basis of singletons, because that makes no sense for interactions, we used the size-relative far spacing trials as a point of comparison for the within-spacing analysis. The calculation of interaction effect strength is detailed in the Methods.

As with main effects, there were no strong latency trends for the onset of interactions across the IT neuronal population. When selectivity was determined by singletons, the time to interaction effect half height across the population of neurons was not significantly affected by inter-element spacing for either large or small elements (linear regression, $F(70) = 2.79$, $p = 0.10$ for small elements, and $F(70) = 0.43$, $p = 0.63$ for large). Likewise, when selectivity was determined separately for each spacing neither small nor large elements exhibited a systematic change in interaction latency as a function of spacing (linear regression, $F(70) = 0.17$, $p = 0.68$ for small elements, and $F(70) = 0.84$, $p = 0.36$ for large). There was also no significant latency effect for monkey 1 (singleton-defined: $F(34) = 0.48$, $p = 0.49$ for small elements and $F(34) = 0.03$, $p = 0.86$ for large elements; within-spacing-defined: $F(34) = 2.10$, $p = 0.15$ for small elements, $F(34) = 0.41$, $p = 0.52$ for large elements) or monkey 2 (singleton-defined: $F(36) = 0.08$, $p = 0.78$ for small elements and $F(36) = 1.65$, $p = 0.21$ for large elements; within-spacing-defined: $F(36) = 0.43$, $p = 0.52$ for small elements, $F(36) = 2.35$, $p = 0.13$ for large elements).

The first order of business was to determine whether qualitative changes in neuronal preferences were manifest in interaction effects. Indeed, this was what we observed for both small (Fig. 20C) and large (Fig. 20D) elements. Similar to main effects, interactions were larger when defined on the basis of far-spacing conditions (compare dashed and solid lines). This relationship was significant for small elements at near spacing (Wilcoxon signed rank test, *z*(70) = 1.94, p = 0.04), mid-near spacing (*z*(70) = 6.39, *p* = 1.7 x 10^{-10}), and mid-far spacing (*z*(70) = 4.90, *p* = 9.6 x 10^{-7}). The same trend was present for large elements (Wilcoxon signed rank test, *z*(70) = 6.91, p = 4.8 x 10^{-12} for near, *z*(70) = 8.78, p = 1.7 x 10^{-18} for mid-near, and *z*(70) = 7.54, p = 4.5 x 10^{-14} for mid-far spacing). Comparing the difference between the two methods across spacing, we saw that for both small and large elements the largest effect was evident for the mid-near element spacing. Compared to near-spacing conditions, mid-near-spacing interactions were significantly larger for both small (Wilcoxon signed rank test, *z*(70) = 6.37, p = 2.0 x 10^{-10}) and large (*z*(70) = 2.20, p = 0.03) elements. The non-monotonic relationship between the qualitative neuronal code change suggests that size-relative mid-range distances between elements are optimal for interactions, whereas the closest spacings may give way to interference.

In contrast with main effects, however, interaction strength actually increased as spacing between elements decreased, but only for large elements (linear regression; *F*(70) = 7.59, *p* = 0.008; Fig. 20D). That the pattern of results was different across sizes for interactions was intriguing because it suggested that perhaps interaction effects underlie crowding. To directly test how spacing impacted interaction strength across the two stimulus sizes, we compared the slopes of the linear regression fits (Fig. 20C,D, gray lines). What we found was that the slope of the fit lines was significantly different across stimulus sizes regardless of how interactions were defined (ANCOVA, *F*(70) = 26.17, *p* = 0.0003 for within-spacing-defined interactions, and *F*(70) = 15.6, *p* = 0.0005 for far-defined).

This pattern was slightly different across monkeys. For monkey 1, the within-spacing-defined interaction effects significantly decreased with spacing for small elements (linear regression, *F*(34) = 4.44, *p* = 0.04), but did not systematically vary for large elements (linear regression, *F*(34) = 1.15, *p* = 0.29). The difference in slope across size was still significant (ANCOVA, *F*(34) = 5.71, *p* = 0.02). For monkey 2, within-spacing-defined interactions did not vary across spacing for small elements (linear regression, *F*(36) = 0.009, *p* = 0.92), but for large elements there was a significant increase in interaction effect magnitude as spacing decreased (linear regression, *F*(36) = 4.39, *p* = 0.04). Again, there was a significant difference in slope across size (ANCOVA, *F*(36) = 6.21, *p* = 0.02). Therefore, in all cases interaction effects were not scale invariant. But to claim that crowding is manifest in interaction effects requires more than simply showing that it is not driven by relative spacing. We must show that absolute spacing is a better predictor.

This is precisely what we found. Absolute interaction strength did not vary significantly when elements were at the same spacing, 2.2° apart (Wilcoxon signed rank test, *z*(70) = 0.98, *p* = 0.33). This was true for both monkey 1 (*z*(34) = 0.56, *p* = 0.58) and monkey 2 (*z*(36) = 1.02, *p* = 0.30). This finding supports the idea that interactions underlie the crowding phenomenon.

3.3.2.4 Swapping Preferences

Main and interaction effects are informative, but also quite removed from the raw data. To gain intuition about what these effects mean for the neuronal code it’s often helpful to see what individual neurons are doing. For the population-level main and interaction effects to change with spacing, individual neurons should be flipping their preferences across spacings. That’s exactly what we saw (Fig. 21).

When looking at the exact same element pair at both the near and far spacing it’s clear that the neuronal code is not static (Fig. 21). Whereas one element pair is preferred at one spacing (Fig. 21A, compared dashed and solid lines), the other pair is preferred at a different spacing (Fig. 21B). In this example, the top element in the pair was held constant to demonstrate that the preference for the two bottom elements truly flipped, simply as the result of the proximity of the other element. For this neuron, the other top element did not generate such a preference flip (Fig. 21C,D). Therefore, this illustrates an example of a single neuron that underwent a qualitative interaction effect change (as seen by the specificity of the preference flip to a particular top element; compare Fig. 21A,B to Fig. 21C,D).

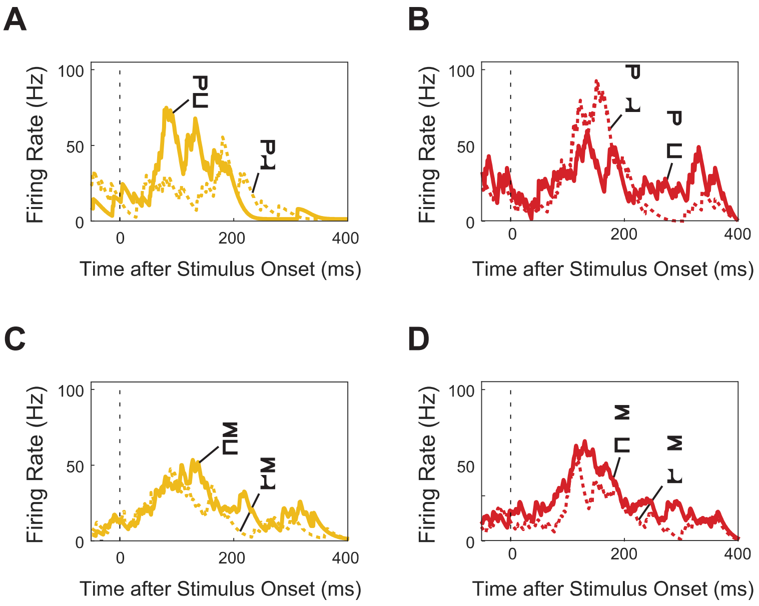

Figure 21. Experiment 2. Example neuron showing flipping preferences as a function of spacing between small elements. **A**, Yellow indicates near spacing. Solid line denotes one element pair while the dashed line denotes another. **B**, Red indicates far spacing. The same line style conventions as in **A**. Notice that the solid line is above the dashed line in **A** whereas in **B** the dashed line is on top. **C**, The same neuron responding to the other two pairs of elements used in this experiment. Near spacing again denoted by yellow. **D**, Far spacing with the same element pairs as in panel **C**. Notice that the order of the dashed and solid lines (indicating specific element pair) is unchanged across panels **C**, and **D**.

3.3.2.5 Main Effects are Correlated across Size, Interactions are Not

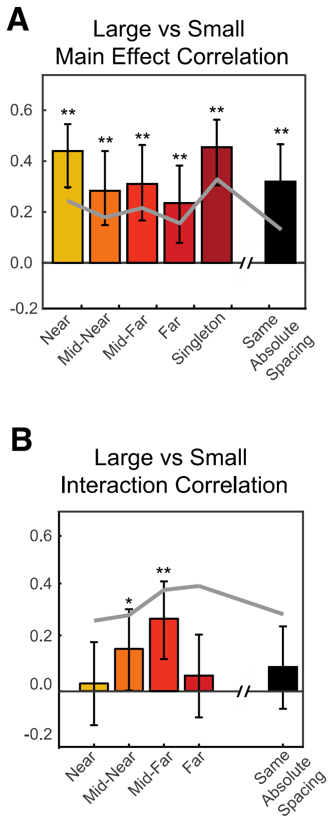

Figure 22. Experiment 2. Correlation across size. **A**, Main effect correlation between element sizes across the population of neurons. Error bars indicate 95% confidence intervals and ** denotes significance ($p < 0.01$). Gray line indicates the across-size minimum lower 95% confidence interval of the Spearman-Brown-corrected correlation between odd and even trials at each spacing. **B**, Interaction effect correlations across size for the population of neurons.

IT neurons are known to maintain rank order across isolated stimuli regardless of size (Ito et al., 1995). Even though population main effect strength was scale-invariant (Fig. 20A,B), without a directly comparing stimulus rank order across sizes we cannot say for certain that the precise neuronal code was the same for large and small elements. To get at this question, we performed a correlation analysis across size for each spacing condition (Fig. 22). Large and small main effects were significantly correlated at all size-relative spacings (Fig. 22A; Pearson correlation, 99% confidence intervals). The magnitude of these correlations was within or above the minimum 95% confidence interval for the correlation between odd and even trials at the same size and spacing (Fig. 22A, gray line), indicating that neuronal preference were fully conserved across size. When absolute spacing was kept constant but scale and relative spacing were different, the correlation was still significantly different from zero (Fig. 22A black bar; Pearson correlation, 95% confidence interval). When the data from each animal was analyzed separately the correlations remained significantly different from zero in monkey 1 (all spacings except near and mid, 95% confidence intervals) as well as monkey 2 (all spacings, 95% confidence intervals). These results reinforce the claim that main effects are scale-invariant.

Interaction effects exhibited a different pattern (Fig. 22B). The correlation between interaction effects across large and small elements at the same relative spacing were significantly different from zero only for mid-near (Pearson correlation, 95% confidence interval), and mid-far (99% confidence interval) spacings, and failed to reach significance for near or far spacings. When the data from the two monkeys were considered separately the same pattern was present, but weaker. For monkey 1, the mid-near and mid-far cross-size correlations were significantly greater than zero. For monkey 2 the mid-far cross-size correlation was the only one to reach statistical significance. Unlike main effects, scale did seem to matter for interactions. However, even though absolute spacing may be a better predictor of interaction effect strength (Fig. 20C,D), there was no significant correlation across size for conditions in which absolute spacing was held constant (Pearson correlation, 95% confidence interval; Fig. 22B, black bar). This contradiction highlights the importance of examining how stimulus rank order changes with experimental variables, rather than merely measuring overall effect strength and raises the question of whether crowding is expressed as interactions among elements in IT neuronal representations.

3.3.3 Experiment 3

Whereas experiment 2 was designed to pose the question of whether the qualitative change in the neuronal code observed under crowding is driven by main or interaction effects, experiment 3 aims to explore how the different parts of crowded objects contribute to crowding. It could be that only the adjacent edges interact with one another or main effects of certain parts are suppressed. Alternatively, all the elements of crowded objects may get tossed together into a mixed-up jumble of alphabet soup. A secondary question pertains to whether interactions between parts depend on the nature of the parts themselves or the location of those parts in the visual field.

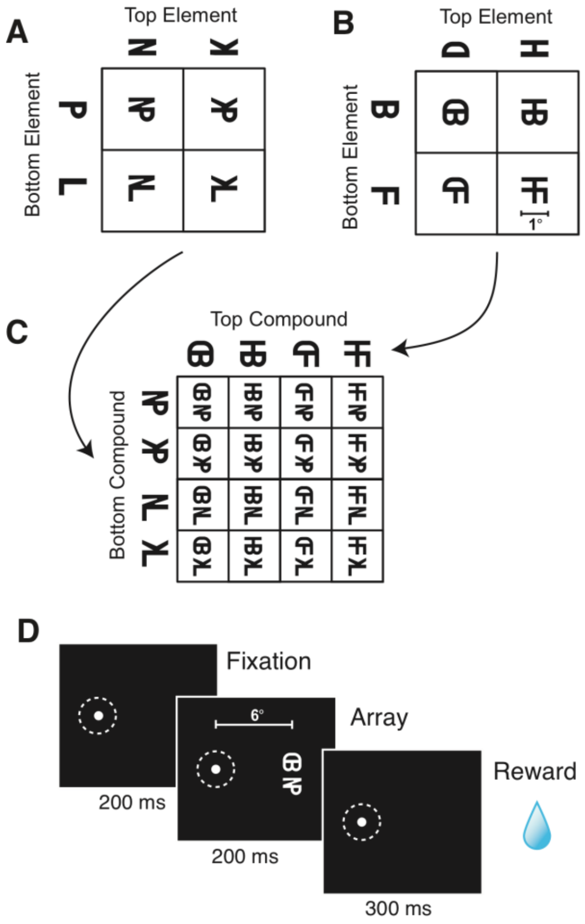

Figure 23. Experiment 3 task and stimulus design. **A**, Top and bottom elements combine in all possible ways to create the top compound. **B**, Top and bottom elements combine in all possible ways to create the bottom compound. **C**, Top and bottom compounds combine in all possible ways to make the full stimulus set. **D**, Series of events during each trial.

3.3.3.1 Task and Stimuli Design

Since we are only interested in the interactions between parts of crowded stimuli, we chose a single center-to-center spacing (1.1°) which matches the most crowded conditions in experiments 1 and 2. As in experiment 2, pairs of stimuli were always vertically-positioned. Each compound stimulus within the pair was made up of two Sloan letters fused together along the line they mutually share (Fig. 23A,B). The aspect ratio of the Sloan letters was halved so that when the two letters came together to form a compound it could have an aspect ratio of one. Line thickness was kept constant across the letters and their shared boundary. Four Sloan letters were combined together in all possible permutations to make the set of four top compounds (Fig. 23A). Another four Sloan letters came together in all possible combinations to make the set of four bottom compounds (23B). Then the four top compounds and the four bottom compounds were combined in all possible ways to form the complete set of 16 conditions (Fig. 23C). Each condition was repeated eight times for a total of 128 trials per session. During each trial the animals fixated a central spot, a pair of compounds was flashed in the periphery, and a fluid reward was given if fixation was held throughout (Fig. 23D).

To decouple stimulus identity from location on the screen we also showed the same set of compound stimuli flipped about the horizontal axis while recording from a subset of neurons. This is an important control because the pattern of main effects and interactions across the elements of the compounds could arise from the position of those particular elements in the display. By inverting the displays we should also invert any patterns that are due solely to stimulus identity, while keeping patterns attributable to retinotopic location the same.

3.3.3.2 Non-Adjacent Elements Interact across Crowded Displays

We recorded from 39 visually-responsive neurons from monkey 1 and 21 from monkey 2. Of those, 13 visually-responsive neurons from monkey 1 and 17 from monkey 2 were tested with both the original stimuli as well as their vertical mirror images.

