Theses and Dissertations

Spring 2015

Summary statistics in vision

Mouna Attarha
University of Iowa

SUMMARY STATISTICS IN VISION

by

Mouna Attarha

A thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Psychology
in the Graduate College of
The University of Iowa

May 2015

Thesis Supervisor: Professor Cathleen M. Moore

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

_____

PH.D. THESIS

_____

This is to certify that the Ph.D. thesis of

Mouna Attarha

has been approved by the Examining Committee
for the thesis requirement for the Doctor of Philosophy
degree in Psychology at the May 2015 graduation.

Thesis Committee: _____
                         Cathleen M. Moore, Thesis Supervisor

                         _____
                         Andrew Hollingworth

                         _____
                         Bob McMurray

                         _____
                         Jonathan T. Mordkoff

                         _____
                         Shaun P. Vecera

Maman va baba, *waves*, daram miam khoone!

ACKNOWLEDGMENTS

Six years ago, I was a different person. I will confess that at the time I thought that belief and intuition equaled knowledge and fact (…or at least that my intuitions were factual even if others' were not). I thought that science revealed absolute truth. I thought that experts were infallible. And, I thought that understanding the emergence of consciousness couldn't possibly be *that* hard.

Graduate school has made my worldview less black and white, and more uncertain, but it has also gifted me with a set of skills for which I am truly grateful. I have become a more critical thinker with a greater sense of curiosity and a deeper appreciation for unsolved problems.

The person who has given me a new lens with which to see the world is my advisor, Dr. Cathleen Moore. I thank her for training me all these years and for being kind, fair, and fun as she did so.

I also thank my other wonderful committee members, Drs. Toby Mordkoff, Shaun Vecera, Andrew Hollingworth, and Bob McMurray, who trained me through their roles as instructors and meeting members. I thank them for inspiring me on a regular basis and for providing indispensible feedback regarding my research, both past and present.

ABSTRACT

The complexity and quantity of information present in the natural world far exceeds the processing abilities of our perceptual and cognitive systems, and yet our perception of the world seems complete. This suggests that the system is equipped with a variety of mechanisms that abstract the large amount of available information in a manner that preserves behaviorally relevant data and minimizes computational load. One such collection of mechanisms are called summary statistics, which refer to the set of processes that generate representations based on the statistical regularities that are often shared among groups of similar items.

Summary statistics are proposed to serve a foundational role in early visual processing as well as in later visual awareness. Many phenomena from crowding to visual search to gist perception are thought to derive from summary representations that abstract a large amount of visual information early within the stream of perceptual processing in a way that avoids all limited-capacity bottlenecks.

This dissertation challenges the view that summary representations hold the key to understanding how we establish a subjectively rich impression of the surrounding world. I demonstrate that summaries cannot complement the limited capacity aspects of our perceptual systems because forming these representations across multiple, disparate areas of the visual field undergo significant interference. I also show that these representations cannot be effortlessly established in unattended areas of the scene because their formation requires attention. These findings indicate that the proposed function of summary representations has been overstated, raising the possibility that they need not be considered a primary component in theories of visual perception after all.

iv

PUBLIC ABSTRACT

It is said that our visual experience is a 'Grand Illusion'. Our brains can only process a fraction of the total information available in the natural world, and yet our subjective impression of that world appears richly detailed and complete. The apparent disparity between our conscious experience of the visual landscape and the precision of our internal representation has suggested to some that our brains are equipped with specialized mechanisms that surmount the inherent limitations of our perceptual and cognitive systems. One proposed set of mechanisms, called summary statistics, processes information in a scene by representing the regularities that are often shared among groups of similar in terms of descriptive statistics. For example, snowflakes blowing in the wind may be represented in terms of their mean direction and speed.

Prevailing views hold that summary statistics may underlie all aspects of our subjective visual experience, inasmuch as such representations are thought to form automatically across multiple visual fields, exhaustively summarizing all available visual features regardless of attention. We challenge this view by showing that summary statistics are mediated by limited-capacity processes and therefore cannot unfold independently across multiple areas of the visual field. We also show that summary statistics require attention and thus cannot account for our sense of visual completeness outside attended visual space. In light of this evidence, we suggest that the application of summary representations to daily perceptual life has been overstated for the past decade. Indeed, many observations interpreted in terms of summary statistics can be accounted for by alternative cognitive processes, such as visual working memory.

v

TABLE OF CONTENTS

CHAPTER 5: ATTENTIONAL DEMANDS OF SUMMARY REPRESENTATIONS 94

CHAPTER 6: GENERAL DISCUSSION 116

REFERENCES 128

LIST OF FIGURES

ix

CHAPTER 1: INTRODUCTION

1.1 The 'Grand Illusion'

Appearing to form the foundation of our richest life experiences and our most enduring memories is the compelling impression that we are surrounded by a visually complete world. It was once thought that this impression arose from a detailed internal representation that rivaled the amount of detail present in the external environment (e.g., Feldman, 1985; Trehub, 1991). Consistent with this view are experiments on iconic memory demonstrating that a large quantity of information can be perceptually available in a picture-like format for a short period of time (Sperling, 1960). Furthermore, studies on scene perception have shown that observers can perceive and recognize literally hundreds of images even after only a single viewing (Shepard, 1967). These findings support the possibility that there exists an internal representation with high-resolution and large capacity that we use to see the world around us.

Although the idea of a photograph-like internal representation seemingly agrees with our phenomenological experience, it is inconsistent with much of the experimental evidence to date. It appears, instead, that the immense detail comprising the environment far exceeds the processing abilities of our sensory, attentional, and cognitive systems (e.g., Kanwisher, 1987; Luck & Vogel, 1997; Mack & Rock, 1998; Nakayama, 1990; Raymond, Shapiro, & Arnell, 1992; Rensink, O'Regan, & Clark, 1997; Simons & Chabris, 1999). Serving as one example are reports that large and remarkably salient changes to a scene oftentimes go unnoticed; observers fail to notice changes to the

identity of their conversation partner during real-world interactions (Simons & Levin, 1998). The logic is that if the system retained precise visual representations of the world, then such failures of processing should never occur. As a consequence of this logic, the view that visual representations were wholly complete required radical revision, leading scientists to develop new frameworks with which to understand perception. One such framework that developed is the view that visual representations are fairly sparse in detail (e.g., Rensink 2000).

An important question that emerges from the claim that internal representations contain few details, then, is how do we construct a representation of the world that seemingly allows us to see "everything" when it is in fact incomplete? How do these representations in turn accommodate behaviors of ecological relevance in daily life, such as finding an important email in a flooded inbox, a favorite coffee mug in a full dishwasher, or a student who skipped class on exam day? How we see and act with success rather than failure, despite processing only a fragment of the available information in the world, are the questions that O'Regan (1992) dubbed the "real mysteries" of visual perception.

One way to reconcile the difference between a sparse internal representation and a vivid subjective experience is to suppose that the brain uses specialized mechanisms that compensate for the inherent limitations of our perceptual and cognitive systems. The output of these mechanisms may ultimately produce our experience of the so-called 'Grand Illusion', or the phenomenological impression of seeing far more information in the visual world than in fact can be processed by the brain (e.g., Noë, 2002; Noë, Pessoa, & Thompson, 2000). The contributions of such mechanisms may fully explain how we

2

develop a sense of continuity and completeness of our surroundings from an impoverished representation of those very surroundings.

1.2 Statistical summary representations

One of the ways in which the human perceptual system may economize the vast amount of information in the natural world is by computing abstract representations of statistical regularities present in the scene, such as the average direction, size, and color of similar groups of items (*statistical summary representations*; SSRs; Ariely, 2001). For example, forming a surface-area ratio of brown-to-yellow for bananas may facilitate choosing the bunch with the fewest bruises more quickly than inspecting each banana in isolation, forming a summary of mean berry size across multiple bushes may lead to finding the most fruitful bush, and forming a representation of the average emotion of students in a classroom may inform an instructor if he should speak more slowly. These summary representations may therefore allow for meaningful interactions between organism and environment by providing a nexus between sensory inputs and behavior.

The study of SSRs began with evidence that humans can form a summary of mean size for a large set of different-sized circles. Observers were asked to compare the perceived mean size of the set of circles to the diameter of a subsequently presented probe circle (Ariely, 2001). The set never included a circle that matched the mean size of the set exactly, yet observers could report whether the size of the probe was smaller or larger than the mean size of the group for diameter differences as small as 4-6%. Critically, when observers were asked to report which of two probe circles had been a

3

member of the set, performance plummeted to chance levels. These results indicate that observers were able to compute the mean size of a set of stimuli quite accurately, even when they failed to either identify or remember the sizes of individual stimuli from the set (see also Brady & Alvarez, 2011; Corbett & Oriet, 2011; Demeyere, Rzeskiewicz, Humphreys, & Humphreys, 2008).

Poor performance for member identity is consistent with the view that there exists a tradeoff between SSRs and the representation of individual stimuli; precise information is lost in favor of a more holistic, coarse summary of the environment (Alvarez, 2011). This might occur in a variety of different specific ways. The identity of individual members might be discarded after a global percept of the group is formed (Ariely, 2001). Alternatively, individual representations might be so noisy that they cannot be used reliably for later identification (Alvarez & Oliva, 2008). Regardless, the inability to remember the properties of individual items within a set is one property that differentiates SSRs from object perception (e.g., Alvarez, 2011; Corbett & Oriet, 2011; Im & Halberda, 2013; Jacoby et al., 2013; Joo et al., 2009).

1.3   Proposed function of summary representations in visual perception

It has been proposed that SSRs guide behavior by reducing processing load, especially within unattended areas of the periphery (e.g., Alvarez, 2011; Alvarez & Oliva, 2008; Joo et al., 2009; Chong & Treisman, 2005a). A system that relies on coarse representations of ensembles will be more efficient than a system that relies on individual items. The idea is that the rich visual experience of the world we experience is produced

by the integration of these two general classes of representations: a representation high in detail produced by sampling individual items at fixation, and a representation low in detail produced by sampling redundant characteristics across many items in unattended and peripheral regions of the visual field (e.g., Chong & Treisman, 2003; Haberman & Whitney, 2009; Whitney, Haberman, & Sweeny, 2014). These broad classes of representations provide a complementary analysis of the external environment; while foveal representations sacrifice generality for more specific analysis, summary representations sacrifice specifics for generality (e.g., Corbett & Oriet, 2011). Because summary statistics allow the system to remain sensitive to behaviorally-relevant events that appear outside areas of focus, it is hypothesized that the function of these statistical representations is to reduce the complexity of information in the environment in a way that optimizes processing for our limited perceptual and cognitive systems (e.g., Alvarez, 2011; Alvarez & Oliva, 2009). Understanding how SSRs are computed then is important for understanding visual perception more generally.

If SSRs play this fundamental role in vision, then it follows that there should be substantial generality in the types of features and object properties that can be summarized. Consistent with this, accurate summaries are found to occur over space and time for both low-level stimuli and more complex objects, including mean brightness (Bauer, 2009), motion speed and direction (e.g., Watamaniuk, Sekular, & Williams, 1989), spatial position (e.g., Alvarez & Oliva, 2008), orientation (e.g., Dakin, 2001), height (Fouriezos, Rubenfeld, & Capstick, 2008), size over space (Ariely, 2001), size over time (Albrecht & Scholl, 2010), length (Weiss & Anderson, 1969), color (Demeyere et al., 2008), inclination (Miller & Sheldon, 1969), biological motion (Sweeny, Haroz, &

Whitney, 2013), facial identity (e.g., de Fockert & Wolfenstein, 2009), facial attractiveness (Walker & Vul, 2014), and facial emotion and gender (e.g., Haberman & Whitney, 2007). Thus, it is clear that SSRs can be formed for a wide range of visual attributes, consistent with the suggestion that establishing SSRs is a fundamental early step in visual processing. In addition, although SSRs of the mean have received the most attention, summaries of other measures of central tendency and dispersion, such as the variance, range, skew, and kurtosis are also possible (Morgan et al., 2008; Peterson & Beach, 1967; Pollard, 1984). Finally, the ability to summarize statistical regularities is not a unique property of the vision system. SSRs are generated in other sensory modalities as well (Albrecht, Scholl, & Chun, 2012; Piazza, Sweeny, Wessel, Silver, & Whitney, 2013) suggesting that these representations may be the output of a general mechanism that pools redundancies over all available sensory information (Alvarez & Oliva, 2008; Alvarez, 2011).

Within the last several years, SSRs have been proposed to underlie a wide range of phenomena and daily tasks. A few examples include peripheral recognition, texture segmentation, perceptual stability, crowding, spatial vision, visual illusions, visual search, change blindness, visual working memory, and gist perception (e.g., Ariely, 2001; Ackerman & Landy, 2014; Balas, Nakano, & Rosenholtz, 2010; Brady & Alvarez, 2011; Cavanagh, 2001; Chong et al., 2008; Corbett & Melcher, 2014; Gillen & Heath, 2014; Rosenholtz, 2011; Whitney, 2009; Whitney, Haberman, & Sweeny, 2014). For example, consider the task of visual search, a necessary skill for efficient interaction with the environment. In this task, one finds a target item (e.g. coffee mug) among distractors (e.g. surrounding kitchen items). Rosenholtz and colleagues propose that under some

6

conditions, a model that predicts search performance based on statistical summary representations of groups of items can be more successful in explaining a host of effects cited in the visual search literature (such as target-distractor discriminability, search asymmetries, and pop-out) than models that predict performance based on individual items alone (e.g., Treisman & Gelade, 1980, Treisman & Souther, 1985; Wolfe, 1994).

In addition, the finding that summaries are represented implicitly has led researchers to conclude that SSRs drive our impression of a complete world despite limited awareness (Cavanaugh, 2001; Chong & Treisman, 2003; Haberman & Whitney, 2009). The idea is that the Grand Illusion (e.g., Noë, 2002; Noë, Pessoa, & Thompson, 2000) may simply be our experience of a coarse representation of feature averages that are established early within the stream of perceptual processing (e.g., Whitney, Haberman, & Sweeny, 2014). This claim implies that these representations play a critical role from early vision to visual awareness (Haberman & Whitney, 2011; Noë, Pessoa, & Thomson, 2000; Whitney, Haberman, & Sweeny, 2014).

To summarize, SSRs are thought to play a key role in abstracting a large amount of visual information in a way that leads to rapid visual scene perception and the subjective impression that we see more than we do (e.g., Whitney, 2009; Rosenholtz, 2011). If true, then understanding SSRs is of considerable importance for theories of visual perception because these representations are necessary for both early vision and visual awareness (e.g., Corbett & Song, 2014; Haberman & Whitney, 2011; Whitney, Haberman, & Sweeny, 2014).

1.4 Motivation and outline of dissertation

Central to the proposed function of summary representations are two key claims. The first claim is that summary representations are capacity free. The second claim is that summaries are established without attention. In the chapters that follow, I reexamine these two widely held beliefs and argue that the evidence supporting these claims is flawed either in terms of methodology or experimental design.

In chapters 2, 3, and 4, I test whether summary representations are truly capacity free. To briefly introduce the issue, a statement that pervades the literature is that SSRs bypass the limited capacity bottleneck (Chong & Treisman, 2005a, p. 899; see also Alvarez, 2011; Alvarez & Oliva, 2008; Ariely, 2001; Brand et al., 2012; Chong & Treisman, 2003, 2005b; Demeyere, Rzeskiewicz, Humphreys, & Humphreys, 2008; Jacoby et al., 2013; Marchant et al., 2013; Oriet & Brand, 2013; Robitaille & Harris, 2011; Whiting & Oriet, 2011). This conclusion is based on studies that compare performance across conditions that vary the number of to-be-summarized items. The idea is that SSRs engage parallel processes to the extent that performance is equal when 4 vs 16 items are summarized. The absence of set size effects in these tasks, however, is equivocal with regard to the issue of processing independence (e.g., Huang & Pashler, 2005; Pashler, 1998), especially given the way in which set size was manipulated in these tasks. I use an extended version of the simultaneous-sequential method to test whether the two most popular SSRs reported in the literature -- mean size and mean orientation -- engage only unlimited capacity processes. Unlike previous methodologies testing this

question, the simultaneous-sequential method makes specific predictions about processing capacity (see Scharff et al., 2011a).

In addition, I discuss the various ways in which one can talk about capacity limitations. The stimuli used in Chapters 2-4 can be defined in terms of the number of available sets, number of items, or number of feature dimensions. In Experiment 1 of these chapters, I demonstrate that capacity is limited with respect to the number of sets that must be established at any given time. The last experiment in these three chapters describes capacity for the number of items that are established. The last experiment of Chapter 4 also reports capacity with regard to the number of feature dimensions.

In chapter 5, I reexamine the claim that summary representations are formed without attention. The current dominant view in the SSR literature is that summaries occur beyond the focus of selective attention (i.e., preattentively). For example, a central grating crowded out of awareness can be integrated with surrounding orientation information even though subjects cannot consciously individuate the central patch (Parkes, Lund, Angelucci, Solomon, & Morgan, 2001). This has suggested to some that averaging is "compulsory" and "mandatory". Similarly, features from an irrelevant set of items can bias reported averages for the relevant set, even to a detriment in performance (Oriet & Brand, 2012). In addition, the centroid position of multiple unattended stimuli can be identified above chance levels (Alvarez & Oliva, 2009). Although these studies suggest that averages of unattended information may be formed when attention to allocated elsewhere, a critical concern with these studies is they ask subjects to report properties of the unattended summary information throughout the experiment. Chapter 5 uses a correlated flankers task to assess the attentional demands of summary extraction.

The key difference between chapter 5 and previous studies is that statistical extraction in measured via conditioned responses. Attention therefore does not need to be directed to the unattended summary stimuli in order to perform the task.

To maintain the view that SSRs are fundamental for early visual processing and later visual awareness, it is necessary to show that these representations are computed over many items in the visual field without limitations in processing capacity. By way of preview, the results of chapters 2-4 suggest that the visual system cannot form summary representations across the entire visual field without engaging limited-capacity processes. These chapters do, however, demonstrate that multiple summary representations can be generated in one area of the environment, insofar as they are established between different feature dimensions. The results from chapter 5 suggest that these single-location summaries require attention or else they fail to be represented. Based on these findings, I challenge the widespread claim that summary representations provide a sense of visual continuity across unattended and peripheral regions of the visual field. It seems instead that the purported functional significance of summary representations has been overstated for over a decade.

CHAPTER 2: CAPACITY OF SIZE SUMMARY REPRESENTATIONS

2.1   Overview

We assessed the processing capacity of establishing statistical summary representations of mean size in visual displays using the simultaneous-sequential method. Four clusters of stimuli, each composed of several circles with various diameters, were presented around fixation. Observers searched for the cluster with the largest or smallest mean size. In the simultaneous condition, all four clusters were presented concurrently; in the sequential condition, the clusters appeared two at a time. We found that the processing capacity of SSRs for multiple sets was as extreme as a fixed-rate bottleneck process (Experiment 1). A control experiment confirmed that this was not caused by having to compare the results of multiple averaging processes (Experiment 2). In contrast to computing SSRs across ensembles, computing SSRs for items within a single ensemble using the same stimuli was consistent with unlimited-capacity processing (Experiment 3). Contrary to existing claims, summary representations appear to be extracted independently for items within single ensembles but not multiple ensembles. A developing understanding of capacity limitations in perceptual processing is discussed.