The goal of this experiment was to determine how main and interaction effects vary across different parts of crowed displays and whether any pattern that might emerge was a result of position in the array versus the parts themselves. We could compute a main effect for each of the four positions by computing the difference between the preferred and non-preferred letter occupying that part of the stimulus. This is the same procedure that we adopted previously (see Methods). Likewise, the six possible pairwise interactions between each of the four parts of the arrays could be computed using the same recipe as before (see Methods). To determine whether the measured strengths were greater than expected by chance we performed a bootstrap analysis where we shuffled the trial labels for each neuron and computed the main and interaction effects on this shuffled data. We repeated this process 1,000 times and took the average across neurons for each run. Then we calculated the 95% and 99% percentiles of this shuffled data and compared those values to the real mean main and interaction effects across neurons to determine whether these effects were significantly stronger than what would occur by chance.

Main effects were significantly greater than chance for all positions in the display during both the original experiment (Fig. 24A; bootstrap, $p < 0.01$) and its vertical mirror inversion (Fig. 24B; bootstrap, $p < 0.01$). Likewise, interaction effects, were greater than expected by chance across all parts of the display for both the original (Fig. 24A; bootstrap, $p < 0.01$) and inverted (Fig. 24B; bootstrap $p < 0.01$) experiments. Therefore, we may conclude that non-adjacent and non-connected parts of crowded objects interact and all parts of crowded displays contribute main effects.

To directly examine the effect of part location on main and interaction effect magnitude, we turned to the sub-population of 30 neurons that were recorded in the context of both experiments. First, to examine whether the pattern of main effects varied as a function of letter

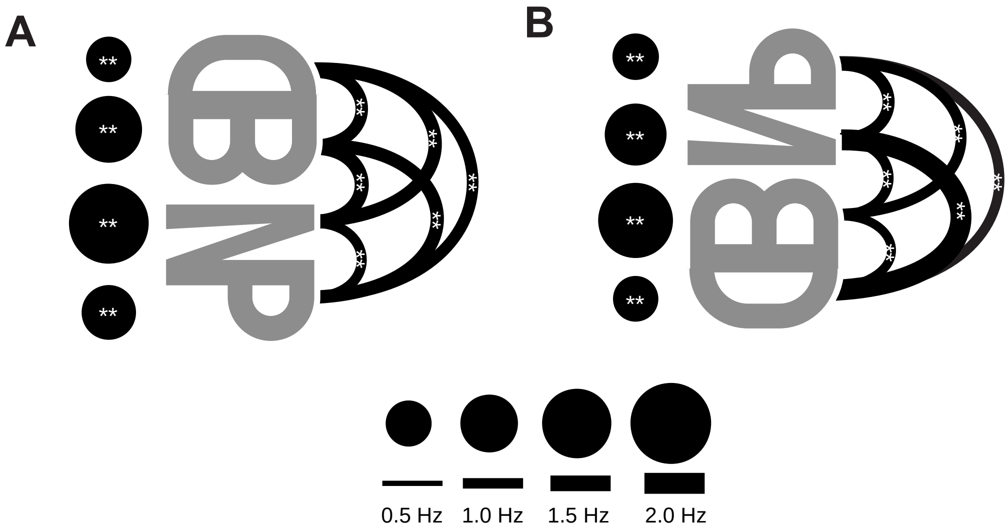

Figure 24. Experiment 3. Non-adjacent interactions are tied to space. **A**, Main effects (circles) and interaction effects (curved lines) are shown as a function of the position of the element in the compound pair. Circle diameter indicates main effect strength and line thickness indicates interaction effect strength. **B**, Main and interaction effects for the vertically inverted stimuli. Conventions as in A. Significance was determined from a bootstrap analysis in which firing rates for all neurons were shuffled 1,000 times and mean interaction and main effects across the population were calculated on each iteration to get a null distribution. The 95th and 99th percentiles were used as cutoffs for * and ** designations, respectively.

position in the array or whether the arrays were presented in either their original configuration or as vertical mirror images of those displays, we set up a nonparametric two-way ANOVA with four levels for the factor representing letter position and two levels for the factor representing vertical mirror images. We found no significant difference of main effect strength either as a function of letter position (two-way repeated measures ANOVA, $F(30) = 0.25$, $p = 0.84$) or vertical mirror images (two-way repeated measures ANOVA, $F(30) = 0.72$, $p = 0.40$). There was also no significant interaction effect between letter position and vertical mirror inversion (two-way repeated measures ANOVA, $F(30) = 0.26$, $p = 0.86$).

Considering the monkeys independently did not alter the results. For monkey 1, main effect strength was the same at all positions in the array (two-way repeated measures ANOVA, $F(13) = 1.24$, $p = 0.27$) and was not significantly affected by vertical mirror inversion of the entire display (two-way repeated measures ANOVA, $F(13) = 0.10$, $p = 0.76$). There was also no significant interaction effect between letter position and vertical mirror inversion (two-way repeated measures ANOVA, $F(13) = 0.78$, $p = 0.52$). Similarly for monkey 2, main effect was not significantly modulated by letter position (two-way repeated measures ANOVA, $F(17) = 0.64$, $p = 0.59$) or vertical mirror images (two-way repeated measures ANOVA, $F(17) = 1.33$, $p = 0.27$). There was also no significant interaction effect between letter position and vertical mirror inversion (two-way repeated measures ANOVA, $F(17) = 0.28$, $p = 0.84$).

As with main effects we can ask whether interaction effects were sensitive to the relative position between pairs in the array or vertically flipping the whole array. The factor representing pair distance had six levels, corresponding to the six possible pairwise interactions. The factor representing vertical mirror flips again had two levels. As with main effects, there were no significant impact of either the distance between parts (two-way repeated measures ANOVA, *F*(30) = 0.25, *p* = 0.94) or vertical mirror inversion (two-way repeated measures ANOVA, *F*(30) = 1.29, *p* =0.27) on interaction effect strength. There was also no significant interaction effect between pairwise position and vertical mirror inversion (two-way repeated measures ANOVA, *F*(30) = 0.78, p = 0.61). What this tells us is that the relative locations of parts within crowded displays do not affect the strength of main and interaction effects in IT neuronal populations.

Again, the pattern of results was not different when the two monkeys were considered independently. For monkey 1, there was no significant effect of distance between parts on the strength of their interaction (two-way repeated measures ANOVA, *F*(13) = 0.75, *p* = 0.59). Neither was there a significant effect of vertical mirror inversion (two-way repeated measures ANOVA, *F*(13) = 0.027, *p* = 0.87). Likewise, there was no significant interaction between part distance and mirror inversion (two-way repeated measures ANOVA, *F*(13) = 1.36, *p* = 0.25). When considering monkey 2 alone, the pattern remained the same. Distance between parts did not modulate interaction strength (two-way repeated measures ANOVA, *F*(17) = 0.48, *p* = 0.79). Vertical mirror inversion also had no significant impact on interaction effects (two-way repeated measures ANOVA, *F*(17) = 1.62, *p* = 0.22). Finally, there was no significant interaction between pairwise distance and mirror inversion (two-way repeated measures ANOVA, *F*(17) = 0.77, *p* = 0.58).

Overall, these results indicate that object parts within crowded displays interact promiscuously with all other parts in the display. This effect was not specific to particular stimuli falling at particular retinotopic locations because inverting displays did not significantly alter the results. Thus, all the parts contribute to the jumbled percept characteristic of crowding.

3.4 DISCUSSION

We have carried out three complimentary experiments in macaque monkeys to determine how visual crowding affects neuronal object representations. The key findings are as follows. First, as the space between stimuli decreased, neuronal selectivity for target letters both weakened quantitatively and changed qualitatively (Fig. 17). A second experiment demonstrated that this qualitative change can be captured in terms of both main (Fig. 20A,B) and interaction effects (Fig. 20C,D), however only interactions followed the size-invariance characteristic of crowding. The strength of interaction effects increased as the absolute spacing between elements decreased and then plateaued. Main effects, on the other hand, were driven by relative spacing between elements. A third experiment revealed that all parts of the stimuli within crowded arrays interact (Fig. 24). These findings are the first to demonstrate how crowding alters object representations at the single neuron level.

3.4.1 Distinguishing between Models of Crowding

These results enable us to test the many competing models of crowding. One popular account of crowding is that signals about closely-spaced peripheral stimuli are averaged together (Parkes et al., 2001). This averaging could easily be instantiated in neuronal populations by summing feedforward inputs across the relatively large receptive fields found in higher levels of the ventral stream (van den Berg et al., 2010). We did observe a weakening in the strength of selectivity as a function of crowding (Fig. 16, 20A), but the qualitative changes we observed in the neuronal code indicate that averaging isn't the whole story (Figs. 17,18, and 20).

Another popular account of crowding is that peripheral clutter is encoded as texture (Balas et al., 2009; Freeman and E. P. Simoncelli, 2011). Intuitively, one can imagine texture, as well as crowded arrays, as “hav[ing] lost form without losing crispness” (Lettvin, 1976). Given this definition, it’s no surprise that IT – an area concerned with form representation – is weakly selective between textures compared to V4 (Rust and Dicarlo, 2010) or V2 (Freeman et al., 2013; Ziemba et al., 2016). If crowded arrays were perceived as texture then we should have observed a weakening of target selection strength, which indeed we did (Fig. 17). But again, the more profound and surprising result we showed was a qualitative change in neuronal preferences. Because no prior neurophysiology study has directly compared the pattern of firing rates evoked by objects to that of their texturized versions (Portilla and E. Simoncelli, 2000) it is not possible to say whether textures would produce the kind of qualitative change in the neuronal code that we observed between crowded and uncrowded displays. As such, we cannot rule out the idea that crowded stimuli are represented as textures. Moreover, our finding that all parts of crowded arrays interact (Fig. 24) seems to support this idea.

Yet another model of crowding centers on attention (He et al., 1996). The idea is that crowding arises when the attentional spotlight is no longer able to isolate a single stimulus (Cavanagh et al., 1999). The recent finding that the N1 EEG component inversely correlates with performance during a crowding task supports this assertion (Herzog et al., 2015). Only one of our animals in one of our experiments was indisputably deploying attention during neuronal recordings, and the results were essentially the same for both animals, so we cannot make strong claims about the role of attention. However, the fact that qualitative neuronal preference changes occurred at absolute spacings beyond the reach of crowding (Fig. 20B) suggests that the brain can overcome code confusion, and thus there may not be an inescapable loss of bottom-up information, as has been suggested by opponents of the attention hypothesis (Freeman and E. P. Simoncelli, 2011). Another piece of evidence in favor of the attention hypothesis is the latent improvement in neuronal selectivity for crowded targets exhibited by both monkeys during experiment 1, in which they were trained to attend the target (Fig.15H,I and 16E,F). If crowding were due solely to feed-forward information loss then no such improvement should have been observed. The target selection latency under crowded conditions is on par with what has been seen previously in IT during an attention task (Chelazzi et al., 1993).

Finally, crowding has been considered to result from feature mislocalization (Wolford and Shum, 1980) or substitution of a flanker for a target (Freeman et al., 2012). Conceptually, this model has been described as features becoming unglued from space (Pelli and Tillman, 2008). Our main finding that the neuronal code changed qualitatively between crowded and uncrowded displays generally supports this model. Feature recombination could seemingly flip a neuron's preference from one stimulus to another (Fig. 21). The fact that interactions occurred across all parts of crowded arrays regardless of which part was where (Fig. 24) also lends support for this model because it relies on a system in which features are pooled across space with equal weight (Wolford, 1975).

At first blush it may appear that our results do not come down strongly either in favor of or against any of these theories, and in fact this is an endemic problem that has plagued the field of crowding since its infancy. Despite recent attempts to find a grand unifying theory of crowding (Harrison and Bex, 2015; Keshvari and Rosenholtz, 2016), it has been called into question whether this quest is misguided (Agaoglu and Chung, 2016). Pitting the competing models against each other, Agaoglu and Chung found that none alone was sufficient to explain the entire crowding phenomenon. Throughout the course of our own study, using the novel approach of neurophysiology, we have reached essentially the same conclusion. Crowding has proven to be far too complex a phenomenon to be captured by any of the existing models.

3.4.2 Size Invariance

While the qualitative change we observed for crowded conditions in experiment 1 seemed to be a reasonable neuronal correlate of crowding (Fig. 17), the lack of size invariance of this effect in experiment 2 (Fig. 20A-D) raised a few concerns. Size invariance is, after all, considered one of the hallmarks of crowding (Pelli et al., 2004). In the previous section, we interpreted these results as support for the attention hypothesis of crowding because presumably the large letters with narrow spacing were above the reach of crowding, so attention should be able to overcome the qualitative changes in main and interaction effects. Without behavior we cannot say this for certain, but the lack of an effect of spacing on latency as well as the equal status of elements as targets argues that attention is not being deployed during this task.

An alternative interpretation of these results is that the qualitative change in main and interaction effects is not a neuronal correlate of crowding after all. Instead, it could simply be a general feature of object representation. Previous investigations of multiple objects or multiple object parts in IT have not reported the breakdown of divisive normalization that we saw as a function of spacing (Sripati and Olson, 2010a; Zoccolan et al., 2005), but this is also the first study to systematically vary the distance between a variety of peripheral stimuli. Perhaps what we observed in experiment 2 is simply a new manifestation of scale invariance, which is a well-known characteristic of IT neurons (Ito et al., 1995). It’s possible that the two-element displays in experiment 2 were interpreted by the animals as a single object, scaled up or down across conditions.

Interaction effects also reflected the qualitative preference change, and much more in line with the behavioral characterization of crowding we did observe differences across scale (Fig. 22B). Like crowding, absolute spacing seemed to dictate within-spacing defined interaction strength, which sharply increased and then plateaued as absolute spacing decreased (Fig. 20C-D), although this plateau occurred at a much larger spacing than what we observed previously in behavior (Crowder and Olson, 2015)(Fig. 15A). There are two possible explanations that make this incongruence less damning. First, given that the task design was different between the first two experiments, it could be the case that the critical spacing over which crowding operates is much larger than in the original experiment. In line with this thought is the finding that using fewer flankers increased critical spacing (Banks et al., 1979). Second, because monkeys were not deploying attention during the second experiment, this might have had the effect of making critical spacing appear larger, just as attention has been said to “shrink” receptive fields (Rolls et al., 2003). However, without a behavioral readout on this task we can only speculate about the link between interaction effects and crowding.

3.4.3 Limitations and Future Directions

Although we uncovered some novel and surprising results, the present study is not without caveats. First, there are many different versions of behavioral tasks designed to show the crowding phenomenon and just as much variance in the findings, so our results may well not generalize. Some researchers use oriented gratings (Anderson et al., 2012; He et al., 1996; Parkes et al., 2001), others oriented letters (Flom et al., 1963a; Harrison and Bex, 2015; Tripathy and Cavanagh, 2002), and still others used upright letters of the English (Bouma, 1970) or Armenian (Freeman and Pelli, 2007) alphabets. Some labs measured crowding in terms of threshold contrast (Pelli et al., 2004;

Strasburger et al., 1991), whereas others relied on the psychometric curve (Chung, 2007; Tripathy and Cavanagh, 2002; Yeshurun and Rashal, 2010). Based on the observation that crowding occurs across visual domains such as orientation, hue, and saturation (van den Berg et al., 2007) so long as flankers and targets are similar (Põder, 2007), we developed a task for nonhuman primates (Fig. 14) that should meet the criteria for crowding. However, experiment 1 in the present study is admittedly limited in scope compared to the vast variations of past studies.

Our first goal with creating a crowding task for nonhuman primates was that the behavior had to resemble that of humans (Crowder and Olson, 2015) and it must also facilitate neuronal recordings in IT. This balance between relevance to the human crowding literature as well as practical considerations inherent in finding selective neurons, compounded by the fact that no one has attempted to study crowding in an awake behaving monkey before, meant that the data from experiment 1 were too complex to draw strong conclusions from on its own.

A subsequent, much simpler task (Fig. 19) was developed to address this concern; however, this task lacked behavior. It was an omission that proved problematic when it came time to understand the size invariance of the qualitative neuronal code change (Fig. 20A,B, Fig. 22A), precise absolute spacing effects for interactions, and most glaringly, how these neuronal findings relate to attention. To a human observer it is clear that the large letters with narrow spacing are discriminable in the periphery, and while the same is probably true of monkeys (Crowder and Olson, 2015), we have no direct evidence in this specific experiment. Also, because the number of distractors varied between experiments 1 and 2 we cannot say for sure whether critical spacing was the same, especially in light of evidence that distractor numerosity has been shown to reduce critical spacing (Banks et al., 1979).

Another limitation of our study is that we designed experiments 2 and 3 to be amenable to main and interaction effect measurements, which we expected to be telling. In fact, their modulation with relative spacing and tenuous connection with absolute spacing proved to be confusing. Future investigations into the neurophysiological basis of crowding should strive to attack the hypotheses we outlined in the previous section head-on so as to eliminate the ones that still remain standing. In particular, it would be interesting to investigate how texture derived from objects (Portilla and E. Simoncelli, 2000) is encoded in both peripheral and foveal vision compared to the original objects. Another potential avenue would be an explicit exploration of the feature mislocalization hypothesis of crowding (Wolford, 1975). One could record neuronal responses while requiring animals to report the object-relative location of features or report the identity of the letter in the center of an array made up of targets and distractors drawn from the same pool.

Finally, our study didn’t touch one of the most striking and neglected characteristics of crowding, which is the radial-tangential anisotropy of crowding zones (Toet and Levi, 1992). These elliptical zones have an uncanny resemblance to saccadic eye movement endpoint scatter (Harrison et al., 2013; Nandy and Tjan, 2012), which also ties into the attention hypothesis of crowding by virtue of the fact that the same brain structures, such as the Frontal Eye Field (FEF), that command saccades also direct spatial attention (Moore, 2003). FEF and IT have strong reciprocal connections (Schall et al., 1995), and inactivating FEF reduces IT neuronal selectivity for peripheral objects falling in the lesion site (Monosov et al., 2011), so it would be interesting to see how either sub-threshold stimulation or inactivation of FEF would affect IT neuronal responses to crowded displays.

3.4.4 Conclusions

This work constitutes the first investigation into the neuronal mechanisms of visual crowding in a nonhuman primate. As the last four decades of crowding research have generally gone, we found the neuronal data to be more puzzling and strange than we ever imagined. Crowding is not merely the reduction in strength of neuronal selectivity for the target object. Instead, the preferences of IT neurons changed qualitatively while preserving a good portion of the selectivity between crowded arrays. Just as with subjective experience, crowded letters don't fade or disappear, but rather they transform. We take this as support for models that explain crowding as a devolution of objects into texture (Rosenholtz et al., 2012) or as mislocalization of object features (Strasburger and Malania, 2013). The latency of target selectivity under crowded conditions suggests that what sets the limit on peripheral resolution is not bottom-up pooling, but rather top-down attention. Having taken this first step toward finding a neuronal correlate of crowding, future work may address models of crowding at the single neuron level more directly.

4.0 INFEROTEMPORAL NEURONS BREAK THE LAW OF SIMPLICITY

The visual system is tasked with the tricky job of taking in sometimes ambiguous information and constructing an internal representation that enables some degree of understanding. One way in which the brain might achieve this is to invoke the Gestalt law of simplicity, which states that the visual system interprets input with from the simplest possible explanation. A classic example of the law of simplicity in action is the case of overlapping shape outlines. The natural tendency is to perceive this composite figure as the set of shapes originally used to construct it, but there are many other possible — albeit more complicated — interpretations. We hypothesized that monkey inferotemporal cortex neurons would also decompose such composite figures into their natural constituents. What we found, however, was exactly the opposite. The neuronal representation most resembling the composite figure was actually the external contour, tracing the overall footprint of the composite figure. This effect was especially true early in the stimulus presentation period. We interpret this finding as more support for the global advantage effect, and argue against an inferotemporal object code based on the law of simplicity.

4.1 INTRODUCTION

Any theory of perception must deal with the basic fact that two-dimesional projections onto the retina permit many alternative interpretations. Therefore, the visual system must impose its own constraints to allow for the convergence on a single explanation for what the eyes are taking in. One such constraint is known as the law of simplicity, which lies at the heart of the Gestalt school of perceptual organization (Wertheimer, 1923).

Simplicity hinges on the belief that the visual system interprets its available information with the simplest explanation possible. For instance, it’s simpler to imagine that a horse occluded by a tree is a single, whole animal rather than two half-animals. The law of simplicity is conjectured to underlie all other laws of perceptual organization (Wertheimer, 1923). From the very beginning, the simplicity principle was put forward in opposition to the likelihood principle suggested earlier by Helmholtz, which has persisted as the predominant competing theory. Helmholtz’s likelihood principle suggests that the visual system interprets input as being derived not from the simplest interpretation, but from the most likely, based on prior experience (1925). Returning to the horse in the forest example, this theory states that the tree is considered an occluder and the horse considered whole because that’s far more likely than any alternative interpretation. Given this rather mundane example, both likelihood and simplicity are congruent, but what happens when we decouple them such that the simplest interpretation is unlikely? Leeuwenberg and Boselie proposed such a Gedankenexperiment by constructing the silhouette of a horse with a head on both ends (1988). The most likely interpretation is that there are two horses standing side by side facing opposite directions. The simplest interpretation, however, is that this is a strange animal with two heads, by virtue of symmetric information being redundant. This example seems to demonstrate that because likelihood and simplicity give rise to different predictions they cannot be reconciled. However, at the level of theory they converge (Chater, 1996).

Likelihood is best and most commonly described within a Bayesian framework. Bayes' rule takes in the prior probabilities of various scenarios as well as the current observations and spits out the most likely interpretation of these data (Bayes et al., 1763). Even though there is no explicit penalty for complexity, Bayes' rule tends to favor simpler interpretations of the data (MacKay, 1992). Coming from the other side, structural information theory (SIT) is explicitly built upon the Gestalt principle of simplicity (Pomerantz and Kubovy, 1986), yet as a side-effect it exhibits the veridicality of stimulus identity that is characteristic of the likelihood model (Wagemans et al., 2012). Bayes rule and SIT can even be shown to be mathematically equivalent (Chater, 1996). So even though simplicity arose as a reaction to the likelihood hypothesis, these two models of perception are not mutually exclusive and may even be interdependent. How then does one explain the two-headed horse (Leeuwenberg and Boselie, 1988)? Even though such an animal is unlikely to occur based on raw frequency of sightings, the lack of depth and boundary cues do suggest the likelihood that this is one continuous object (Chater, 1996). However, this argument is purely theoretical and the empirical evidence to settle the debate is still lacking.

Many studies have found evidence for various gestalt laws subsidiary to simplicity, such as grouping, illusory contours, common fate, similarity, and "seeing the forest before the trees" (Bona et al., 2014; Lee and Nguyen, 2001; Martin and Heydt, 2015; Sáry et al., 2007; Sripati and

Olson, 2009; Wannig et al., 2011; Zaretskaya et al., 2013; Zhou et al., 2000), but none have explicitly tested the neuronal basis for the law of simplicity.

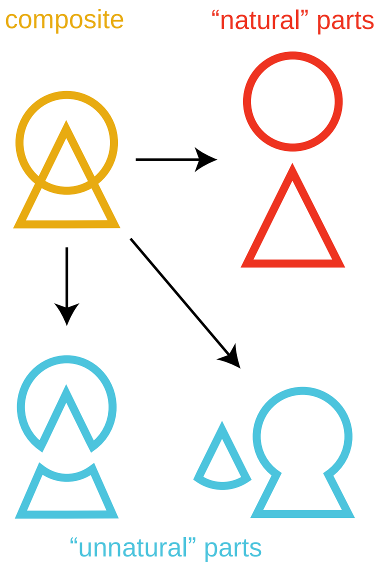

Figure 25. The law of simplicity. There is a natural tendency to interpret a composite figure made of overlapping shapes (yellow) as a collection of the “natural” parts used to construct it (red). However, there are other possible interpretations, which involve decomposing the composite into “unnatural” parts (blue). The tendency to perceive natural parts and overlook unnatural parts is evidence for the Gestalt law of simplicity.

To tackle this mother of all Gestalt laws head-on we chose to leverage the problem that inspired the SIT model in the first place (Pomerantz and Kubovy, 1986). When shape outlines overlap the composite (Fig. 25A, yellow) humans still tend to perceive the “natural” parts used to construct it in the first place (Metzger, 1936). However, this is not the only possible interpretation. One can also imagine decomposing this composite figure into the “unnatural” parts typically hidden from conscious perception (Fig. 25, blue).

Since neurons in IT tend to track conscious visual perception while subjects view and report on bistable stimuli (Sheinberg and Logothetis, 1997), we hypothesized that these neurons might also preferentially encode the natural parts of compound stimuli, and thus following the law of simplicity. We tested this by comparing the responses to each part and set of parts against the response to the composite. What we found was quite the opposite. The representation most resembling the composite was actually the external contour, tracing the global footprint of the composite (Sripati and Olson, 2010b), especially early in the trial. We take this as more support for the global advantage effect (Sripati and Olson, 2009), and argue against an inferotemporal object code based on the law of simplicity.

4.2 MATERIALS AND METHODS

4.2.1 Animals and Equipment

Two adult male rhesus macaque monkeys (Macaca mulatta) were used in these experiments (monkey 1 and monkey 2). Experimental procedures were approved by the Carnegie Mellon University Institutional Animal Care and Use Committee and were in compliance with the United States Public Health Service Guide for the Care and Use of Laboratory Animals. Before the recording period, each monkey was surgically fitted with a cranial implant and headpost (Crist Instrument). After initial training, a 2 cm-diameter vertically oriented cylindrical recording chamber (Crist) was implanted over the left hemisphere in both monkeys. In both animals, MRI brain scans were used to position the chamber mediolaterally above the superior temporal sulcus and rostrocaudally above anterior medial temporal sulcus.

For behavioral testing, each monkey was seated in a primate chair with the head stabilized using the headpost. Events during each trial were controlled by Cortex software (NIMH). Visual stimuli were presented on a 17” LCD screen with 1024 x 768 pixels of resolution positioned 18” from the animal’s eyes. The precise time at which images appeared on the screen was recorded using a photodetector circuit (designed by NIMH and built in-house). Eye position was tracked by an infrared system (ISCAN). The system was calibrated by requiring the monkey, at the beginning of each block of trials, to fixate a small target presented successively at four locations corresponding to the corners of a 14° x 14° square centered on the screen. Offline, the readings on each trial were converted to degrees of visual angle by performing a linear transformation based on the stored calibration voltages.

Each day's recording session would begin with the insertion of a varnish-coated tungsten microelectrode with an initial impedance of 1.0 MΩ at 1 kHz (FHC) into the temporal lobe through a transdural guide tube advanced using a hydraulic microdrive (Narishige). When mapping a new track electrodes were lowered to a depth such that its tip was 10 mm above the superior temporal sulcus, as estimated from MRI images of each animal's brain. Using a grid inside the chamber with 1mm spacing between holes (Crist) the electrode could be advanced reproducibly along the same tracks day to day. The action potentials of a single neuron were isolated online by means of a commercially available spike-sorting system (Plexon). All waveforms were recorded during the experiments and final spike sorting was performed manually offline.

Neurons were probed first with a set of 32 colorful photographs of objects to see whether they were visually-responsive. If so, they were further tested with the four composite stimuli (Fig. 26A, yellow shapes), which ensure that neurons and stimuli were selected in a way that remained agnostic to experimental questions. Stimuli were chosen to maximize both mean firing rate. The composite that elicited the highest firing rate was chosen, along with all of its constituent parts (Fig 26A).

4.2.2 Task and Training

Monkeys were trained to maintain fixation within a 2° by 2° window around a central fixation point while shapes flashed on the screen at 2° eccentricity in the right visual hemifield. After fixating for 200ms a stimulus appeared for 200ms, followed by a 500ms gap (Fig. 26B). If fixation was maintained throughout the stimulus period the animals received a small juice reward and were allowed to look around freely until the fixation spot reappeared. All stimuli were presented in a pseudorandom order so that each would be shown 8 times. Incomplete trials were repeated later in the block.

Compound stimuli were formed from two overlapping shape outlines, which together spanned 2.5° horizontally and 3.5° vertically (Fig. 26A, yellow). Then those compound stimuli were decomposed into all of the constituent closed parts possible, so as not to introduce closure as a confound. The parts could either be the ones that the compound was initially composed of, which we dubbed “natural” parts (Fig. 26A, red), whereas other parts were labeled “unnatural” (Fig. 26A, blue). The natural parts were always 2.5° by 2.5° across whereas the unnatural parts could vary in size. The first category of unnatural parts was made by bisecting the compound stimulus along the vertical axis where the constituent shapes intersect (Fig. 26A, leftmost blue). The second kind of unnatural decomposition was made by separating the external contour of the compound from the internal contour (Fig. 26A, rightmost blue). The key difference between the natural and unnatural part designations is that to get the natural parts from the compound one must invoke the gestalt law of simplicity, which states that people tend to interpret ambiguous or complex images as being composed of the simplest forms possible (Metzger, 1936). The natural parts are more simple by virtue of having fewer abrupt direction changes (i.e., corners) that the unnatural shapes. Parts were always presented at the same location on the screen as they appeared in the composite, which was intended to avoid any influence that spatial location of particular features may have on the neuronal responses.