The experiments described in this chapter are published under the following citation: Attarha, M., Moore, C.M., & Vecera, S.P. (2014). Summary statistics of size: Fixed processing capacity for multiple ensembles but unlimited processing capacity for single ensembles. *Journal of Experimental Psychology: Human Perception and Performance, 40*(4), 1440-1449. DOI: http://dx.doi.org/10.1037/a0036206

2.2 Introduction

A focus in the literature has been that SSRs are established in parallel across the visual field. Consistent with these claims, Ariely (2001) found that discrimination thresholds of mean size were unaffected by the number of items (4 or 16) in the to-be-averaged set, and Chong and Treisman found that mean-size estimates for a group of heterogeneously sized circles were as accurate as those for single circles. Finally in a later study, Chong & Treisman (2005) showed that observers were able to report the mean of one of two interspersed sets of circles that were defined by color, and that performance was unaffected by whether the color of the to-be-reported subset was pre-cued or post-cued relative to the display. Observers in this study were also able to report the mean size of one of two colored subsets of stimuli as well as they could for individually presented sets, but Brand et al., (2012) were unable to replicate this aspect of the results. Together these findings have lead researchers to conclude that SSRs for multiple sets are established through processes that "…precede the limited capacity bottleneck" (Chong & Treisman, 2005), and by implication are established through unlimited-capacity processes (e.g., Oriet & Brand, 2013; Robitaille & Harris, 2011).

The goal of the current study was to test the hypothesis that SSRs are established through unlimited-capacity processes. Unlimited-capacity models state that processing occurs independently (i.e., without interference) across stimuli. These models therefore predict that performance will not vary with the number of stimuli that must be processed simultaneously. In contrast, limited-capacity models state that the processing of one stimulus is compromised by having to process other stimuli simultaneously. These

12

models therefore predict that performance will decline with increasing numbers of simultaneous stimuli.

Although the absence of set size effects in the studies reviewed above is consistent with an unlimited-capacity model of SSRs (e.g., Chong & Treisman, 2003), the evidence is equivocal with regard to the issue of interference because of the way in which set size was manipulated. Specifically, in order to maintain given average sizes, Ariely (2001) varied set size between 4 and 16 items by varying the frequency of only four distinct circle sizes. Observers therefore did not have to sample all of the stimuli in a set to do the task. They could instead sample from only a portion (e.g., an average of 4 items), effectively nullifying the set-size manipulation. When size regularity across items was minimized, forcing observers to sample from the whole set, significant set size effects were observed (Marchant et al., 2013; Myczek & Simons, 2008; but cf. Ariely, 2008; Robitaille & Harris, 2011).

Based on the large set size effects found in Marchant et al. (2013), it is unclear whether statistical extraction occurs with or without interference across stimuli. This is because set size manipulations generally simultaneously vary aspects of the task other than the number of to-be-processed stimuli, such as statistical decision noise, eye movements, exposure duration, and the ratio of relevant to irrelevant stimuli (Eckstein et al., 2000; Palmer, 1994; Shaw, 1980; Townsend, 1990). In the case of statistical decision noise, for example, set size confounds the number stimuli that must be processed with the number of perceptual representations that contribute to the task decision. Because every representation is associated with a certain amount of noise, a greater number of representations implies a greater amount of noise that is fed into decision processes.

13

Poorer performance with larger set sizes could reflect this difference alone. One strategy for handling this confound has been to develop specific models of the task in question and use them to make quantitative predictions regarding how large of an effect the increased noise should have on performance. These predictions can then be compared to the observed effect of set size on performance, which will either be more or less the same as that predicted by increased decision noise alone or not (e.g., Palmer, 1994; Shaw, 1984). Though effective within specific contexts, this strategy is limited in that it is dependent on the development of specific processing models that require, and these models require specific, and often ancillary to the question of interest, assumptions about how processing unfolds. It is for this and similar reasons that set size effects are not ideal for assessing the issue of processing independence (e.g., Huang & Pashler, 2005; Pashler, 1998; Wolfe, 1998). We turn to the simultaneous–sequential method instead.

2.3.1 Simultaneous–sequential method

The simultaneous–sequential method was developed to test the capacity limitations of perceptual processing in a way that avoids many of the problems associated with set size manipulations (Eriksen & Spencer, 1969; Shiffrin & Gardner, 1972). The overall number of to-be-processed stimuli remains constant in this method. Because of this fixed overall set size, decision factors and most sensory factors also remain constant and therefore cannot drive any observed differences in performance that occur. The factor that is varied in the simultaneous–sequential method is how many stimuli must be processed at any given time. In the simultaneous condition, all stimuli onset concurrently

14

in a single frame and must be processed at the same time to perform the task. In contrast, the sequential condition presents half of the same display across two temporal frames, and therefore fewer stimuli require processing at any given time. Importantly every display is presented for the same amount of time in the simultaneous and sequential conditions (see *Figure 2.1*). Furthermore, the quick exposure duration of the critical displays serve to minimize eye movements and sequential shifts of attention. A direct comparison of accuracy performance between the simultaneous and sequential conditions, therefore, can then be made because the amount of time available for processing each item is constant between conditions and because the duration is fast enough to limit performance.

The simultaneous–sequential method tests the (in)dependence of processing multiple relevant stimuli. Unlimited-capacity models predict equal accuracy across the simultaneous and sequential conditions. This follows because if processing unfolds completely independently across multiple stimuli, then it should make no difference how many stimuli require processing. The quality or speed of processing will be constant. In contrast, limited-capacity models predict an advantage in accuracy for sequential over simultaneous presentation because the sequential condition allows fewer stimuli to engage the process at any one time. Processing is compromised by having to process additional items at the same time. Scharff et al. (2011a) has formulized these predictions.

In the current study, observers viewed four clusters of circles and reported whether the mean size of one of the clusters was larger or smaller than the others (*Figure 2.1*). The four clusters were presented all at once in the simultaneous condition (*Figure 2.1A*) or in subsets of two in the sequential condition (*Figure 2.1B*). If SSRs unfold

15

independently across groups of stimuli (i.e., they are established through unlimited-capacity processes), then performance should be just as good in the simultaneous condition as in the sequential condition. In contrast, if computing SSRs interfere with each other across clusters (i.e., they require limited-capacity processes), then performance should be better in the sequential condition than in the simultaneous condition because fewer items require processing at any given time.

Finally following Scharff et al. (2011a), we included a repeated condition (*Figure 2.1C*). This was just like the simultaneous condition, except that it presents the entire array of items twice across two temporal frames. Assuming there is room for improvement over what can be processed during the single simultaneous display, performance should be better in the repeated condition when each item is available for twice the duration. This provided two advantages over the basic design. First, it allowed us to confirm that if SSRs do involve limited-capacity processes, our conditions were such that observers could have taken advantage of the sequential condition. If, for example, the stimulus duration that we used was too long, SSRs for all clusters could be established by shifting processing within that one display period, then performance might be equal across the simultaneous and sequential conditions, despite SSRs depending on limited-capacity processes. If performance is equal across the simultaneous and sequential conditions but better in the repeated condition, however, then we can be assured that this was not the case. Another advantage of including the repeated condition is that in the event that processing is limited capacity, it allows one to assess a particular limited-capacity model -- *fixed capacity* -- which states that processing is limited to a fixed amount of information per unit time. A serial model (i.e., one cluster at a time) is a

specific example of a fixed-capacity model. A fixed-capacity model predicts not only that performance will be higher in the sequential condition than in the simultaneous condition, but also that it will be as good as performance in the repeated condition. Formal details of these predictions are given in Scharff et al., 2011a.

As a preview of our results, we found that computing SSRs for multiple ensembles of stimuli was inconsistent with unlimited-capacity processing, and consistent with fixed-capacity processing (Experiment 1). A control experiment confirmed that this was not caused by having to compare the results of multiple averaging processes (Experiment 2). In contrast to computing SSRs across ensembles, computing SSRs for a single ensemble was consistent with unlimited-capacity processing (Experiment 3). The striking contrast in results for computing SSRs across multiple ensembles (fixed capacity) versus computing an SSR for a single ensemble (unlimited capacity) provides an explanation for apparently conflicting results and conclusions regarding the processing limitations of SSRs within the literature.

2.4   Experiment 1: Establishing multiple summaries of mean size

2.4.1   Methods

An N* power analysis, which calculates the number of subjects necessary to have at least 80% power for every factor (Cohen, 1988), determined the minimum number of observers needed in our experiments. Effect size estimates for this analysis were based on a pilot run of the experiment with 3 subjects. This analysis indicated that at least 7

17

subjects were necessary to detect effects in this design if they were there. For good measure, we increased this number by 5 prior to running any study. Twelve undergraduate volunteers from the University of Iowa participated in exchange for course credit (6 male, 6 female, age range: 18 – 26 years, all right-handed). All observers were naïve as to the purpose of the study and all reported normal visual acuity and color vision.

Stimuli were displayed on a flat-screen cathode ray tube monitor (19-inch ViewSonic G90fB) controlled by a Macintosh Pro (Mac OS X) with a 512MB NVIDIA GeForce 8800 GT graphics card (1024 by 768 pixels, viewing distance of 61.5 cm, refresh rate of 100 Hz). Stimuli were generated using the Psychophysics Toolbox Version 3.0.8 (Brainard, 1997; Pelli, 1997) for MATLAB (Version 7.5, Mathworks, MA). Observers sat in a height-adjustable chair and used an adjustable chin rest to maintain a constant viewing distance from the monitor.

Displays consisted of sixteen filled circles of various sizes (*Figure 2.1*), which were presented as luminance increments (43.03 cd/m^2) on an achromatic background (39.45 cd/m^2). The circles were configured to give rise to the perception of four clusters centered 4.19° from fixation. Each cluster was composed of four circles whose sizes were chosen from a target or distractor distribution. The center of the circle closest to fixation was 3.26° away, while the circle furthest from fixation was 5.59° away. Clusters were separated horizontally and vertically by 6.05° center-to-center.