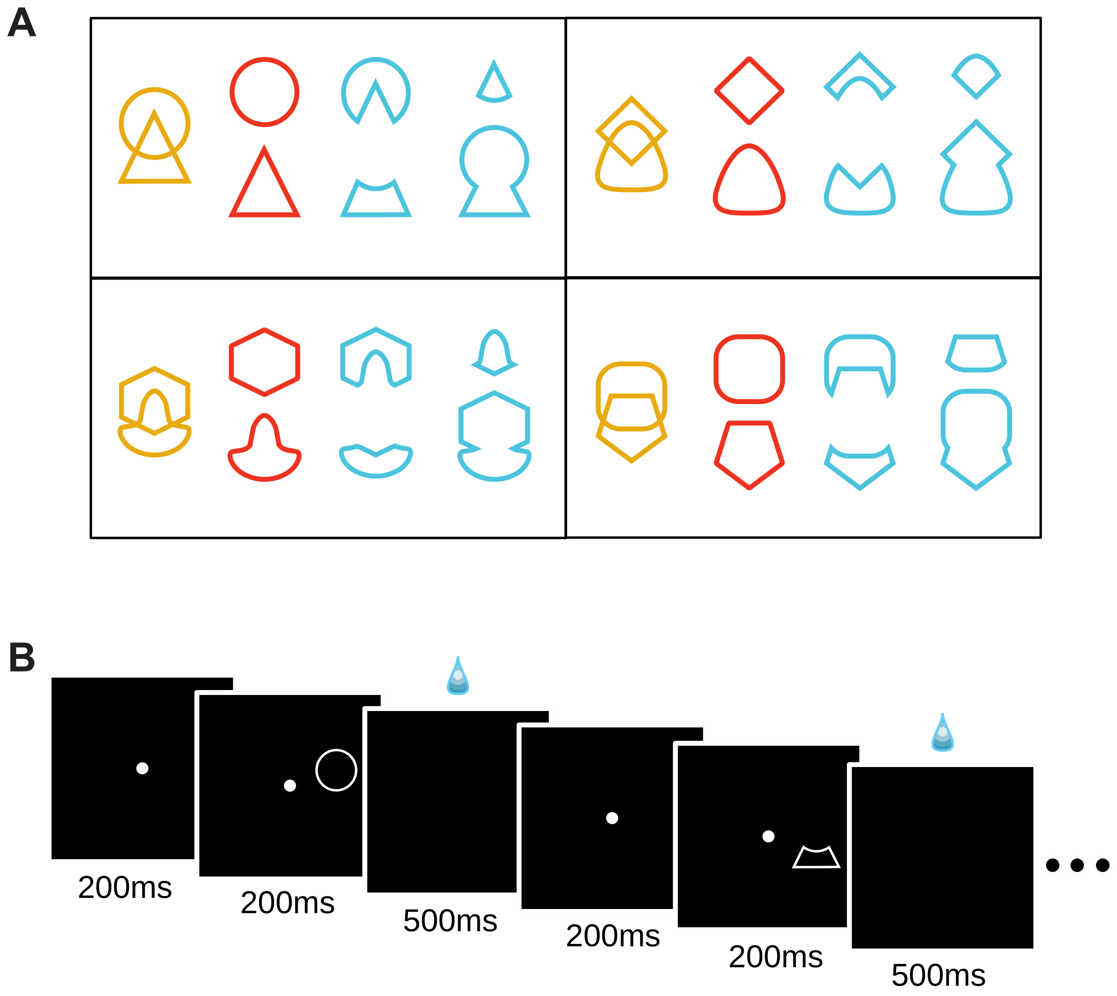

Figure 26. Passive fixation task. **A**, Four possible stimulus sets, each comprised of the composite (yellow) as well as its corresponding natural (red) and unnatural (blue) parts. **B**, Task sequence, which repeated with pseudorandom order until each shape had been presented 8 times.

4.2.3 Data Analysis

Neurons were only considered for analysis if they fired at a significantly higher rate in the period 70 to 270 ms after stimulus onset compared to the baseline period, -100 to 50 ms. Significance was assessed using a Wilcoxon signed rank test with $\alpha = 0.05$. Unless otherwise noted, this same epoch, 70 to 270ms after stimulus onset, was used for all analyses.

Because the central question asks whether IT neurons encode composites as if they are made up of the natural parts, we need a way to measure how similar the population of neurons considers these images to be. To do this we measured the Euclidean distance between each pair of images in multidimensional neuronal activation space. The larger the distance, the less similar the population considers those images to be. The advantage of the neuronal activation space approach is that it gets around the issue of specific neuronal preferences and interrogates the population as a whole, much the way a downstream brain area might.

Euclidean distance (ED) can be visualized in a time-varying, stimulus-aligned manner similar in spirit to a peristimulus time histogram (PSTHs). To do this, firing rates were first aligned to stimulus onset and then the ED between pairs of images was computed for each 5ms bin. Another way we visualized similarity between the stimuli was with multidimensional scaling, which finds a set of 2D coordinates for each stimulus such that the 2D distances are as close as possible to the distances measured in the full activation space. This was implemented with a built-in function in MATLAB (mdscale) after first standardizing the firing rates into z-scores, as required by the model. The quality of this 2D representation was assessed as percent of total variance across the multidimensional space captured by the first two dimensions. This can be computed by dividing the sum of the eigenvalues for the first two dimensions into the sum of all the eigenvalues.

Another way to examine similarity between stimulus representations is with a dendrogram. For this we used another built-in MATLAB function (linkage) to convert z-scored mean firing rates into hierarchical clusters and then another function (dendrogram) was used to visualize these clusters. A dendrogram consists of a hierarchical series of inverted U-shaped connectors between clusters. The shorter the U (dubbed cophenic distance), the closer the clusters. To determine whether a given dendrogram was a good representation of the actual similarity between neuronal representations, we computed the Pearson correlation between the cophenic distances in the tree and the actual distances in the full neuronal activation space (cophenet in MATLAB). Like any correlation, this value could vary from 0 (poor dendrogram) to 1 (perfect dendrogram).

To assess the degree to which the firing rate of the composite reflected the sum of its various pairs of parts, we first computed the average firing rate of the composite for each neuron. Then we computed average firing rate for one part as well as the average firing rate evoked by its complement. If the whole is equal to the sum of the parts then these values for each neuron should lie close to the unity line, which we assessed using a Wilcoxon signed-rank test.

We also examined the degree to which the composite representation resembled that of the parts at different epochs following visual stimulation. We defined three epochs on the basis of the ED versus time plots. Epoch 1 spanned 80 to 150ms. Epoch 2 spanned 150 to 270ms. Epoch 3 spanned 270 to 300ms. The epochs were initially selected visually, but each was confirmed to reflect a significant difference between the ED for the outline of the composite and the ED of the other parts. Within each epoch, we computed z-scored mean firing rates of each neuron for each part and then calculated the correlation coefficient between those values and the z-scored mean firing rates evoked by the composite during the same epochs.

To completely rid our measures of distance in neuronal activation space from possible firing rate confounds, we also computed distance in terms of the angle between population firing rate vectors. Distance between two stimulus conditions could then be calculated as the shortest great circle path between the corresponding population vectors. We used the following equation:

$$\theta_{i,j} = \cos^{-1}(\hat{r}_i \cdot \hat{r}_j) \qquad \text{Eq. 4}$$

where $\hat{r}_i$ is the unit vector of the population firing rate for stimulus i, and $\hat{r}_j$ is the unit vector of the population firing rate for stimulus j. The output, $\theta_{i,j}$, is the angular distance between the two population vectors, and it must lie within 0 to π radians.

4.3 RESULTS

We recorded 37 visually-responsive neurons from monkey 1 and 30 from monkey 2. The main question concerned whether the representation of the composite image more closely resembled that of the natural parts, which would indicate adherence to the law of simplicity. Our first major test was whether the average Euclidean distance in neuronal activation space between the natural parts and composite was smaller than the distance between the unnatural parts and the composite. This turned out not to be true (Fig. 27A). The distance between the neuronal representation of the composite and its constituent parts was not significantly different for natural versus unnatural decompositions either in the composite data (Wilcoxon signed rank, *z*(67) = 0.55, *p* = 0.58) or for monkey 1 (*z*(37) = 0.46, *p* = 0.73) or monkey 2 (*z*(30) = 0.02, *p* = 0.98) considered separately.

Lacking evidence for the law of simplicity, we turned to the classic Gestalt adage that "the whole is different from the sum of the parts" (Koffka, 1935). We did observe this to be true for both natural (Fig. 27B) and unnatural (Fig. 27C,D) parts, with significant deviation from the sum in all cases (Wilcoxon signed rank, *z*(67) = 5.89, p = 4.0 x 10^{-9} for natural parts, *z*(67) = 6.34, p = 2.3 x 10^{-10} for vertically bisected unnatural parts, and *z*(67) = 6.35, *p* = 2.2 x 10^{-10} for nested contours). Rather than the sum, the average of the individual part responses was not significantly different from the response to the whole, regardless of whether it was deconstructed into natural parts (Wilcoxon signed rank, *z*(67) = 0.68, *p* = 0.50), vertically-bisected unnatural parts (*z*(67) = 0.14, *p* = 0.89), or nested contours (*z*(67) = 1.12, *p* = 0.26). This finding is in line with previous work using non-overlapping shapes (Sripati and Olson, 2010a; Zoccolan et al., 2005).

The same trend was present when both monkeys were considered separately. For monkey 1, the neuronal responses to the composite significantly deviated from the responses to the sum of any complementary set of parts (Wilcoxon signed rank, *z*(34) = 4.35, p = 1.4 x 10^{-5} for natural parts, *z*(34) = 4.76, p = 1.9 x 10^{-6} for vertically bisected unnatural parts, and *z*(34) = 4.78, *p* = 1.7 x 10^{-6} for nested contours). In contrast, the average was a much more reasonable model (Wilcoxon signed rank, *z*(34) = 1.12, p = 0.26 for natural parts, *z*(34) = 1.03, p = 0.30 for vertically bisected unnatural parts, and *z*(34) = 0.13, *p* = 0.89 for nested contours). For monkey 2, the neuronal responses to the sum was still a poor model (Wilcoxon signed rank, *z*(30) = 4.02, p = 5.6 x 10^{-5} for natural parts, *z*(30) = 4.23, p = 2.4 x 10^{-5} for vertically bisected unnatural parts, and *z*(30) = 4.17, *p* = 3.0 x 10^{-5} for nested contours). In contrast, the average was a much more reasonable model (Wilcoxon signed rank, *z*(30) = 0.26, p = 0.78 for natural parts, *z*(30) = 1.09, p = 0.28 for vertically bisected unnatural parts, and *z*(30) = 1.49, *p* = 0.14 for nested contours).

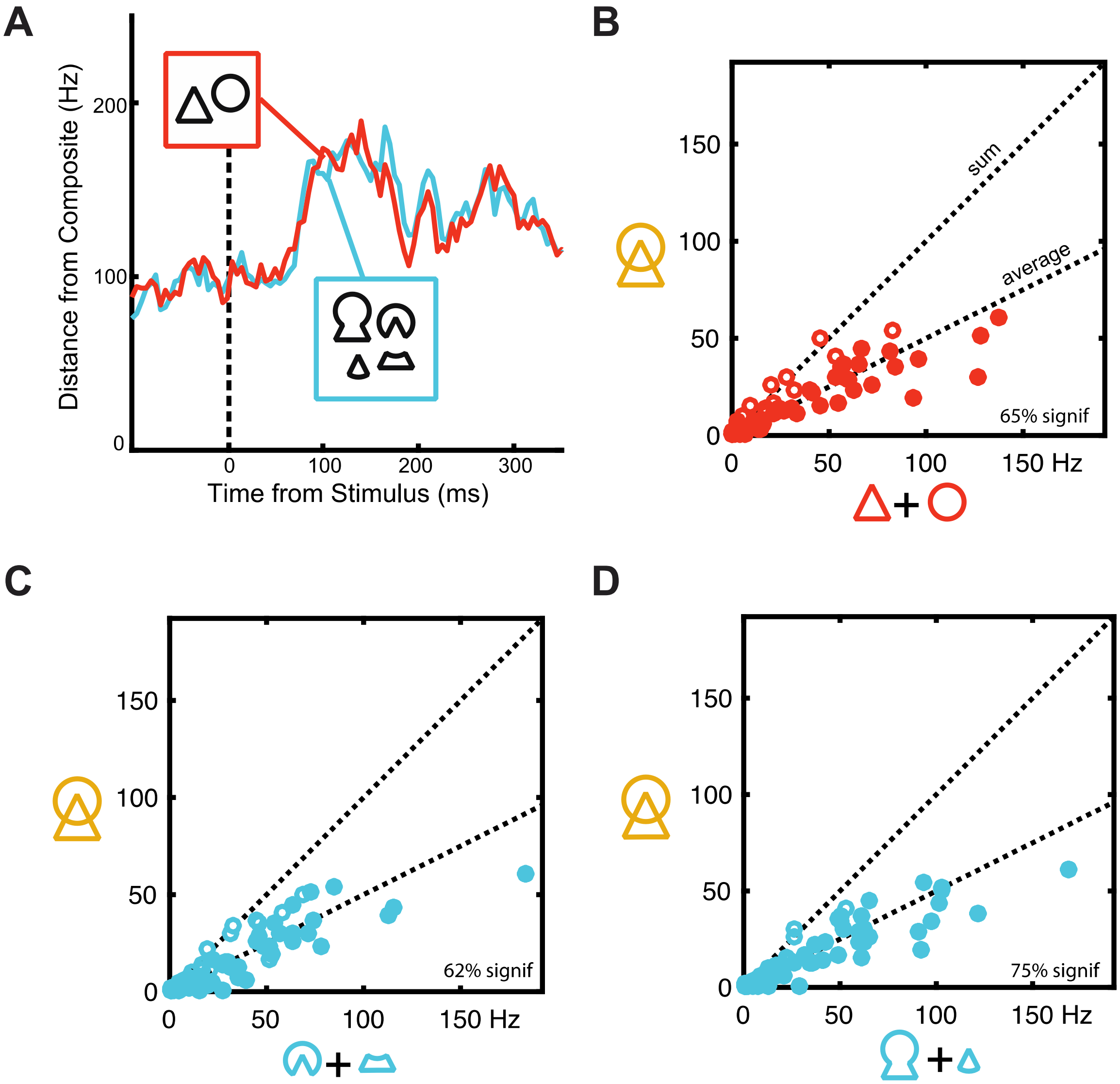

Figure 27. Composites are no more represented as natural parts than unnatural parts. **A**, Average euclidean distance in neuronal activation space between the composite and natural (red) versus unnatural (blue) parts as a function of time since stimulus onset. **B**, The firing rates evoked by the composite are not equal to the sum of the responses to the natural parts. Filled circles indicate neurons whose firing rates significantly deviated from the dashed line representing a summation model (Wilcoxon signed rank test, $p < 0.05$). The percentage of significant neurons is indicated in the lower right corner. **C**, The same analysis as in **B**, except comparing the unnatural parts created by separating the composite into upper and lower components. **D**, The same analysis again except using the unnatural parts consisting of the external and internal contours.

Having ruled out simplicity as the guiding rule behind the inferotemporal shape code, we took a step back and asked how the response to each individual part compared to that of the whole. Using multidimensional scaling (see Methods) to render inter-object distances on a 2D plane, we saw that the composite was closest to the external contour in neuronal activation space (Fig. 28A). For monkey 1 considered alone, the results were virtually identical, with 68% varance explained by the reduced dimensions. For monkey 2, the composite was still closest to the external contour, but the natural shapes no longer appeared to be the next closest (61% of variance explained).

Because collapsing the full 67-dimension neuronal activation space onto a 2D plane is bound to lose and distort information, we also analyzed the clustering in the full activation space using a dendrogram (see Methods). Just as before, the composite most closely clustered with the contour (Fig. 28B). The dendrogram representation correlated strongly with the distances between shapes in the full neuronal activation space (cophenic correlation, $r = 0.73$). When the data for the two monkeys was analyzed separately, monkey 1 again was nearly a carbon copy of the combined data (cophenic correlation = 0.58) while monkey 2 had some deviations in which parts clustered together (cophenic correlation = 0.78). Nevertheless, in both cases, the composite clustered most closely with the external contour.

One possible explanation for this effect is that the composite and external contour may simply evoke higher mean firing rates compared to the other parts due to their relatively large size and overall number of photon emissions. However, that doesn't seem to be the case because mean firing rate was not significantly different across conditions (Kruskal-Wallis test, $H(67) = 3.93$, $p = 0.69$; Fig. 28C). The same was true for monkey 1 alone ($H(37) = 3.92$, $p = 0.69$) as well as monkey 2 ($H(30) = 1.15$, $p = 0.98$).

Overall, composites are encoded more like contours than any other part, but does this effect vary over time? The propensity to perceive global shape over local details (Navon, 1977) or for IT neurons to encode global shape first (Sripati and Olson, 2009) suggests that this might be the case. If so, we also want to know whether all the other parts are represented equally later in the trial.