18
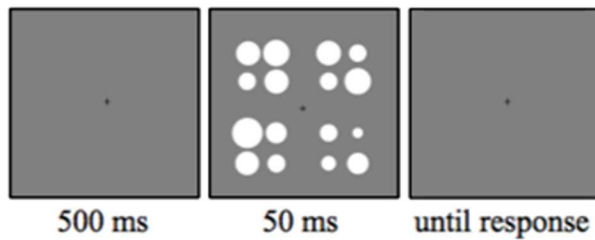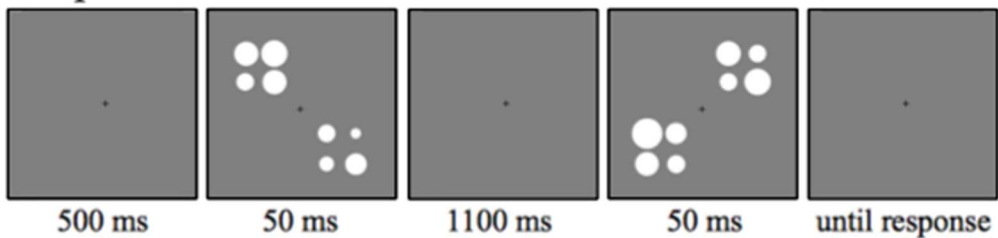
Figure 2.1. Trial events for the (A) simultaneous, (B) sequential, and (C) repeated conditions in Experiment 1. Observers saw four clusters of differently sized circles (1 target, 3 distractors) and reported whether the mean size of the target cluster was relatively smaller or larger than the mean of the distractor clusters. In this example, the target cluster is smaller and presented in the lower right.

On every trial, the sizes of circles within three of the four clusters were randomly chosen from a uniform distractor distribution (Range: 1.09° - 1.96°), while the sizes of circles within the fourth cluster were equally chosen from either a uniform small-target distribution (Range: 0.34° - 1.38°) or a uniform large-target distribution (Range: 1.40° - 2.21°). Each distribution contained 122 possible diameter sizes. The sizes were equally

spaced on a power function with an exponent of 0.76, identified by Teghtsoonian (1965) as the psychological scale for size (see also Chong & Treisman, 2003). The lower bound of the distractor distribution was the median of the small-target distribution while the upper bound of the distractor distribution was the median of the large-target distribution. The heavy overlap between target and distractor distributions minimized the degree to which observers could bypass the averaging process by using size information of individual circles to perform the task. While this potential strategy is not eliminated in the current experiment (i.e., the distributions did not fully overlap), it is only a concern if evidence of unlimited capacity is obtained. Stated another way, simply using feature information to determine target identity predicts equal performance between the simultaneous and sequential conditions because extreme sizes exclusive to the target distributions would "pop-out" and would be processed with parallel, unlimited capacity (Huang & Pashler, 2005).

Observers completed one 45-minute session that consisted of a practice block of 30 trials, followed by 6 experimental blocks of 48 trials each (96 observations per condition, 288 experimental observations per subject). Practice trials were excluded from all analyses.

Trials began with a centrally located black fixation cross ($0.25° \times 0.25°$) for 500 ms. In the simultaneous condition, this was followed by the four clusters for 50 ms, and then a blank screen until response (*Figure 2.1A*). In the sequential condition, fixation was followed by two clusters for 50 ms presented along either the positive or negative diagonal, a blank ISI of 1,100 ms, the other two clusters for 50 ms presented along the opposite diagonal, and a blank screen until response (*Figure 2.1B*). The repeated

condition was the same as the sequential condition except that all four clusters appeared in both of the two 50 ms displays (*Figure 2.1C*). Written feedback was given for 1,000 ms in the form of words "correct" or "incorrect" at fixation following each response. The next trial automatically began 1,000 ms after the presentation of feedback.

Display type (simultaneous, sequential, repeated), target type (small, large), and target position (upper-left, upper-right, lower-left, lower-right) were randomly mixed within blocks of trials and appeared equally often. Which of the two diagonally opposite positions were presented first in the sequential display was constant for a given observer but varied across observers. The purpose of this was to eliminate uncertainty of presentation positions.

The task was to find the cluster of circles that had a different mean size than the other three, and to report whether it was smaller or larger than the others by pressing the "F" or "J" key, respectively. Observers were instructed to maintain central fixation and respond as accurately as possible.

All theoretical models assume a reliable advantage in the repeated condition relative to the simultaneous condition. Subjects who did not meet this criterion were omitted from all analyses and replaced until a total of 12 subjects in each experiment were collected. One, one, and seven subjects failed to show a repeated advantage in Experiments 1, 2, and 3.[1] In the experiments that follow, accuracy data were transformed to arcsin values to normalize their distributions. The underlying assumptions of all statistical tests were confirmed and corrections were made if needed. Violations of normality and sphericity were confirmed using a one-sample Kolmogorov-Smirnov test

---

[1] By reviewer request, we later ensured that the inclusion of filtered subjects did not alter our conclusions in any substantive way.

21

and Mauchly's test. Violations of sphericity were corrected using the Greenhouse-Geisser epsilon. Follow-up *t*-tests were used after significance of the final model was verified.

2.4.2   Results and discussion

Figure 2.2 shows the mean percent correct as a function of condition, collapsed across observers. Performance was higher in the sequential condition than the simultaneous condition, which is inconsistent with an unlimited-capacity model of SSRs but consistent with a limited-capacity model. Moreover, because performance was as high in the sequential condition as in the repeated condition, the results are consistent with the fixed capacity version of the limited-capacity model. Inferential statistics confirmed these descriptive patterns.

Because the sizes of circles were chosen randomly from partially-overlapping distributions, a small percentage of trials would by chance include a distractor cluster whose mean size was either greater than (or less than) the mean size of the large or small target cluster, respectively. As a result, the cluster that appeared to be the target might in fact be a distractor cluster. We omitted these trials from the reported analyses. Out of 3,456 experimental trials across all observers, a total of 248 (7%) were omitted for this reason. Elimination of these trials did not alter the results qualitatively.
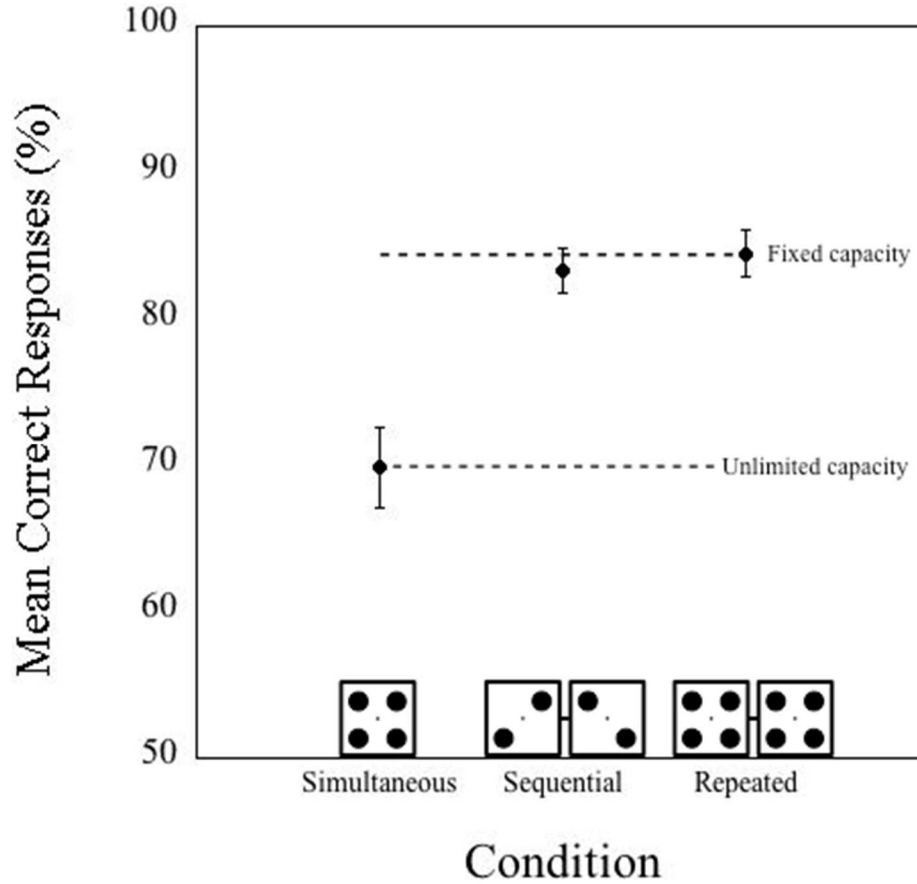
Figure 2.2. Mean correct responses (%) as a function of display collapsed across observers in Experiment 1. We found equal performance between the sequential and repeated conditions, and a reliable difference in the simultaneous condition. These results suggest that summary statistic representations engage fixed-capacity processes when multiple clusters require averaging. Error bars are within-subject standard errors (Cousineau, 2005; Moray, 2008).

Arcsin transformed values were submitted to a one-way repeated-measures ANOVA with condition as the within-subjects factor (all Kolmogorov-Smirnov $p > .918$; Mauchly's $p = .027$, Greenhouse-Geisser epsilon = .661). The final model was significant, $F(1.32, 14.54) = 16.91$, $p < .001$, $p\eta^2 = .606$, $MSE = .010$. As predicted by fixed-capacity processing, accuracy was not reliably greater in the sequential condition (83.4%) than in the repeated condition (84.5%), $t(11) = 1.06$, $p = .313$. However,

performance in the sequential condition was significantly greater than in the simultaneous condition (69.9%), $t(11) = 4.22$, $p = .001$.

An assumption of the simultaneous-sequential method is that the conditions differ only with respect to how many stimuli must be processed simultaneously. They did necessarily differ, however, in when the target appeared within the trial sequence. In the simultaneous condition the target always appeared in the "first" frame because that was the only frame, whereas in the sequential condition, the target could appear in either the first frame or in the second frame. This difference might provide an advantage to the simultaneous condition if there are any memory differences across the two conditions. To assess this possibility, we compared performance in the sequential condition for trials on which the target appeared in the first and second frames. No reliable differences were observed: 82.3% (first frame) vs. 83.7% (second frame), $F(1,11) = 0.25$, $p = .625$, pη^2 = .023, $MSE = .005$ (all Kolmogorov-Smirnov $p > .898$).

The results of this experiment indicate that establishing SSRs of size engage limited-capacity processes, and, in particular, that only a fixed amount of information can be processed per unit time. Furthermore, the reliable difference between the simultaneous and sequential conditions shows this experiment had the power to detect an unlimited-capacity result. In summary, the results of Experiment 1 are consistent with a limited-capacity model of SSRs for multiple ensembles, and not with an unlimited-capacity model.

2.5 Experiment 2: Control experiment

We have interpreted the results of Experiment 1 as evidence that SSRs of mean size involve limited-capacity processes. Successful performance, however, required that observers not only compute the mean size of each cluster, but also compare those means and determine whether the mean furthest away, in numerical terms, was relatively smaller or larger. To rule out the possibility that it was some other aspect of the task that caused performance to be limited capacity, we conducted a control experiment in which the task required all of the same processes except computing mean size. Subjects were shown clusters of homogeneous circles, the size of each determined by the mean of their respective cluster from Experiment 1. Averaging was no longer required since the mean of each cluster was directly provided and since all circles within a cluster were of equal size (*Figure 2.3*). The task was the same otherwise. If the limited capacity results of Experiment 1 were caused by limited capacity SSR formation and nothing else, then we should find evidence of unlimited-capacity processing in this second experiment.

2.5.1 Methods

All aspects of the stimuli and procedure were identical to Experiment 1, with the following exceptions.

Twelve undergraduate volunteers from the University of Iowa participated in exchange for course credit (4 male, 8 female, age range: 17 – 20 years, 11 right-handed).
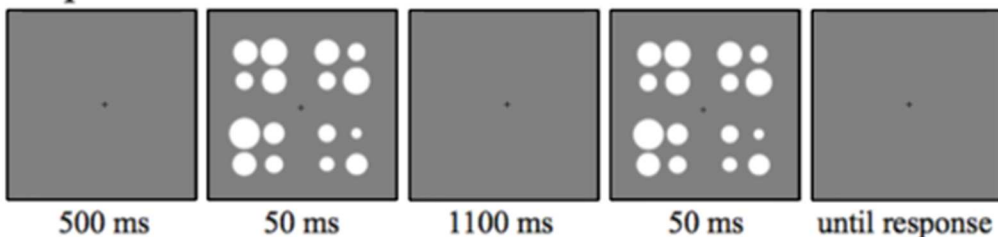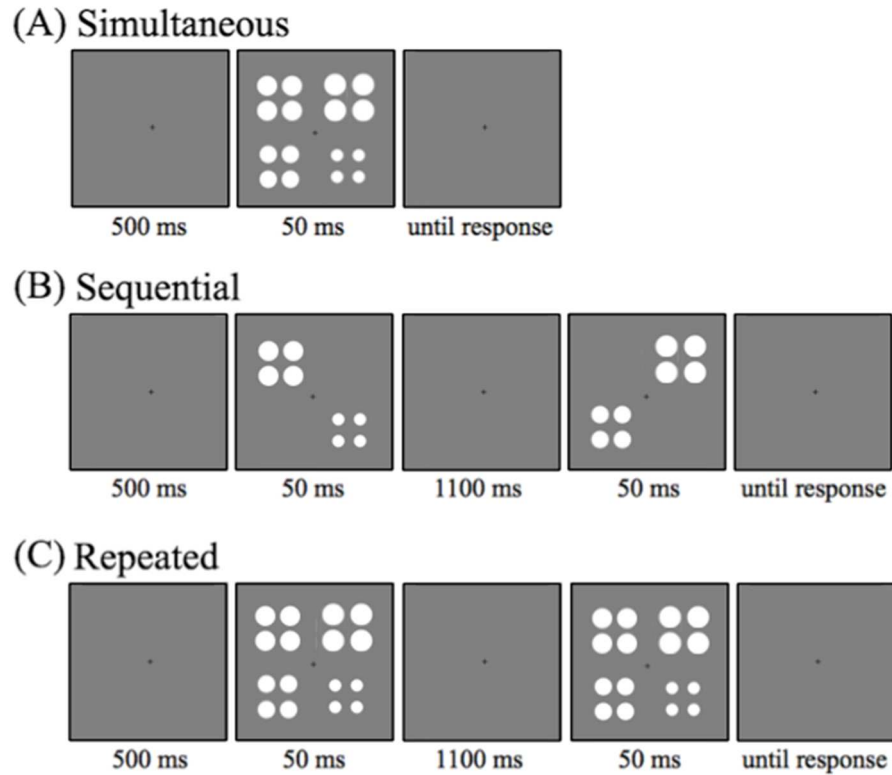
Figure 2.3. Trial events for the (A) simultaneous, (B) sequential, and (C) repeated conditions in Experiment 2. The size of all circles in a given cluster reflected that cluster's mean. Computing the mean for each cluster was no longer required to perform the task since the circles were of equal size. The target cluster is smaller and presented in the lower right in this example.

As in Experiment 1, the sizes of sixteen circles were randomly chosen from the appropriate target or distractor distribution. But before the stimuli were presented, the mean size for each cluster was computed. The size of all circles within a given cluster was adjusted according to that cluster's mean prior to presentation (*Figure 2.3*). As a result, subjects could circumvent the averaging process by directly comparing individual circles to determine whether the oddball cluster was relatively larger or smaller than the others.

2.5.2   Results and discussion

Figure 2.4 shows the mean percent correct as a function of condition collapsed across observers. Performance was no different in the sequential condition than in the simultaneous condition, but it was higher in the repeated condition than the other two. This pattern is consistent with an unlimited-capacity model and inconsistent with a limited-capacity model. The inferential statistics confirmed this pattern.
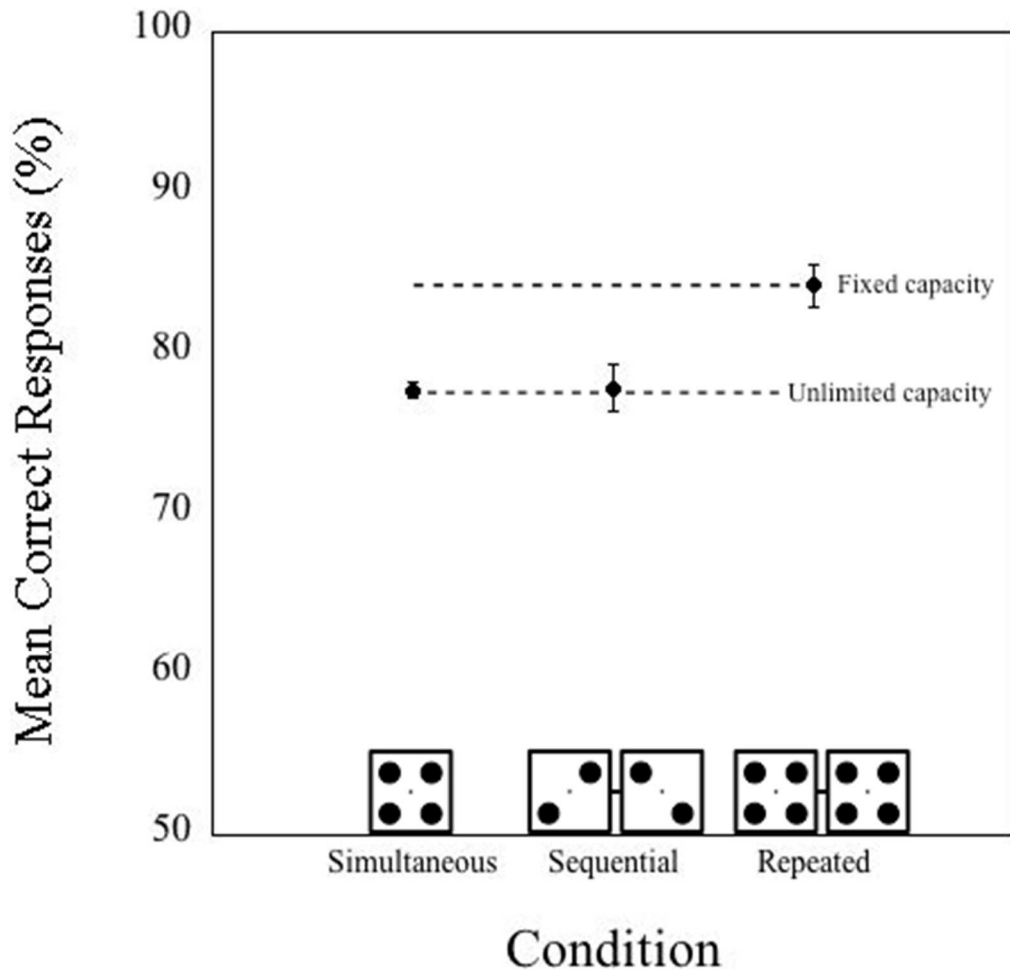
Figure 2.4. Mean correct responses (%) as a function of display collapsed across observers in Experiment 2. Evidence consistent with unlimited capacity was obtained when the task no longer required that subjects compute the average of each cluster. Error bars are within-subject standard errors (Cousineau, 2005; Moray, 2008).

27

As in Experiment 1, we filtered trials in which the perceptually correct response may have led to an "incorrect" feedback message (241 trials of 3,456 total trials across observers, for 7%). Arcsin transformed values were submitted to a one-way repeated-measures ANOVA with condition as the within-subjects factor (all Kolmogorov-Smirnov $p > .895$; Mauchly's $p = .003$, Greenhouse-Geisser epsilon = .591). The final model was significant, $F(1.18, 13.01) = 10.89$, $p = .004$, $p\eta^2 = .498$, $MSE = .004$. As predicted by unlimited-capacity processing, accuracy was not reliably greater in the sequential condition (77.8%) than in the simultaneous condition (77.7%), $t(11) = 0.37$, $p = .722$. However, performance in the sequential condition was significantly lower than performance in the repeated condition (84.3%), $t(11) = 2.88$, $p = .015$.