What we found is that the composite is indeed represented similarly to the contour in the early portion of the trial (Fig. 29A). During the first epoch, from 80 to 150ms after stimulus onset, the squared difference in mean firing rate between the composite and its contour was smaller than the mean squared firing rate difference between the composite and all the other parts (Wilcoxon signed rank test, *z*(67) = 2.53, *p* = 0.006). For the second epoch, from 150 to 270ms after stimulus onset, the contour was actually significantly farther away in neuronal activation space than the other parts (Wilcoxon signed rank test, *z*(67) = 1.86, *p* = 0.03). For the third epoch, from 270 to 300ms, the system

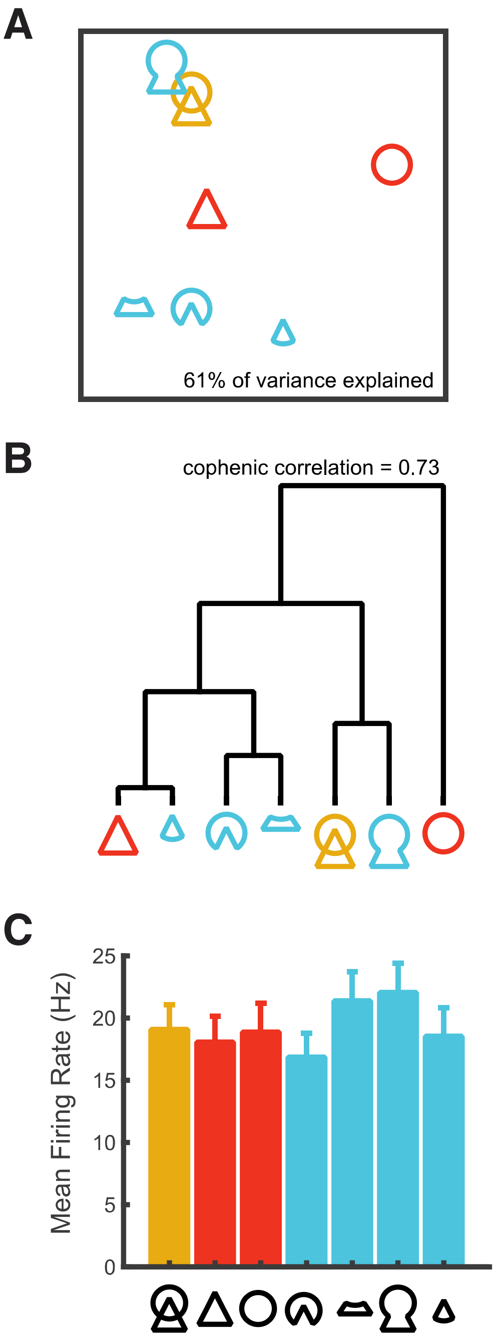

Figure 28. Composites more closely resemble the external contour than natural parts. **A**, Distance in neuronal activation space collapsed onto two-dimensions using multidimensional scaling. The external contour (blue keyhole) is closest to the composite (yellow). **B**, Dendrogram showing the same result. **C**, The above results cannot be explained by mean firing rate because it is not significantly different across conditions.

oscillated back toward a contour-heavy interpretation of the composite stimulus (Wilcoxon signed rank test, $z(67) = 2.70$, $p = 0.004$).

When considering the data from monkey 1 alone, the results followed the same trend and the difference between the curves was significant for epoch 1 (Wilcoxon signed rank test, $z(37) = 2.07$, $p = 0.02$), epoch 2 (Wilcoxon signed rank test, $z(37) = 1.76$, $p = 0.04$), and epoch 3 (Wilcoxon signed rank test, $z(37) = 2.29$, $p = 0.01$). For monkey 2 alone the data again followed the same overall trend, but the effect was just shy of statistical significance for all epochs (Wilcoxon signed rank test, epoch 1: $z(30) = 1.43$, $p = 0.08$, epoch 2: $z(30) = 0.91$, $p = 0.18$, epoch 3: $z(30) = 1.55$, $p = 0.06$).

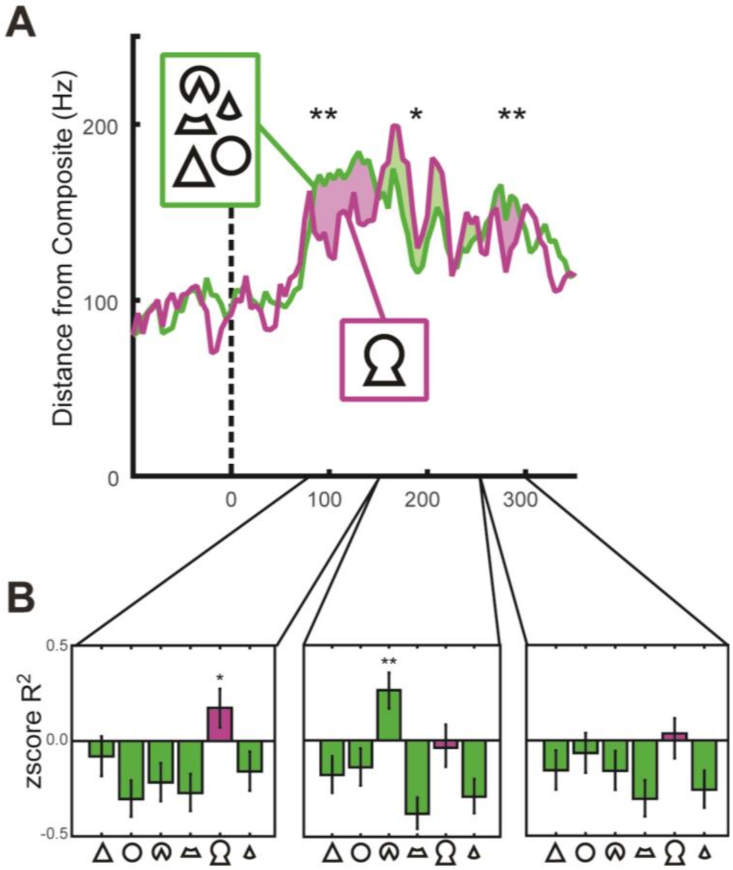

Figure 29. Composites resemble different parts over time. **A**, Euclidean distance as a function of the time after stimulus onset. The purple curve reflects distance between the composite while the green curve reflects the average distance between the composite and remaining parts. When the contour distance is less, the difference between the curves is highlighted in light purple. Likewise, when the distance is less for the remaining parts the area between the curves is highlighted in light green. Significant differences between the curves, pooled over shaded regions, are denoted with ** ($p < 0.01$). **B**, The correlation between z-scored firing rates evoked by each part versus the composite. Error bars represent the 95% confidence intervals. Significant positive correlations are denoted by * ($p < 0.05$) and ** ($p < 0.01$).

To see whether particular parts matched the composite better than others during the different epochs, we computed the correlation coefficient between the z-scored firing rate evoked by each part and that of the composite (Fig. 29B). For the first epoch, only the external contour

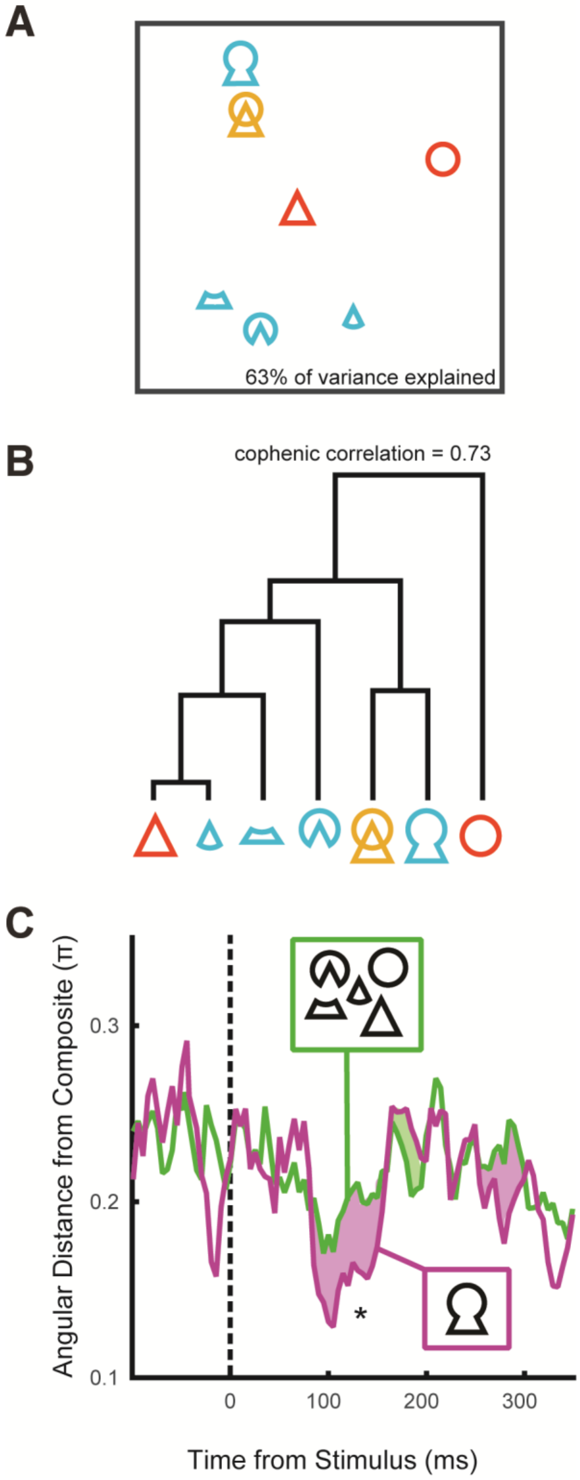

Figure 30. Population vector angle yields the same results. **A**, Multidimensional scaling reflects approximate population vector angle between stimuli. As in Fig. 28A the composite is closest to the external contour. **B**, Dendrogram of angular distances between stimuli. Compare to Fig. 28B. **C**, Timecourse of the angular distances between the composite and the external contour (purple) and the mean angular distance between the composite and the remaining parts (green).

was significantly positively correlated with the composite (Pearson correlation, $r(67) = 0.17$ $p = 0.03$). During the second epoch, the only part significantly positively correlated with the composite was the one corresponding to the top half of the composite figure (Pearson correlation, $r(67) = 0.25$ $p = 0.006$). For the third epoch, none of the parts were significantly positively correlated with the composite. There were significant negative correlations during all three epochs, but the interpretation of this finding is unclear.

Even though there were no significant differences in mean firing rate across stimulus conditions we were still concerned that firing rate covariation may be confounding our results. To completely remove firing rate from the equation we set aside Euclidean distance in favor of a distance measure that was instead based on the angle between the population firing rate vectors (Eq. 4). Since the angle and magnitude of population firing rate vectors are independent, we could be confident that any findings produced by this method were not due to firing rate contamination.

First, we repeated the multidimensional scaling analysis (Fig. 28A) using angular distance instead of Euclidean distance. Our results looked remarkably similar to the original analysis (Fig. 30A). The composite was still situated closest to the external contour. Next we constructed a dendrogram with these correlation-derived distances, and arrived at the same essential outcome as with the ED-based clustering (Fig. 28B). The representation of the composite was closer to that of the external contour compared to other parts (Fig. 30B).

Finally, we wanted to determine whether the temporal precedence of the contour-like representation was still present when examining angular distance between population vectors. The temporal dynamics remained intact (Fig. 30C), compared to the previous analysis (Fig. 29A). What differed between the two methods is that while the Euclidean distance between the composite and its constituent parts rose from baseline, the angular distance actually decreased from baseline. This is not surprising, but it serves to reinforce the rationale behind using angular distance as a metric unbiased by firing rate. Using a bootstrap analysis, we determined that the angular distance between composite and contour was significantly less than that of the other parts for the early epoch (1,000 permutations, $p = 0.04$) and trending toward significance for late epochs (1,000 permutations, $p = 0.07$). For the middle epoch, there was no significant difference in angular distance (1,000 permutations, $p = 0.59$). The early epoch effect remained significant when considering only the data from monkey 1 ($n = 37$, $p = 0.05$), but only trended toward significance for monkey 2 ($n = 30$, $p = 0.16$). In both animals the trend was always in favor of a smaller angular distance between the composite and the external contour.

4.4 DISCUSSION

We studied the responses of IT neurons to composite shapes and all of their constituent parts shown in isolation. These parts can be categorized as either natural or unnatural based on the Gestalt law of simplicity (Fig. 25). The natural parts are those whose overlapping contours were used to construct the composites in the first place, and when taken as pairs they have fewer corners compared to the unnatural shapes. Humans tend to regard the unnatural parts as invisible unless they are explicitly pointed out (Metzger, 1936).

What we found was that as a population IT neurons no more encoded the composite like the natural parts than like the unnatural parts (Fig. 27A). Instead, these overlapping shapes more closely resembled the average of any pair of complementary parts (Fig. 27B-D). One of the unnatural parts did come close to the composite in neuronal activation space: the external contour. This effect was prominent early in the trial (Fig. 29A).

4.4.1 Relation to Occlusion

Previous work using overlapping shape outlines demonstrated mean firing rate suppression to the composites compared to the individual natural parts (Missal et al., 1999). Our results did not replicate this finding. Instead, we observed that mean firing rates was the same across conditions (Fig. 28C) and on a neuron-by-neuron basis the composite response was equal to the average of the responses to the parts (Fig. 27B-D).

Our results are also at odds with a prior finding in V4 that the representation of accidental contours – those generated by occlusion of solid shapes – were suppressed (Bushnell et al., 2011). The analog of accidental contours in our study were the unnatural parts. We saw no evidence that the responses to unnatural parts were suppressed compared to natural parts or composites. However, the comparison between accidental contours created by opaque occluders and the creation of unnatural parts by overalapping shape outlines are not exactly the same, so it is unclear whether our results truly stand in opposition.

Instead it may be the case that opaque occluders more forcefully assert the percept of 3D overlap whereas line drawings are perceived as 2D intersections. However, Missal and colleagues also explored the effect of opaque occluders and found no significant differences between those conditions and overlapping shape outlines (1999), suggesting that the lack of suppression we observed was not due to our use of shape outlines. Neither these studies nor ours incorporated behavior so we can't say for sure whether the neuronal difference reflect genuine difference in mental percepts.

4.4.2 The Whole is equal to the Average of the Parts

Our result that the whole is equal to the average of the parts is in agreement with prior studies involving IT neurons (Sripati and Olson, 2010a) and visual search behavior (Pramod and Arun, 2016a). Under these previous designs, the parts were segregated to separate poles of the objects, and the local features anchoring the parts to the object never changed, so it is less surprising that IT neurons would encode the whole as a linear combination of the parts in this context.

One exception to this rule of part summation was the presence of symmetry, which increased the perceived dissimilarity between objects (Pramod and Arun, 2016a). Symmetry has long been a fixture of the Gestalt theories of vision, and it ties into the law of simplicity by the virtue that symmetric objects are simpler than their asymmetric counterparts so they attain configural superiority such that they “pop out” of visual arrays (Pomerantz et al., 1977). Another interesting parallel between this behavioral work and the present experiment is that when the targets of visual search were varied in their natural parts – in this case defined as the top and bottom halves of a wasp-waisted figure – there was a slight advantage over when the unnatural parts – defined as the parts separated by a vertical bisection – were varied across search targets (Pramod and Arun, 2016a). Because of this behavioral evidence that the Gestalt of objects affects their mental representation, it was surprising that there were no nonlinearities present in our results.

Using the kind of 2D line drawings that we employed, it has been demonstrated that humans more readily perceive natural compared to unnatural parts (Mens and Leeuwenberg, 1988). Because our results stand in opposition to this behavior, we can postulate three explanations. First, it may be the case that neuronal segregation of natural from unnatural parts may require conscious effort that our monkeys were not inspired to put forth. Second, parsing composites into natural parts may be unique tendency of humans, perhaps as a result of prior history with interpreting abstract shapes. Third, the breakdown of composites into natural versus unnatural parts may occur somewhere outside of IT cortex. While IT activity correlates with spontaneous perceptual oscillations of a bistable image (Sheinberg and Logothetis, 1997), task instructions not seem to modulate IT responses evoked by single objects (Vogels et al., 1995). The supposition that monkeys don’t perceptually parse composites into their natural parts remain plausible because no one has explicitly tested this in monkeys behaviorally. While functional imaging demonstrates activation in human LOC when displays possess the Gestalt rule of good continuation (Kuai et al., 2016), monkey IT is not a perfect analog. Therefore, we cannot conclude from our negative finding that the law of simplicity is not reflected elsewhere in the primate brain.

4.4.3 Dissociating External Contour from Internal Parts

While we did not find evidence for natural part preference, we did observe that one part stood above the rest. Of all the parts we tested, the external contour came closest to capturing the representation of the composite (Fig. 29A). This is similar to the finding that IT neurons encode image dissimilarity according to the global footprint of objects (Sripati and Olson, 2010b).

Many attempts have been made in the past to describe the high level neuronal representation of shapes in terms of the external contour of objects. So-called boundary-based models first convert an object into a silhouette, discarding internal detail. Four boundary-based models were compared against human psychophysics during a visual search task and found to be wholly inadequate to describe the perception of shape dissimilarity (Pramod and Arun, 2016b). One of these models, Fourier descriptor filters was also tested against shape representations in IT and found to be a poor fit in that system as well (Albright and Gross, 1990).

What we observed is that especially early in the trial (Fig. 29), the composite was encoded more like the external contour than any other part. Later in the trial the parts – in particular the part created from the top half of the composite – became more equally represented. We take this as evidence for the global advantage effect, in which global shape is perceived before local details (Navon, 1977), which has previously been demonstrated in IT with Navon-style hierarchical stimuli (Sripati and Olson, 2009).

4.4.4 Likelihood versus Simplicity

Finally, we return to an age-old debate. Are visual stimuli interpreted in the simplest way possible (Wertheimer, 1923), or as the most likely interpretation (Helmholtz, 1925)? Our results are more consistent with the latter theory. Each part was equally likely to appear under our experimental design. The simplicity of parts conferred no bias on neuronal responses. Composites were instead encoded as a linear combination of their constituent parts.

By contrast, prior studies have shown that IT neurons robustly encode familiarity (Anderson et al., 2008; Freedman et al., 2005; Li et al., 1993; Mruczek and Sheinberg, 2005). Therefore, a neuronal mechanism exists in inferotemporal cortex whereby the input could be interpreted based on its likelihood, which is in turn derived from past experience. This is the idea behind hierarchical Bayesian models of visual cortex (Lee and Mumford, 2003). Since our experiment did not vary the likelihood that any part would appear we cannot say that our results provide evidence for the likelihood model, only that our results do not support its rival, the simplicity model.

4.4.5 Conclusions

Ultimately what we conclude from these results is that inferotemporal cortex does not abide by the Gestalt law of simplicity when encoding two-dimensional overlapping shapes. Such composites were merely represented as the average of any pair of constituent parts, regardless of how complex or how unnatural they were. Furthermore, the whole was represented especially well by the external contour alone, particularly in the first 70ms of the visual burst. We take this as additional evidence that global form is encoded before local features and that this phenomenon occurs in inferotemporal cortex.

5.0 GENERAL DISCUSSION

The purpose of the experiments described in this dissertation was to investigate the neuronal mechanisms that contribute to object representation. The experiment presented in chapter 2 investigated whether macaque monkeys experience the phenomenon of crowding. We demonstrated that, like humans, macaque monkeys have greater trouble discriminating peripheral letters when other letters are nearby, and that the interference zone was dependent solely on eccentricity. These are the hallmarks of crowding, and they provide a new behavioral paradigm for investigating this phenomenon at the neuronal level, which has never before been done. The experiments presented in chapter 3 did just that. We investigated the effects of crowding on visual object representations in inferotemporal cortex. What we found was that crowding both quantitatively weakens the neuronal selectivity between crowded objects and qualitatively alters the neuronal code. In chapter 4 we continued the theme of investigating how parts of objects interact, this time in the context of overlapping shapes that create new parts. What we found is that rather than following Gestalt principle of simplicity, inferotemporal cortex neurons encoded composite stimuli as the average of any complete set of their constituent parts. The significance of our findings was discussed at a more technical level at the end of the previous chapters. Here, we provide a more general overview of the relevance of our findings to the field of visual neuroscience at large.

5.1 A MONKEY MODEL OF VISUAL CROWDING

Much of what we think we know about the human nervous system comes from the study of nonhuman animals. The emphasis on animal models in understandable given the limitations of experimental tools available for use in humans, particularly when it comes to understanding mechanisms at or below the level of single neurons. For many cognitive and behavioral neuroscientists, the animal of choice has been the macaque monkey (Macaca mulatta). The role of macaques as proxies for humans is especially apparent in vision research. When textbooks turn from human visual function to its neuronal basis, they switch from human to macaque studies so casually that one could easily lose sight of the fact that these species are not neurologically equivalent.

Old World monkeys, the group to which macaques belong, diverged from apes and humans approximately 25–30 million years ago (Wilkinson et al., 2010). When considering the evolutionary relationship between any pair of species, it is important to keep in mind that evolutionary changes can and do occur along both branches emanating from the most recent common ancestor. Thus, it is not a given that homologous structures and functions are conserved.

While it has been asserted that monkeys see what humans see (Kaas, 1992), when the two species were tested in side-by-side psychophysical tasks, researchers found “important, nontrivial differences between the data for monkeys and humans” (Harwerth and E. L. Smith, 1985). Whereas humans showed the greatest sensitivity to light in the red part of the spectrum and the lowest sensitivity to blue, macaques followed the opposite trend, which the researchers attribute to pre-retinal light loss due to higher macular pigment in humans (Harwerth and E. L. Smith, 1985). Humans also possess superior acuity compared to macaques, as measured by orientation discrimination (Vazquez et al., 2000), Vernier acuity (Kiorpes et al., 1993), or the peak spatial frequency of luminance contrast sensitivity (Harwerth and E. L. Smith, 1985). Harwerth and Smith chalked this acuity difference up to differences in the size of the eye. Additionally, humans had larger and more profoundly inhibitory “perceptive fields” – the regions of visual space over which an annulus interferes with the detection of a spot of light within in (Spillmann et al., 1987) – which may have implications for attention allocation (Kaas and Collins, 2003). This result is also relevant to the present study, which is concerned with crowding.

With the ultimate goal of understanding the neuronal mechanisms underlying crowding we did not want to assume that monkeys exhibit crowding or that a task suitable for monkey neurophysiology would produce crowding. What we observed was that macaque monkeys do exhibit the essential hallmarks of crowding.

At the most basic level, nearby flankers impeded the discrimination of peripheral targets, and as the spacing between targets and flankers increased performance improved. Most importantly, the critical spacing over which flankers interfered with target recognition scaled with eccentricity, independent of target/flanker size. Human behavior under identical task conditions reflected these same overall patterns, although critical spacings were on average a little smaller in humans. This finding establishes a benchmark for the extent and spread of crowding in a nonhuman primate that can be invoked for future neurophysiological investigations. Our results also support the use of the macaque monkey as a relevant model organism for future studies of crowding. Crowding researchers can now avail themselves of the powerful invasive tools not available for human studies.

5.2 CROWDING AND OBJECT REPRESENTATION

At the core of most vision research is implicitly or explicitly a hierarchical, feedforward model, in which visual processing proceeds from the analysis of basic features to more and more complex ones (Fukushima, 1980; Guclu and van Gerven, 2015; Riesenhuber and Poggio, 1999; Serre et al., 2007). Neurons in the primary visual cortex V1 act as filters to detect the lines and edges of visual images (Hubel and Wiesel, 1959). Neurons in V2 pool information from V1 neurons to encode more complex features, such as contours (Hegde and Van Essen, 2000) or texture (Freeman et al., 2013). This coding principle of filtering and pooling proceeds along the visual hierarchy to V4 and ultimately IT. The beauty and main goal of these models is to replace subjective terms, such as the Gestalt laws, by truly mechanistic theory, but there's an inherent danger that the theory will be over-simplified.

Let's consider two important characteristics of this hierarchical feedforward theoretical framework. First, if information is lost at the early stages, it is irretrievably lost, since processing at each level is fully determined by convergent inputs from the previous level. Second, receptive field size increases along the visual hierarchy because integration over progressively larger parts of the visual scene is necessary for representing faces and objects, as IT does. Although these hierarchical feedforward models don't generally account for differences between foveal and peripheral vision, we know empirically that receptive fields also get larger as a function of eccentricity (Gattass et al., 1981; 1988).

So, what does any of this have to do with crowding? Several researchers have pointed to large peripheral receptive fields as the fundamental neuronal mechanism underlying crowding (Freeman and E. P. Simoncelli, 2011; Rosenholtz, 2012). By this line of reasoning, peripheral object recognition becomes difficult when objects are embedded in clutter because irrelevant elements comingle with relevant ones when they all fall within the same receptive field. Past behavioral experiments, as well as the results presented in chapter 3, highlight the flaws of this account of crowding, and by extension, the purely hierarchical feedforward model of object recognition in general.

In particular, crowding is not an inevitable bottleneck, as was once thought (Levi, 2008). Adding more flankers (Banks et al., 1979), planning a saccadic eye movement (Harrison et al., 2013), and arranging flankers such that they exhibit grouping (Livne, 2010) can release peripheral stimuli from crowding. In our own research, we observed that information apparently lost in the feedforward sweep may not be entirely irretrievable under crowding. When attention was deployed to crowded targets, neuronal selectivity improved over time (Fig. 15, 16), which should not have occurred if target information was actually lost. Furthermore, we observed qualitative changes in the neuronal preferences for targets (Fig. 17, 18), main effects (Fig. 20A,B), and interaction effects (Fig. 20C,D) as a function of the spacing between peripheral stimuli. This finding doesn’t follow from a feedforward, hierarchical model that pools information across receptive fields. Therefore, these models fail to explain crowding and thus they cannot explain object recognition in general.

Overall, crowding offers a powerful tool for breaking the normal processes of object recognition, and in turn highlighting ways in which models of the brain come up short. By probing crowding at the level of single neurons we were able to uncover a novel neuronal behavior – swapping stimulus preferences – which future models of vision should be able to replicate.

Besides crowding, hierarchical feedforward models of vision also fail to account for the primacy of the whole – seeing the forest before the trees – as well as the laws of vision put forth by the Gestalt school of psychology. While the Gestalt system was light on mechanistic theory, the phenomena they sought to explain do not go away. In the next section, we investigate how the

Gestalt law of simplicity relates to the encoding of parts and wholes by inferotemporal cortex neurons.

5.3 THE WHOLE EQUALS THE AVERAGE OF THE PARTS

Since it was first committed to the page a century ago, researchers have attempted to operationalize the fuzzy Gestalt notion that the whole is not simply the sum of its constituent parts. The goal of operationalization was to create quantitative and falsifiable hypotheses, and ultimately to develop a theoretical framework that captures the perceptual phenomena that inspired the Gestaltists in the first place.

Perhaps the most famous attempt at translating Gestalt psychology into a quantitative model came in the form of structural information theory (SIT), which is based on Shannon's information theory (1948). Rather than quantifying information by the probability of occurrence – as Shannon did – SIT quantifies the information load of a visual stimulus by the number of parameters needed to specify its content. For instance, two parallel curves in an object, let's call them *c*, could be represented as 2*c*, whereas two different curves, would have to be represented with their own parameters, say *d* and *e*. When objects can be represented according to multiple coding schemes – such as when occlusion introduces accidental contours – SIT offers a mechanism to quantify the information load of each potential representation. Following the Gestaltist elevation of simplicity as the core principal of visual representation, proponents of SIT postulated that stimuli are perceptually organized according to the simplest – i.e., lowest information load – representation possible (Simon, 1972).

We sought to test this hypothesis by constructing composite stimuli that could be decomposed in several ways, with varying degrees of complexity. What we observed was that composite stimuli were not preferentially represented as a combination of the simplest possible set of parts (Fig. 27A). Rather, the composite was encoded as the average of any complete set of parts (Fig. 27B-D). This finding is in line with the idea that even though hidden figures tend not to rise to the level of conscious awareness they are still present in the subconscious (Mens and Leeuwenberg, 1988). Despite the lack of evidence for SIT, we did find support for another operationalization of holistic processing: primacy of the whole (Fig. 29A, 30C).

Primacy of the whole – the idea that stimuli are processed holistically before their local features are perceived – has been previously demonstrated both behaviorally (Navon, 1977) and at the level of single neurons (Sripati and Olson, 2009). What these results imply is that visual processing progresses hierarchically down a decision tree, with global form at the top and local features in the branches. This is like the predominant hierarchical, feedforward models of the visual system, except played out in reverse (Hochstein and Ahissar, 2002).

Our findings reinforced the notion of primacy of the whole in the sense that the external contour of a composite stimulus was represented first, before the other parts. However, the more precise hierarchical conceptualization of this theory found less support in our data. We constructed hierarchical trees depicting the relationships between the various parts and the composite (Fig. 28B). While the external contour continued to cluster with the composite in this analysis, and the parts formed a secondary cluster, it was the curvy part of the set – depicted as a circle in Fig. 28B – that stood atop the hierarchy, not the external contour or the composite. Yet again, this operationalization of Gestalt principles was not fully borne out in the neuronal data.

While our simple experiment with composite shapes and various decompositions is not sufficient to reject Gestalt theory outright, it does cast doubt on the notion that Gestalt laws are innately present in the visual system. It could be that holistic neuronal representations in the visual system only come with practice (Baker et al., 2002). Or perhaps monkeys lack the level of holistic perception that humans possess (Bruce, 1982), such that to the nonhuman primate brain the whole really is simply a combination of its parts (Sripati and Olson, 2010a). Since our experiment lacked behavior, we cannot make any strong claims about how our animals perceived the compound stimuli.

Ultimately, the behavioral evidence for Gestalt laws is robust and real (Elder and Zucker, 1994; Pomerantz et al., 1977; Pomerantz and Portillo, 2011; Sekuler and Palmer, 1992). Introspectively, these rules of perception seem effortless, almost inescapable (Metzger, 1936). Thus, despite the absence of mechanistic explanations, we should not reject the Gestaltists' intuitions. They protect us from falling back on the reductionist view that the representation of integrated, coherent forms can be understood by studying local processing alone.

5.4 SUMMARY AND CONCLUSIONS

The work described in this dissertation has resulted in three contributions to our understanding of the neuronal mechanisms of object recognition. The first discovery is that macaque monkeys exhibit the behavioral hallmarks of visual crowding. This is the first demonstration of this phenomenon in a non-human primate and it establishes a new experimental paradigm for future investigation of the elusive neuronal mechanisms underlying crowding. The second discovery is that crowding both quantitatively weakens and qualitatively changes the neuronal code in inferotemporal cortex. This finding rules out an explanation of crowding based solely on signal averaging. The third discovery is that neurons in inferotemporal cortex do not follow the Gestalt law of simplicity. Instead, they encode composite shapes as the average of any set of its constituent parts, not just those that appear “natural” or “simple” (i.e., possess the fewest corners). Taken together, these results provide novel insights into how the representation of object parts interfere and cohere in inferotemporal cortex.

6.0 REFERENCES

Agaoglu, M.N., Chung, S.T.L., 2016. Can (should) theories of crowding be unified? Journal of Vision 16, 10. doi:10.1167/16.15.10

Albright, T.D., Gross, C.G., 1990. Do inferior temporal cortex neurons encode shape by acting as Fourier descriptor filters?, in:. Presented at the Proceedings of the International Conference on Fuzzy Logic & Neural Networks, pp. 375–378.