We again compared performance within sequential trials when the target was presented in the first frame versus the second frame. Again, we found that performance across both frames were statistically equal, 78.5% (first frame) vs. 76.4% (second frame), $F(1,11) = 0.42$, $p = .532$, $p\eta^2 = .036$, $MSE = .010$ (all Kolmogorov-Smirnov $p > .877$), suggesting that targets presented first did not suffer from more memory loss than targets presented closer in time to response.

In summary, the results of Experiment 2 indicate that when the task no longer requires the computation of averages, processing becomes unlimited capacity. The fact that the results of Experiment 1 indicated fixed capacity can be confidently interpreted as evidence that SSRs depend on limited averaging processes, and not on limited comparison or decision processes. The crowding of items within each cluster also cannot explain the reported limitation since the stimulus spacing in Experiment 1 was preserved in Experiment 2 (Banno & Saiki, 2012; Bouma, 1970).

2.6 Experiment 3: Establishing a single summary of mean size

Experiment 1 showed that computing SSRs of size for multiple ensembles engages fixed-capacity processes. But what about computing an SSR for a single ensemble of stimuli? Previous studies asking about SSRs have rarely made this distinction. Some have used tasks in which SSRs are computed across a single ensemble (e.g., Ariely, 2001, Robitaille & Harris, 2011), whereas others have used tasks in which SSRs are computed across multiple ensembles (e.g., Banno & Saiki, 2012; Oriet & Brand, 2013). It is possible that computing SSRs is limited by the number of ensembles for which an SSR is extracted, but that computing a single SSR is not limited by the number of stimuli across which the summary is made. If that were the case, then not distinguishing between tasks that depend on SSRs of single versus multiple ensembles could lead to apparently conflicting conclusions about whether computing SSRs is limited capacity. Indeed such conflicting conclusions exist. For example, the fixed-capacity conclusion drawn from Experiment 1 of this study, which involved multiple ensembles, contrasts with the unlimited-capacity conclusion drawn from a study reported by Robitaille and Harris (2011), which focused on single ensembles. Experiment 3 used the simultaneous-sequential method to test the capacity limitations for a task that depended on only a single-ensemble SSR.

2.6.1 Methods

All aspects of the stimuli and procedure were identical to Experiment 1, with the following exceptions.

Twelve undergraduate volunteers from the University of Iowa participated in exchange for course credit (8 male, 4 female, age range: 18 – 30 years, all right-handed).

Sixteen filled circles of various sizes were placed on a square grid spaced horizontally and vertically by 2.21° and centered at fixation (*Figure 2.5*). *Procedure*. On each practice trial, a black probe circle (1.39°) appeared on the response screen after the simultaneous, sequential, and repeated displays. The size of the probe circle was fixed and subjects were instructed to report whether the average of all sixteen circles was smaller or larger than the size of the probe circle appearing afterward. After the practice block, subjects were told that while the probe circle would not appear on experimental trials, their task remained the same because the probe remained a fixed size. This kept the trial events consistent across all three experiments. The size of the probe circle was a unique value in the distractor distribution; it did not match any of the sizes falling in either the small or large target distributions. On trials in which the target was "small", the average of the entire cluster was shifted lower than the size of the probe circle; conversely, on trials in which the target was "large", the average of the entire cluster was shifted higher than the size of the probe circle.

The task was to report whether the average of the single set was smaller ("F" key) or larger ("J" key) than the probe circle that had been presented throughout the practice block.
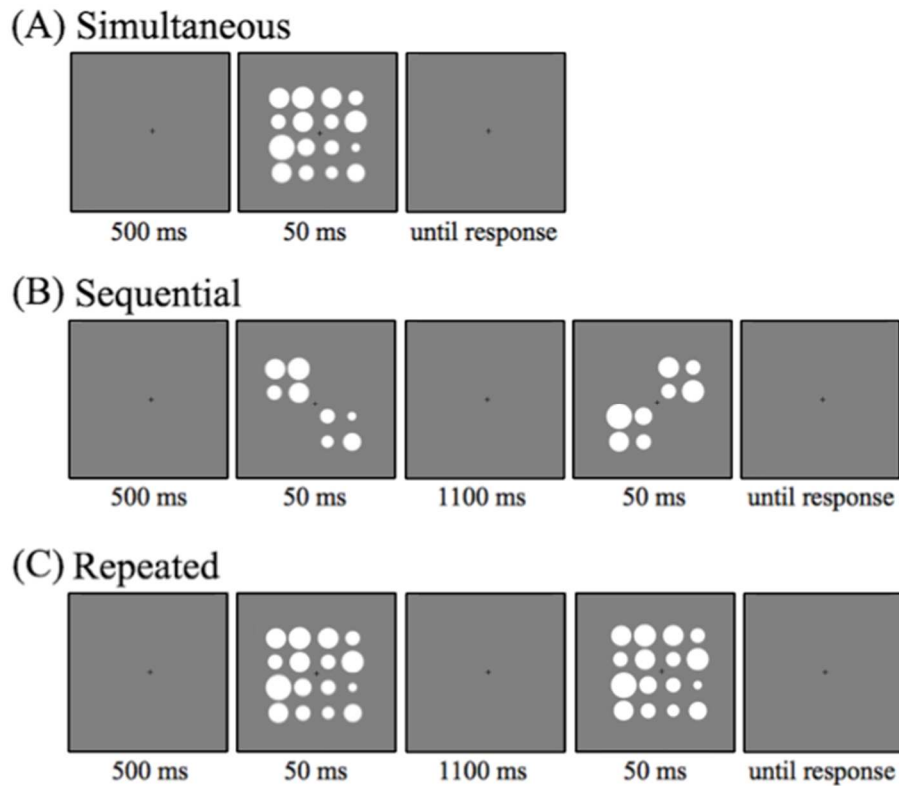
Figure 2.5. Trial events for the (A) simultaneous, (B) sequential, and (C) repeated conditions in Experiment 3. The 4 clusters presented in Experiment 1 were presented on an equally-spaced grid to produce the perception of a single cluster with 16 items. During practice trials (not pictured) a probe circle appeared on the response screen and subjects reported whether the mean size of the single cluster was larger or smaller than the size of the probe circle. In the real experiment (pictured), presentation of the probe circle was removed because it remained the same size on every trial. In this example, the correct response is "smaller".

2.6.2   Results and discussion

Figure 2.6 shows the mean percent correct as a function of condition collapsed across observers. Performance was no different in the sequential condition than in the

simultaneous condition, but it was higher in the repeated condition than the other two. This pattern is consistent with an unlimited-capacity model and inconsistent with a limited-capacity model. The inferential statistics confirmed this pattern.

As before, we filtered trials in which the perceptually correct response may have led to an "incorrect" feedback message (233 trials of 3,456 total trials across observers, for 7%). Arcsin transformed values were submitted to a one-way repeated-measures ANOVA with condition as the within-subjects factor (all Kolmogorov-Smirnov $p > .711$; Mauchly's $p = .093$, Greenhouse-Geisser epsilon = .725). The final model was significant, $F(1.45,15.96) = 14.72$, $p = .001$, $p\eta^2 = .572$, $MSE = .003$. As predicted by unlimited-capacity processing, accuracy was not reliably greater in the sequential condition (79.8%) than in the simultaneous condition (78.7%), $t(11) = 0.59$, $p = .569$. However, performance in the sequential condition was significantly lower than performance in the repeated condition (85.1%), $t(11) = 3.55$, $p = .005$.

Performance across both frames in the sequential condition were statistically equal, 81.1% (first frame) vs. 80.2% (second frame), $F(1,11) = 0.23$, $p = .644$, $p\eta^2 = .020$, $MSE = .008$ (all Kolmogorov-Smirnov $p > .520$), suggesting that targets presented first did not suffer from more memory loss than targets presented closer in time to response.
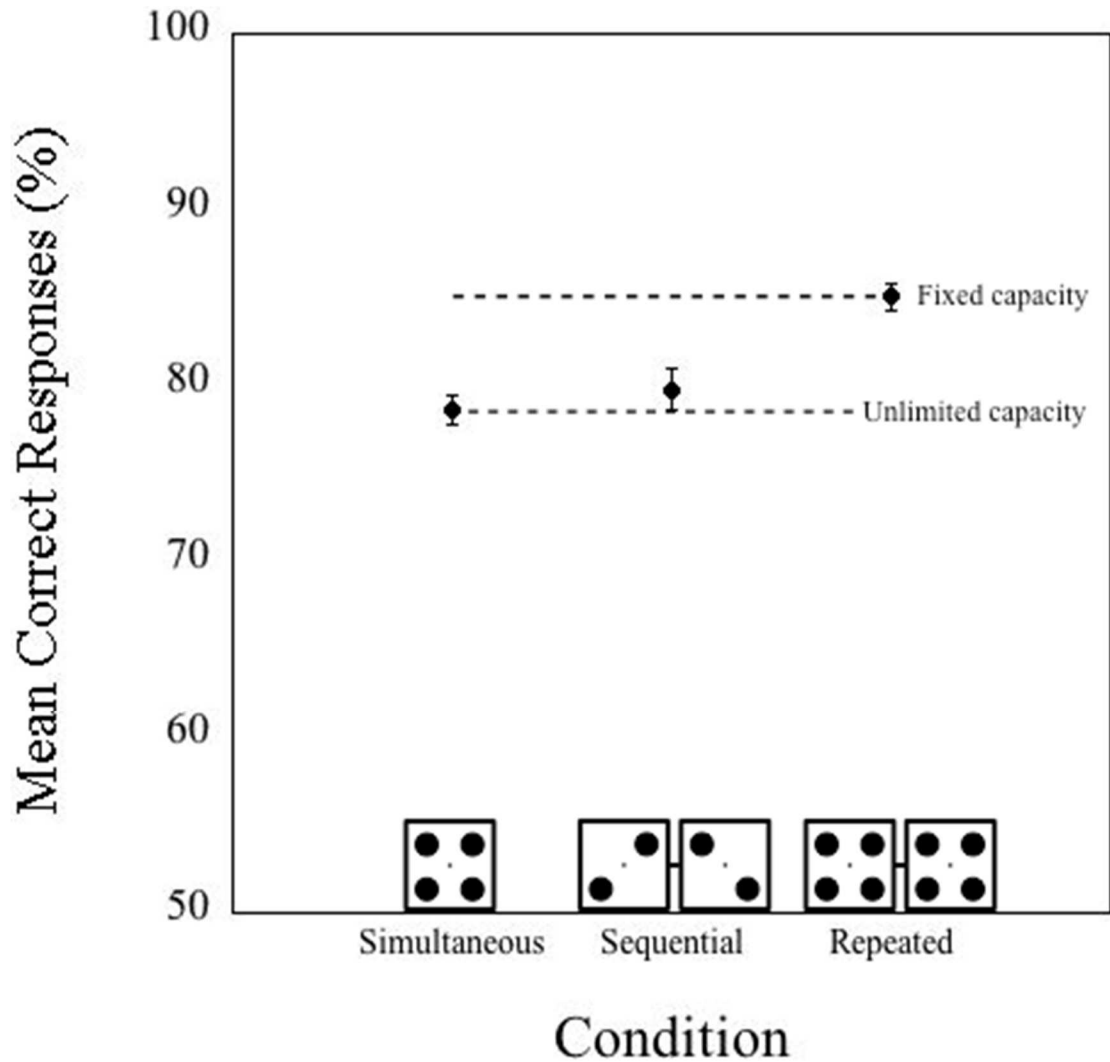
Figure 2.6. Mean correct responses (%) as a function of display collapsed across observers in Experiment 3. Evidence consistent with unlimited capacity was obtained when summary statistics were computed for a single cluster. Error bars are within-subject standard errors (Cousineau, 2005; Moray, 2008).

In summary, the results of Experiment 3 indicate that SSRs of mean size for single ensembles engage only unlimited-capacity processes. When the same sixteen items were grouped into four clusters in Experiment 1, the results were consistent with the opposite processing extreme. Computing summary representations for multiple ensembles introduces interference unlike single ensembles.

33

2.7   General discussion

Applying the simultaneous-sequential method to test the capacity limitations of SSRs, we found evidence that was consistent with a limited-capacity model for the formation of SSRs for multiple sets and an unlimited-capacity model for the formation of multiple items within a single set. Specifically, observers were poorer at responding on the basis of mean size of four clusters of circles when the clusters were all presented simultaneously compared to when they were presented two at a time sequentially. In fact, because performance was equally good in the sequential condition as in the repeated condition, the results suggest that computing mean size involve fixed-capacity processing, an extreme version of the limited-capacity model (Scharff et al., 2011a). When the same items were presented as a single perceptual unit, mean size was computed through unlimited-capacity processes and performance was equally good across the simultaneous and sequential conditions. One large set can be averaged more efficiently than multiple smaller sets.

The current results indicate that the formation of SSRs across multiple ensembles of stimuli depends on limited-capacity processes (i.e., ensembles are not processed independently), whereas the formation of a single SSR for one ensemble of stimuli seems to be unlimited capacity (i.e., stimuli within an ensemble are processed independently). It has been proposed that a compressed representation of the environment that is established through the formation of multiple SSRs bypasses limited-capacity components of our perceptual and cognitive systems and serves to guide later visual

processes. We suggest that this cannot be the case given the highly limited nature of forming multiple SSRs.

We would like to note that Chong and Treisman (2003, experiment 1) tested thresholds for mean size summaries for two sets of circles that appeared either simultaneously or sequentially. Performance was equal between both presentation types and it was concluded that two averages could be generated just as well as one. At first glance, this design appears to be the simultaneous-sequential method; however, Chong and Treisman did not intend to use this method and thus a critical aspect of the design, namely, the use of constant exposure durations, was violated. The duration of the simultaneous condition was twice as long (200 ms) as the duration of each frame in the sequential condition (100 ms). The simultaneous condition in Chong and Treisman is therefore most similar to the repeated condition in Experiment 1 of the current paper. We conclude that their results are actually consistent with ours; performance in the sequential condition achieved that of the double-duration condition, suggesting that the statistical extraction of multiple ensembles engages at least some fixed-capacity processes (see Scharff et al., 2011).

Finally, we end with a discussion of the contrast between processing capacity (the degree to which a process can be engaged independently by multiple stimuli; Broadbent, 1958; Estes & Taylor, 1964; Rumelhart, 1970; Shiffrin & Gardner, 1972) and storage capacity (the amount of information that can be maintained in memory; Alvarez & Cavanagh, 2004; Awh, Barton, & Vogel, 2007; Luck & Vogel, 1997). Many recent studies have investigated the storage capacity of visual working memory. In an initial paper, Luck and Vogel (1997) used a simple change-detection method to estimate that

observers were able to hold approximately 3 stimuli in visual working memory. This study lead to a flurry of follow-up studies asking questions about the nature of this capacity limitation, such as whether it is limited by the number of objects that can be held or the degree of precision with which stimuli can be remembered, or both. Because estimates of storage capacity from these studies tend to be on the order of 2.5 - 4 items (see Brady, Konkle, & Alvarez, 2011 for a review), there has been a tendency to criticize the use of the simultaneous-sequential method with conditions that vary from two-at-a-time presentations (sequential) to four-at-a-time presentations (simultaneous) because both 2 and 4 fall within the range of most people's 'capacity'. It is critical to note, however, that the simultaneous-sequential method is assessing processing independence versus dependence, not storage capacity. If stimulus presentation conditions are such that performance is limited by how much information can be extracted from the display (e.g., because stimuli are presented briefly), then limited-capacity processing predicts a difference between simultaneous versus sequential even for one versus two items. Two versus four has been used in order to minimize contamination from differences in eye movements across conditions and to minimize contamination from sensory effects like crowding, but the logic is identical. Finally, if the criticism regarding four items is too small were valid, evidence of limited-capacity should never attain. Yet it has for many different tasks, including shape identification, spatial configuration, object categorization, and word categorization (Huang & Pashler, 2005; Scharff et al., 2011a, 2011b, 2013). Thus, while Experiment 3 clearly demonstrates unlimited capacity even for 16 items, it is important to note that the logic of the simultaneous-sequential method does not depend on this extension.

3.3 Overview

The simultaneous–sequential method was used to test the processing capacity of establishing mean orientation summaries. Four clusters of oriented Gabor patches were presented in the peripheral visual field. One of the clusters had a mean orientation that was tilted either left or right while the mean orientations of the other three clusters were roughly vertical. All four clusters were presented at the same time in the simultaneous condition whereas the clusters appeared in temporal subsets of two in the sequential condition. Performance was lower when the means of all four clusters had to be processed concurrently than when only two had to be processed in the same amount of time. The advantage for establishing fewer summaries at a given time indicates that processing multiple sets of mean orientation engages limited-capacity processes (Experiment 1). This limitation cannot be attributed to crowding, low target-distractor discriminability, or a limited-capacity comparison process (Experiments 2 and 3). In contrast to the limitations of establishing multiple summary representations, establishing a single summary representation unfolds without interference (Experiment 4). When interpreted in the context of recent work on the capacity of summary statistics, these findings encourage reevaluation of the view that early visual perception consists of SSRs that unfold independently across multiple areas of the visual field.

3.4   Introduction

The view that SSRs are a fundamental aspect of early visual processing is dependent on the claim that summaries are computed over many items in the visual field independently. That is, they are assumed to depend entirely on unlimited-capacity processes. In the current study, we applied the extended simultaneous–sequential method (Scharff et al., 2011a) to ask whether establishing SSRs of mean orientation depends on limited-capacity processes or whether they can be established entirely through unlimited-capacity processes. In a recent study, we addressed this question for the establishment of mean size and found that representing mean size for multiple ensembles depended on limited-capacity processes (Attarha et al., 2014b). This finding presents a challenge to the hypothesis that the functional role of SSRs is to reduce complex information across the visual field to support later processes and the sense of perceptual continuity (e.g., Alvarez, 2011; Chong & Treisman, 2005a; Whitney, Haberman, & Sweeny, 2014).

Why follow up with orientation? One reason for considering the processing limitations of establishing SSRs for orientation, in particular, is that the visual search literature suggests that orientation information may be processed in a manner that is qualitatively different from other simple features. For example, when within-feature conjunctions are configured in a whole-part structure, attention can be guided by size

38

(and color) but not by orientation. One possible explanation is that orientation may not be processed hierarchically to the same extent as other features (Bilsky & Wolfe, 1995; Wolfe, Friedman-Hill, & Bilsky, 1994; Wolfe et al., 1990). The results of this study and others (e.g., Cavanagh, Arguin, & Treisman, 1990; Lüschow & Nothdurft, 1993) suggest that orientation processing may be unique and thus it follows that any limitations or advantages observed for size may not generalize to orientation. If mean orientation SSRs can be established through unlimited-capacity processes, then it would provide evidence that at least some summary representations might serve in the role of abstracted information in the support of later visual processes (e.g., Alvarez, 2011; Rosenholtz et al., 2012). Alternatively, finding that orientation SSRs also depend on limited-capacity processes would challenge the widespread claim that SSRs precede or bypass the limited-capacity bottleneck.