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, 1199–1206. doi:10.1016/j.cub.2012.04.063

Anderson, Mruczek, R.E., Kawasaki, K., Sheinberg, D., 2008. Effects of Familiarity on Neural Activity in Monkey Inferior Temporal Lobe. Cerebral Cortex 18, 2540–2552. doi:10.1093/cercor/bhn015

Baker, C.I., Behrmann, M., Olson, C.R., 2002. Impact of learning on representation of parts and wholes in monkey inferotemporal cortex. Nature Neuroscience 5, 1210–1216. doi:10.1038/nn960

Balas, B., Nakano, L., Rosenholtz, R., 2009. A summary-statistic representation in peripheral vision explains visual crowding. Journal of Vision 9, 13–13. doi:10.1167/9.12.13

Banks, W.P., Larson, D.W., Prinzmetal, W., 1979. Asymmetry of visual interference. Percept Psychophys 25, 447–456.

Bayes, T., Price, R., Canton, J., 1763. An essay towards solving a problem in the doctrine of chances.

Bi, T., Cai, P., Zhou, T., Fang, F., 2009. The effect of crowding on orientation-selective adaptation in human early visual cortex. Journal of Vision 9, 13–13. doi:10.1167/9.11.13

Bona, S., Herbert, A., Toneatto, C., Silvanto, J., Cattaneo, Z., 2014. ScienceDirect. Cortex 51, 46–55. doi:10.1016/j.cortex.2013.11.004

Bouma, H., 1970. Interaction effects in parafoveal letter recognition. Nature 226, 177–178.

Brincat, S.L., Connor, C.E., 2004. Underlying principles of visual shape selectivity in posterior inferotemporal cortex. Nature Publishing Group 7, 880–886. doi:10.1038/nn1278

Brown, W., 1910. Some experimental results in the correlation of mental abilities. British Journal of Psychology 3, 296–322.

Bruce, C., 1982. Face recognition by monkeys: absence of an inversion effect. Neuropsychologia 20, 515–521.

Bushnell, B.N., Harding, P.J., Kosai, Y., Pasupathy, A., 2011. Partial Occlusion Modulates Contour-Based Shape Encoding in Primate Area V4. J Neurosci 31, 4012–4024. doi:10.1523/JNEUROSCI.4766-10.2011

Carandini, M., Heeger, D.J., 2011. Normalization as a canonical neural computation. Nat Rev Neurosci 13, 51–62. doi:10.1038/nrn3136

Cavanagh, P., He, S., Intriligator, J., 1999. Attentional resolution: the grain and locus of visual awareness. Neuronal basis and psychological aspects of consciousness. Singapore: World Scientific 41–52.

Chastain, G., 1982. Confusability and interference between members of parafoveal letter pairs. Percept Psychophys 32, 576–580.

Chater, N., 1996. Reconciling simplicity and likelihood principles in perceptual organization. Psychol Rev 103, 566–581.

Chelazzi, L., Duncan, J., Miller, E.K., Desimone, R., 1998. Responses of neurons in inferior temporal cortex during memory-guided visual search. J Neurophysiol 80, 2918–2940.

Chelazzi, L., Miller, E.K., Duncan, J., Desimone, R., 1993. A neural basis for visual search in inferior temporal cortex. Nature 363, 345–347. doi:10.1038/363345a0

Chen, J., He, Y., Zhu, Z., Zhou, T., Peng, Y., Zhang, X., Fang, F., 2014. Attention-Dependent Early Cortical Suppression Contributes to Crowding. J Neurosci 34, 10465–10474. doi:10.1523/JNEUROSCI.1140-14.2014

Chung, S.T.L., 2007. Learning to identify crowded letters: Does it improve reading speed? Vision Res 47, 3150–3159. doi:10.1016/j.visres.2007.08.017

Chung, S.T.L., Levi, D.M., Legge, G.E., 2001. Spatial-frequency and contrast properties of crowding. Vision Res 41, 1833–1850.

Chung, S.T.L., Li, R.W., Levi, D.M., 2007. Crowding between first- and second-order letter stimuli in normal foveal and peripheral vision. Journal of Vision 7, 10–10. doi:10.1167/7.2.10

Cowey, A., Gross, C.G., 1970. Effects of foveal prestriate and inferotemporal lesions on visual discrimination by rhesus monkeys. Exp Brain Res 11, 128–144.

Cowey, A., Rolls, E.T., 1974. Human cortical magnification factor and its relation to visual acuity. Exp Brain Res 21, 447–454.

Crowder, E.A., Olson, C.R., 2015. Macaque monkeys experience visual crowding. Journal of Vision 15, 14. doi:10.1167/15.5.14

Dakin, S.C., Bex, P.J., Cass, J.R., Watt, R.J., 2009. Dissociable effects of attention and crowding on orientation averaging. Journal of Vision 9, 28–28. doi:10.1167/9.11.28

De Weerd, P., Peralta, M.R., Desimone, R., Ungerleider, L.G., 1999. Loss of attentional stimulus selection after extrastriate cortical lesions in macaques. Nature Publishing Group 2, 753–758. doi:10.1038/11234

Desimone, R., Albright, T.D., Gross, C.G., Bruce, C., 1984. Stimulus-selective properties of inferior temporal neurons in the macaque. J Neurosci 4, 2051–2062.

Dicarlo, J.J., Zoccolan, D., Rust, N.C., 2012. How Does the Brain Solve Visual Object Recognition? Neuron 73, 415–434. doi:10.1016/j.neuron.2012.01.010

Elder, J., Zucker, S., 1994. A measure of closure. Vision Res 34, 3361–3369.

Emadi, N., Esteky, H., 2013. Neural representation of ambiguous visual objects in the inferior temporal cortex. PLoS ONE 8, e76856. doi:10.1371/journal.pone.0076856

Fang, F., He, S., 2008. Crowding alters the spatial distribution of attention modulation in human primary visual cortex. Journal of Vision 8, 6–6. doi:10.1167/8.9.6

Farah, M.J., 2004. Visual Agnosia, 2nd ed. MIT Press.

Felleman, D.J., Van Essen, D.C., 1991. Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex 1, 1–47.

Flom, M.C., Heath, G.G., Takahashi, E., 1963a. Contour interaction and visual resolution: Contralateral effects. Science 142, 979–980.

Flom, M.C., Weymouth, F.W., Kahneman, D., 1963b. Visual resolution and contour interaction. JOSA 53, 1026–1032.

Freedman, D.J., Riesenhuber, M., Poggio, T., Miller, E., 2005. Experience-Dependent Sharpening of Visual Shape Selectivity in Inferior Temporal Cortex. Cerebral Cortex 16, 1631–1644. doi:10.1093/cercor/bhj100

Freedman, D.J., Riesenhuber, M., Poggio, T., Miller, E.K., 2003. A comparison of primate prefrontal and inferior temporal cortices during visual categorization. J Neurosci 23, 5235–5246.

Freeman, J., Chakravarthi, R., Pelli, D.G., 2012. Substitution and pooling in crowding. Attention, Perception & Psychophysics 74, 379–396. doi:10.3758/s13414-011-0229-0

Freeman, J., Pelli, D.G., 2007. An escape from crowding. Journal of Vision 7, 22–22. doi:10.1167/7.2.22

Freeman, J., Simoncelli, E.P., 2011. Metamers of the ventral stream. Nature Neuroscience 14, 1195–1201. doi:10.1038/nn.2889

Freeman, J., Ziemba, C.M., Heeger, D.J., Simoncelli, E.P., Movshon, J.A., 2013. A functional and perceptual signature of the second visual area in primates. Nature Neuroscience. doi:10.1038/nn.3402

Fukushima, K., 1980. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biological Cybernetics 36, 193–202.

Gattass, R., Gross, C.G., Sandell, J.H., 1981. Visual topography of V2 in the macaque. J Comp Neurol 201, 519–539. doi:10.1002/cne.902010405

Gattass, R., Sousa, A.P., Gross, C.G., 1988. Visuotopic organization and extent of V3 and V4 of the macaque. The Journal of neuroscience 8, 1831–1845.

Gray, H., 1918. Gray's Anatomy of the Human Body, 20 ed. Lea & Febinger, Philadelphia.

Greenwood, J.A., Bex, P.J., Dakin, S.C., 2010. Crowding Changes Appearance. Current Biology 20, 496–501. doi:10.1016/j.cub.2010.01.023

Greenwood, J.A., Bex, P.J., Dakin, S.C., 2009. Positional averaging explains crowding with letter-like stimuli. Proc Natl Acad Sci USA 106, 13130–13135. doi:10.1073/pnas.0901352106

Grill-Spector, K., Kourtzi, Z., Kanwisher, N., 2001. The lateral occipital complex and its role in object recognition. Vision Res 41, 1409–1422.

Gross, C.G., Bender, D.B., Rocha-Miranda, C.E., 1969. Visual receptive fields of neurons in inferotemporal cortex of the monkey. Science 166, 1303–1306.

Gross, C.G., Rocha-Miranda, C.E., Bender, D.B., 1972. Visual properties of neurons in inferotemporal cortex of the Macaque. J Neurophysiol 35, 96–111.

Guclu, U., van Gerven, M.A.J., 2015. Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream. J Neurosci 35, 10005–10014. doi:10.1523/JNEUROSCI.5023-14.2015

Hanes, D.P., Thompson, K.G., Schall, J.D., 1995. Relationship of Presaccadic Activity in Frontal Eye Field and Supplementary Eye Field to Saccade Initiation in Macaque - Poisson Spike Train Analysis. Exp Brain Res 103, 85–96.

Hanus, D., Vul, E., 2013. Quantifying error distributions in crowding. Journal of Vision 13, 17–17. doi:10.1167/13.4.17

Harrison, W.J., Bex, P.J., 2015. A Unifying Model of Orientation Crowding in Peripheral Vision. Curr Biol 25, 3213–3219. doi:10.1016/j.cub.2015.10.052

Harrison, W.J., Mattingley, J.B., Remington, R.W., 2013. Eye Movement Targets Are Released from Visual Crowding. J Neurosci 33, 2927–2933. doi:10.1523/JNEUROSCI.4172-12.2013

Harwerth, R.S., Smith, E.L., 1985. Rhesus monkey as a model for normal vision of humans. Am J Optom Physiol Opt 62, 633–641.

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

Hegde, J., Van Essen, D.C., 2000. Selectivity for complex shapes in primate visual area V2. J Neurosci 20, RC61–RC61.

Heinen, S.J., Skavenski, A.A., 1991. Recovery of Visual Responses in Foveal V1 Neurons Following Bilateral Foveal Lesions in Adult Monkey. Exp Brain Res 83, 670–674.

Helmholtz, H., 1925. Perceptions of Vision. The Optical Society of America, Menasha, Wisconsin.

Herzog, M.H., Sayim, B., Chicherov, V., Manassi, M., 2015. Crowding, grouping, and object recognition: A matter of appearance. Journal of Vision 15, 5. doi:10.1167/15.6.5

Hochstein, S., Ahissar, M., 2002. View from the top: Hierarchies and reverse hierarchies in the visual system. Neuron.

Hubel, D.H., Wiesel, T.N., 1965. Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. J Neurophysiol 28, 229–289.

Hubel, D.H., Wiesel, T.N., 1959. Receptive fields of single neurones in the cat's striate cortex. J Physiol (Lond) 148, 574.

Intriligator, J., Cavanagh, P., 2001. The Spatial Resolution of Visual Attention. Cogn Psychol 43, 171–216. doi:10.1006/cogp.2001.0755

Ito, M., Tamura, H., Fujita, I., Tanaka, K., 1995. Size and position invariance of neuronal responses in monkey inferotemporal cortex. J Neurophysiol 73, 218–226.

James, T.W., Culham, J., Humphrey, G.K., Milner, A.D., Goodale, M.A., 2003. Ventral occipital lesions impair object recognition but not object-directed grasping: an fMRI study. Brain 126, 2463–2475. doi:10.1093/brain/awg248

Kaas, J.H., 1992. Do humans see what monkeys see? Trends in neurosciences.

Kaas, J.H., Collins, C.E., 2003. The Primate Visual System. CRC Press.

Keshvari, S., Rosenholtz, R., 2016. Pooling of continuous features provides a unifying account of crowding. Journal of Vision 16, 39. doi:10.1167/16.3.39

Kiorpes, L., Kiper, D.C., Movshon, J.A., 1993. Contrast sensitivity and vernier acuity in amblyopic monkeys. Vision Res 33, 2301–2311.

Kobatake, E., Tanaka, K., 1994. Neuronal Selectivities to Complex Object Features in the Ventral Visual Pathway of the Macaque Cerebral-Cortex. J Neurophysiol 71, 856–867.

Koerner, F., Teuber, H.L., 1973. Visual field defects after missile injuries to the geniculo-striate pathway in man. Exp Brain Res 18, 88–113.

Koffka, K., 1935. Principles of Gestalt Psychology. Lund Humphries, London.

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, 255–279.

Kosai, Y., El-Shamayleh, Y., Fyall, A.M., Pasupathy, A., 2014. The Role of Visual Area V4 in the Discrimination of Partially Occluded Shapes. J Neurosci 34, 8570–8584. doi:10.1523/JNEUROSCI.1375-14.2014

Krumhansl, C.L., Thomas, E., 1977. Effect of Level of Confusability on Reporting Letters From Briefly Presented Visual-Displays. Percept Psychophys 21, 269–279.

Kuai, S.-G., Li, W., Yu, C., Kourtzi, Z., 2016. Contour Integration over Time: Psychophysical and fMRI Evidence. Cerebral Cortex bhw147. doi:10.1093/cercor/bhw147

Lee, J., Maunsell, J.H., 2009. A normalization model of attentional modulation of single unit responses. PLoS ONE 4, e4651. doi:10.1371/journal.pone.0004651.g001

Lee, T.S., Mumford, D., 2003. Hierarchical Bayesian inference in the visual cortex. J Opt Soc Am A Opt Image Sci Vis 20, 1434–1448.

Lee, T.S., Nguyen, M., 2001. Dynamics of subjective contour formation in the early visual cortex. Proc Natl Acad Sci USA 98, 1907–1911. doi:10.1073/pnas.031579998

Leeuwenberg, E., Boselie, F., 1988. Against the Likelihood Principle in Visual Form Perception. Psychol Rev 95, 485–491.

Leopold, D.A., Logothetis, N.K., 1996. Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. Nature 379, 549–553. doi:10.1038/379549a0

Lettvin, J.Y., 1976. On seeing sidelong. The Sciences.

Levi, D.M., 2008. Crowding—An essential bottleneck for object recognition: A mini-review. Vision Res 48, 635–654. doi:10.1016/j.visres.2007.12.009

Levi, D.M., Hariharan, S., Klein, S.A., 2002a. Suppressive and facilitatory spatial interactions in peripheral vision: Peripheral crowding is neither size invariant nor simple contrast masking. Journal of Vision 2, 167–177. doi:10.1167/2.2.3

Levi, D.M., Klein, S.A., Hariharan, S., 2002b. Suppressive and facilitatory spatial interactions in foveal vision: foveal crowding is simple contrast masking. Journal of Vision 2, 140–166. doi:10:1167/2.2.2

Li, Miller, E.K., Desimone, R., 1993. The representation of stimulus familiarity in anterior inferior temporal cortex. J Neurophysiol 69, 1918–1929.

Livne, T., 2010. How do flankers' relations affect crowding? Journal of Vision 10, 1–14. doi:10.1167/10.3.1

MacKay, D., 1992. Bayesian interpolation. Neural computation.

Mareschal, I., Morgan, M.J., Solomon, J.A., 2010. Cortical distance determines whether flankers cause crowding or the tilt illusion. Journal of Vision 10, 13–13. doi:10.1167/10.8.13

Martin, A.B., Heydt, von der, R., 2015. Spike Synchrony Reveals Emergence of Proto-Objects in Visual Cortex. J Neurosci 35, 6860–6870. doi:10.1523/JNEUROSCI.3590-14.2015

McAdams, C.J., Maunsell, J., 1999. Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. The Journal of neuroscience 19, 431–441.

McMahon, D.B.T., Olson, C.R., 2007. Repetition suppression in monkey inferotemporal cortex: relation to behavioral priming. J Neurophysiol 97, 3532–3543. doi:10.1152/jn.01042.2006

Mens, L.H., Leeuwenberg, E.L., 1988. Hidden figures are ever present. J Exp Psychol Hum Percept Perform 14, 561–571.

Merigan, W.H., Nealey, T.A., Maunsell, J., 1993. Visual Effects of Lesions of Cortical Area V2 in Macaques. The Journal of neuroscience 13, 3180–3191.

Metzger, W., 1936. Laws of Seeing. MIT Press. (Original work published 1936), Cambridge, MA.

Miller, M., Pasik, P., Pasik, T., 1980. Extrageniculostriate vision in the monkey. VII. Contrast sensitivity functions. J Neurophysiol 43, 1510–1526.

Millin, R., Arman, A.C., Chung, S.T.L., Tjan, B.S., 2013. Visual Crowding in V1. Cerebral Cortex. doi:10.1093/cercor/bht159

Missal, M., Vogels, R., Li, C.Y., Orban, G.A., 1999. Shape interactions in macaque inferior temporal neurons. J Neurophysiol 82, 131–142.

Miyashita, Y., 1988. Neuronal correlate of visual associative long-term memory in the primate temporal cortex. Nature 335, 817–820. doi:10.1038/335817a0

Monosov, I.E., Sheinberg, D.L., Thompson, K.G., 2011. The Effects of Prefrontal Cortex Inactivation on Object Responses of Single Neurons in the Inferotemporal Cortex during Visual Search. J Neurosci 31, 15956–15961. doi:10.1523/JNEUROSCI.2995-11.2011

Moore, T., 2003. Microstimulation of the Frontal Eye Field and Its Effects on Covert Spatial Attention. J Neurophysiol 91, 152–162. doi:10.1152/jn.00741.2002

Moran, J., Desimone, R., 1985. Selective attention gates visual processing in the extrastriate cortex. Science 229, 782–784.

Motter, B.C., 1994. Neural Correlates of Feature Selective Memory and Pop-Out in Extrastriate Area V4. The Journal of neuroscience 14, 2190–2199.

Mruczek, R., Sheinberg, D.L., 2005. Distractor familiarity leads to more efficient visual search for complex stimuli. Percept Psychophys 67, 1016–1031.

Mruczek, R.E.B., Sheinberg, D.L., 2007. Activity of Inferior Temporal Cortical Neurons Predicts Recognition Choice Behavior and Recognition Time during Visual Search. J Neurosci 27, 2825–2836. doi:10.1523/JNEUROSCI.4102-06.2007

Nandy, A.S., Tjan, B.S., 2012. Saccade-confounded image statistics explain visual crowding. Nature Neuroscience 15, 463–469. doi:10.1038/nn.3021

Navon, D., 1977. Forest before trees: The precedence of global features in visual perception. Cogn Psychol.

Op De Beeck, H., Vogels, R., 2000. Spatial sensitivity of macaque inferior temporal neurons. J Comp Neurol 426, 505–518.

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

Pelli, D.G., 2008. Crowding: a cortical constraint on object recognition. Curr Opin Neurobiol 18, 445–451. doi:10.1016/j.conb.2008.09.008

Pelli, D.G., Palomares, M., Majaj, N.J., 2004. Crowding is unlike ordinary masking: Distinguishing feature integration from detection. Journal of Vision 4, 1136:1168. doi:10.1167/4.12.12

Pelli, D.G., Tillman, K.A., 2008. The uncrowded window of object recognition. Nature Neuroscience 1129–1135. doi:10.1038/nn.2187

Pelli, D.G., Tillman, K.A., Freeman, J., Su, M., Berger, T.D., Majaj, N.J., 2007. Crowding and eccentricity determine reading rate. Journal of Vision 7, 20–20. doi:10.1167/7.2.20

Petrov, Y., Popple, A.V., McKee, S.P., 2007. Crowding and surround suppression: Not to be confused. Journal of Vision 7, 12–12. doi:10.1167/7.2.12

Pomerantz, J.R., Kubovy, M., 1986. Theoretical approaches to perceptual organization: Simplicity and likelihood principles. Organization.

Pomerantz, J.R., Portillo, M.C., 2011. Grouping and emergent features in vision: Toward a theory of basic Gestalts. J Exp Psychol Hum Percept Perform 37, 1331–1349. doi:10.1037/a0024330

Pomerantz, J.R., Sager, L.C., Stoever, R.J., 1977. Perception of wholes and of their component parts: some configural superiority effects. J Exp Psychol Hum Percept Perform 3, 422–435.

Portilla, J., Simoncelli, E., 2000. A parametric texture model based on joint statistics of complex wavelet coefficients. Int J Comput Vision 40, 49–71.

Põder, E., 2007. Effect of colour pop-out on the recognition of letters in crowding conditions. Psychological Research 71, 641–645. doi:10.1007/s00426-006-0053-7

Põder, E., Wagemans, J., 2007. Crowding with conjunctions of simple features. Journal of Vision 7, 23.1–12. doi:10.1167/7.2.23

Pramod, R.T., Arun, S.P., 2016a. Object attributes combine additively in visual search. Journal of Vision 16, 8. doi:10.1167/16.5.8

Pramod, R.T., Arun, S.P., 2016b. Do computational models differ systematically from human object perception?, in:. Presented at the Proceedings of the IEEE Conference on ….

Ratan Murty, N.A., Arun, S.P., 2015. Dynamics of 3D view invariance in monkey inferotemporal cortex. J Neurophysiol 113, 2180–2194. doi:10.1152/jn.00810.2014

Reynolds, J.H., Heeger, D.J., 2009. The normalization model of attention. Neuron 61, 168–185.

Riesenhuber, M., Poggio, T., 1999. Hierarchical models of object recognition in cortex. Nature Neuroscience 2, 1019–1025. doi:10.1038/14819

Rolls, E.T., Aggelopoulos, N.C., Zheng, F., 2003. The receptive fields of inferior temporal cortex neurons in natural scenes. J Neurosci 23, 339–348.

Rolls, E.T., Tovee, M.J., Purcell, D.G., Stewart, A.L., Azzopardi, P., 1994. The responses of neurons in the temporal cortex of primates, and face identification and detection. Exp Brain Res 101, 473–484.

Ronconi, L., Bertoni, S., Marotti, R.B., 2016. ScienceDirect. Cortex 79, 87–98. doi:10.1016/j.cortex.2016.03.005

Rosenholtz, R., 2012. Rethinking the role of top-down attention in vision: effects attributable to a lossy representation in peripheral vision 1–15. doi:10.3389/fpsyg.2012.00013/abstract

Rosenholtz, R., Huang, J., Raj, A., Balas, B.J., Ilie, L., 2012. A summary statistic representation in peripheral vision explains visual search. Journal of Vision 12, 14–14. doi:10.1167/12.4.14

Rust, N.C., Dicarlo, J.J., 2010. Selectivity and Tolerance (“Invariance”) Both Increase as Visual Information Propagates from Cortical Area V4 to IT. J Neurosci 30, 12978–12995. doi:10.1523/JNEUROSCI.0179-10.2010

Sáry, G., Chadaide, Z., Tompa, T., Köteles, K., Kovács, G.Y., Benedek, G., 2007. Illusory shape representation in the monkey inferior temporal cortex. European Journal of Neuroscience 25, 2558–2564. doi:10.1111/j.1460-9568.2007.05494.x

Schall, J.D., Morel, A., King, D.J., Bullier, J., 1995. Topography of visual cortex connections with frontal eye field in macaque: convergence and segregation of processing streams. J Neurosci 15, 4464–4487.

Sekuler, A.B., Palmer, S.E., 1992. Perception of partly occluded objects: A microgenetic analysis. Journal of Experimental Psychology.

Serre, T., Oliva, A., Poggio, T., 2007. A feedforward architecture accounts for rapid categorization. Proc Natl Acad Sci USA 104, 6424–6429. doi:10.1073/pnas.0700622104

Shannon, C.E., 1948. A mathematical theory of communication. Bell System Technical Journal.

Sheinberg, D.L., Logothetis, N.K., 2001. Noticing familiar objects in real world scenes: the role of temporal cortical neurons in natural vision. J Neurosci 21, 1340–1350.

Sheinberg, D.L., Logothetis, N.K., 1997. The role of temporal cortical areas in perceptual organization. Proc Natl Acad Sci USA 94, 3408–3413.

Simon, H.A., 1972. Complexity and the representation of patterned sequences of symbols. Psychol Rev.

Smith, M.A., Bair, W., Movshon, J.A., 2006. Dynamics of suppression in macaque primary visual cortex. J Neurosci 26, 4826–4834. doi:10.1523/JNEUROSCI.5542-06.2006

Spearman, C., 1910. Correlation calculated from faulty data. British Journal of Psychology 3, 271–295. doi:10.1111/j.2044-8295.1910.tb00206.x

Spillmann, L., Ransom-Hogg, Oehler, R., 1987. A Comparison of Perceptive and Receptive-Fields in Man and Monkey. Hum Neurobiol 6, 51–62.

Sripati, A.P., Olson, C.R., 2010a. Responses to compound objects in monkey inferotemporal cortex: the whole is equal to the sum of the discrete parts. J Neurosci 30, 7948–7960. doi:10.1523/JNEUROSCI.0016-10.2010

Sripati, A.P., Olson, C.R., 2010b. Global Image Dissimilarity in Macaque Inferotemporal Cortex

Predicts Human Visual Search Efficiency. The Journal of neuroscience 30, 1258–1269. doi:10.1523/JNEUROSCI.1908-09.2010

Sripati, A.P., Olson, C.R., 2009. Representing the forest before the trees: a global advantage effect in monkey inferotemporal cortex. J Neurosci 29, 7788–7796. doi:10.1523/JNEUROSCI.5766-08.2009

Strappini, F., Pelli, D.G., Di Pace, E., Martelli, M., 2017. Agnosic vision is like peripheral vision, which is limited by crowding. Cortex. doi:10.1016/j.cortex.2017.01.012

Strasburger, H., 2005. Unfocussed spatial attention underlies the crowding effect in indirect form vision. Journal of Vision 5, 8–8. doi:10.1167/5.11.8

Strasburger, H., Harvey, L.O., Rentschler, I., 1991. Contrast thresholds for identification of numeric characters in direct and eccentric view. Percept Psychophys 49, 495–508.

Strasburger, H., Malania, M., 2013. Source confusion is a major cause of crowding. Journal of Vision 13, 24–24. doi:10.1167/13.1.24

Tanaka, K., Saito, H., Fukada, Y., Moriya, M., 1991. Coding visual images of objects in the inferotemporal cortex of the macaque monkey. J Neurophysiol 66, 170–189.

Toet, A., Levi, D.M., 1992. The two-dimensional shape of spatial interaction zones in the parafovea. Vision Res 32, 1349–1357.

Tripathy, S.P., Cavanagh, P., 2002. The extent of crowding in peripheral vision does not scale with target size. Vision Res 42, 2357–2369.

Tsao, D.Y., Freiwald, W.A., Knutsen, T.A., Mandeville, J.B., Tootell, R.B.H., 2003. Faces and objects in macaque cerebral cortex. Nature Publishing Group 6, 989–995. doi:10.1038/nn1111

Ungerleider, L.G., Mishkin, M., 1982. Two cortical visual systems. MIT Press, Cambridge.

van den Berg, R., Roerdink, J.B.T.M., Cornelissen, F.W., 2010. A Neurophysiologically Plausible Population Code Model for Feature Integration Explains Visual Crowding. PLoS Comput Biol 6, e1000646. doi:10.1371/journal.pcbi.1000646.s005

van den Berg, R., Roerdink, J.B.T.M., Cornelissen, F.W., 2007. On the generality of crowding: Visual crowding in size, saturation, and hue compared to orientation. Journal of Vision 7, 14–14. doi:10.1167/7.2.14

Vazquez, P., Cano, M., Acuna, C., 2000. Discrimination of line orientation in humans and monkeys. J Neurophysiol 83, 2639–2648.

Vogels, R., 1999. Effect of image scrambling on inferior temporal cortical responses. Neuroreport 10, 1811–1816.

Vogels, R., Sáry, G., Orban, G.A., 1995. How task-related are the responses of inferior temporal neurons? Vis Neurosci 12, 207–214.

Wagemans, J., Elder, J.H., Kubovy, M., Palmer, S.E., Peterson, M.A., Singh, M., Heydt, von der, R., 2012. A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. Psychological Bulletin 138, 1172–1217. doi:10.1037/a0029333

Wannig, A., Stanisor, L., Roelfsema, P.R., 2011. Automatic spread of attentional response modulation along Gestalt criteria in primary visual cortex. Nature Neuroscience 14, 1243–1244. doi:10.1038/nn.2910

Wapner, W., Judd, T., Gardner, H., 1978. Visual agnosia in an artist. Cortex 14, 343–364.

Wässle, H., Grünert, U., Röhrenbeck, J., Boycott, B.B., 1989. Cortical magnification factor and the ganglion cell density of the primate retina. Nature 341, 643–646. doi:10.1038/341643a0

Wertheimer, M., 1923. Laws of organization in perceptual forms. Routledge & Kegan Paul.

Whitney, D., Levi, D.M., 2011. Visual crowding: a fundamental limit on conscious perception and object recognition. Trends Cogn Sci (Regul Ed) 15, 160–168. doi:10.1016/j.tics.2011.02.005

Wilkinson, R.D., Steiper, M.E., Soligo, C., Martin, R.D., Yang, Z., Tavare, S., 2010. Dating Primate Divergences through an Integrated Analysis of Palaeontological and Molecular Data. Systematic Biology 60, 16–31. doi:10.1093/sysbio/syq054

Wolford, G., 1975. Perturbation Model for Letter Identification. Psychol Rev 82, 184–199.

Wolford, G., Shum, K.H., 1980. Evidence for feature perturbations. Percept Psychophys 27, 409–420.

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

Zaretskaya, N., Anstis, S., Bartels, A., 2013. Parietal Cortex Mediates Conscious Perception of Illusory Gestalt. J Neurosci 33, 523–531. doi:10.1523/JNEUROSCI.2905-12.2013

Zhang, Y., Meyers, E.M., Bichot, N.P., Serre, T., Poggio, T.A., Desimone, R., 2011. Object decoding with attention in inferior temporal cortex. Proc Natl Acad Sci USA 8850–8855. doi:10.1073/pnas.1100999108

Zhou, H., Friedman, H.S., Heydt, von der, R., 2000. Coding of border ownership in monkey visual cortex. The Journal of neuroscience 20, 6594–6611.

Ziemba, C.M., Freeman, J., Movshon, J.A., Simoncelli, E.P., 2016. Selectivity and tolerance for visual texture in macaque V2. Proceedings of the National Academy of Sciences 113, E3140–E3149. doi:10.1073/pnas.1510847113

Zoccolan, D., Cox, D.D., Dicarlo, J.J., 2005. Multiple object response normalization in monkey inferotemporal cortex. J Neurosci 25, 8150–8164. doi:10.1523/JNEUROSCI.2058-05.2005