A second, related, reason for considering the capacity limitations of establishing SSRs for orientations concerns a theoretical account of SSRs according to which summaries are generated at multiple levels and within separate pathways of the visual system (Haberman & Whitney, 2009; Haberman & Whitney, 2011; Whitney et al., 2014). According to this view, averages for some low-level surface features, such as orientation and brightness, may be established at the earliest stages of processing whereas SSRs for other attributes may not be established until later stages (Whitney et al., 2014; p. 702). Average object size and shape, for example, may be processed further along the ventral stream than mean orientation. Similarly, mean direction of motion and mean spatial position may be processed further along the dorsal stream than orientation. Still, other summary representations (e.g., biological motion or facial expression) may not be

39

processed until after the ventral and dorsal pathways converge.

Under this multiple-site view of SSR formation, different SSRs will engage different subsets of processes; some may involve limited-capacity processing, whereas others may bypass all limited-capacity processes. For example, summaries of low-level features may be mediated by physiological mechanisms that pool the activity of a population of early feature channels in parallel, while summaries of more complex representations may involve more complex algorithms (e.g., this issue is discussed in Myzczek & Simons, 2008, p. 773; see also Marchant et al., 2013, p. 245). Although the algorithms by which summary statistics operate are currently unknown, linear pooling models have shown promise (Haberman & Whitney, 2011; Parkes et al., 2001). Specifically, for features that are explicitly represented in early visual stages, such as orientation, pooling mechanisms may combine the outputs of orientation-selective cells into a Gaussian-shaped population code, the center of which could be the basis of a summary percept (e.g., Suzuki, 2005; Whitney et al., 2014). Averaging across low-level feature detectors in this way may be an intrinsic aspect of visual processing that proceeds without capacity limitations. In contrast, more complex summaries (e.g., facial averaging) may require an additional step wherein summaries of multiple component feature populations are integrated into a superordinate population code. The additional step of integrating subordinate summaries may produce an information-processing bottleneck, thus limiting the processing capacity of such complex summaries. According to this framework, orientation averaging is a likely candidate for unlimited-capacity processing (Dakin, 2001; Dakin & Watt, 1997; see also Hubel & Wiesel, 1962; Webster & De Valois, 1985), whereas facial averaging is a likely candidate for limited-capacity

processes.

By way of preview, the results from the current study are inconsistent with the hypothesis that orientation SSRs are established entirely through unlimited-capacity processes. That is, like size, the establishment of a representation of mean orientation cannot be done for multiple ensembles without interference. So far, there is little evidence that any SSRs bypass limited-capacity processes. As such, SSRs do not seem to be good candidates for the computation-saving representations that they are believed to serve as, at least not the versions tested so far using this method.

3.5 Experiment 1: Establishing multiple summaries of mean orientation

3.5.1   Methods

Twelve undergraduate volunteers from the University of Iowa participated in exchange for course credit (5 male, 7 female, age range: 18 – 28 years, 10 right-handed). A power analysis (N*; Cohen, 1988) based on a pilot run of this experiment indicated that only five subjects were needed to achieve at least 80% power. We made an a priori decision to run twelve to be consistent with a similar study that tested the capacity limitations of mean size summaries (Attarha et al., 2014b). All observers reported normal visual acuity and color vision.

Stimuli were displayed on a cathode ray tube monitor (19-inch ViewSonic G90fB) controlled by a Macintosh Pro (Mac OS X) with a 512MB NVIDIA GeForce 8800 GT graphics card (1024 by 768 pixels, viewing distance of 61.5 cm, refresh rate of

100 Hz). Stimuli were generated using the Psychophysics Toolbox Version 3.0.11 (Brainard, 1997; Pelli, 1997) for MATLAB (Version 8.2, Mathworks, MA). Observers sat in a height-adjustable chair and used an adjustable chin rest to maintain a constant viewing distance from the monitor. The room was dimly lit.

Thirty-six Gabor patches (Gabor, 1946) of various orientations were presented on a neutral gray background (37.14 cd/m^2) at the maximum contrast that could be produced by the monitor (50.06 cd/m^2) (*Figure 3.1*). It has been previously established that orientation averaging can operate over Gabor stimuli (e.g., Dakin, 2001; Dakin & Watt, 1997; Parkes et al., 2001). All sinusoidal patches (1.58° in diameter) had a spatial frequency of 3 cycles per degree and were windowed by a symmetric Gaussian envelope with a spatial constant of 7 pixels. The Gabors were spatially grouped to give rise to the perception of four clusters, each centered on a corner of an imaginary square approximately 6.24° from fixation. The center of the Gabor closest to fixation was 2.89° away, while the center of the Gabor furthest from fixation was 9.94° away. A distance of 9.11° separated the clusters horizontally and vertically, center-to-center.

42
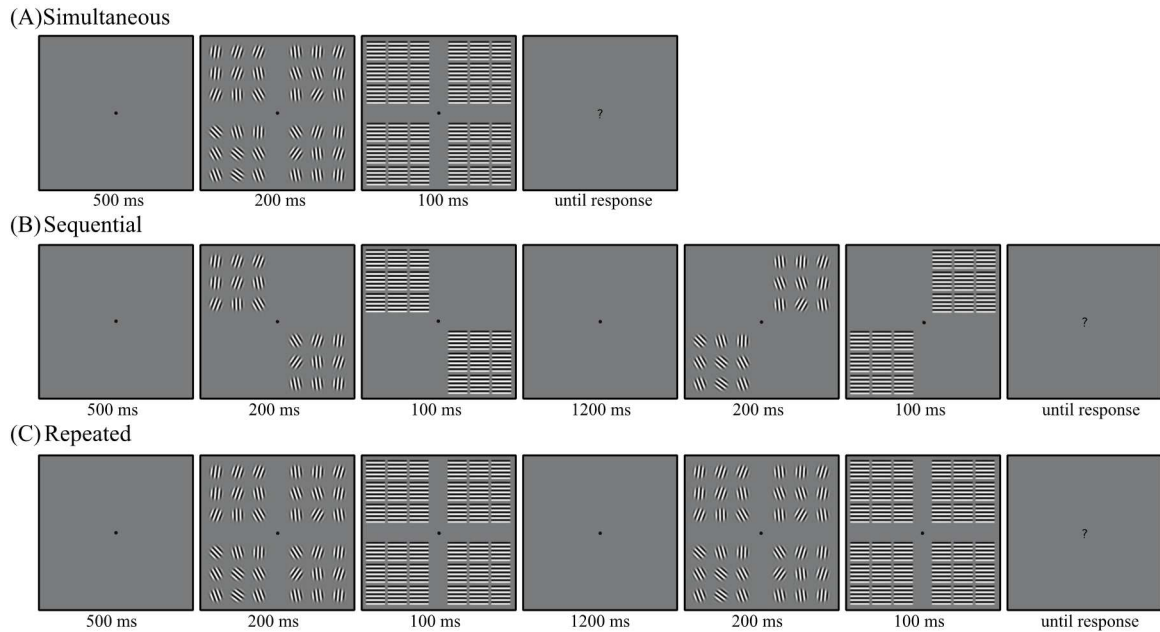
Figure 3.1. Trial events for the (A) simultaneous, (B) sequential, and (C) repeated conditions in Experiment 1. Observers saw four clusters of Gabor patches. One cluster consisted of tilted Gabors randomly sampled from a target distribution of orientations while the other three clusters consisted of Gabors sampled from a distractor distribution. Observers reported whether the mean orientation of the oddball cluster was tilted left or right relative to the others. The target cluster is tilted left and presented in the lower left corner in this example.

On every trial, the orientations of the Gabor patches within each cluster were chosen from a target or distractor distribution. Three of the four clusters were chosen randomly from a Gaussian distractor distribution ($\mu = 0°$; $\sigma = 15°$), while the orientations of Gabors within the fourth cluster were chosen equally from either a Gaussian tilted-left distribution ($\mu = -30°$; $\sigma = 15°$), or a Gaussian tilted-right distribution ($\mu = 30°$; $\sigma = 15°$). Vertical was 0°.

Observers completed one 30-minute session. The session began with a practice block of 30 trials, followed by 6 experimental blocks of 48 trials each (96 observations per display type, 288 experimental observations per subject). Practice trials were

excluded from all analyses.

All trials began with a centrally located fixation dot (2 pixel diameter) colored in black for 500 ms. Observers were instructed to maintain central fixation throughout the experiment. In the simultaneous condition, the fixation display was followed by the four clusters of Gabors for 200 ms. Each Gabor was subsequently masked by a square-shaped Gabor patch that was oriented horizontally at 90° (2.05° × 2.05°) for 100 ms. A blank screen with a question mark ("?") at fixation followed the mask display and remained on the screen until a response was made (*Figure 3.1A*). In the sequential condition, fixation was followed by two clusters for 200 ms presented along either the positive or negative diagonal, masks for 100 ms, a blank ISI of 1,200 ms, the other two clusters for 200 ms presented along the opposite diagonal, masks again for 100 ms, and a blank screen with a question mark until response (*Figure 3.1B*). The repeated condition was the same as the sequential condition except that all four clusters appeared in both of the two 200 ms displays (*Figure 3.1C*). Written feedback ("correct" / "incorrect") was given at fixation following each response for 500 ms. The next trial automatically began 1,000 ms after the feedback display.

The default exposure duration was 200 ms (see Whiting & Oriet, 2011). A coarse tracking procedure altered the exposure duration, block-by-block, on the basis of performance in the simultaneous condition only. If performance in the simultaneous condition was more than 90% on a given block, then the exposure duration for the simultaneous, sequential, and repeated conditions was decreased by 10 ms on the next block. Moreover, if performance was less than 60% in the simultaneous condition, then the exposure duration in all three conditions increased by 10 ms. The average adjusted

44

exposure duration across all subjects was 190 ms.

The full factorial combination of *display type* (simultaneous, sequential, repeated), *target type* (tilted left, tilted right), and *target position* (upper-left, upper-right, lower-left, lower-right) were randomly mixed within blocks of trials and appeared equally often. Which of the two diagonally opposite positions were presented first in the sequential display was constant for a given observer but varied across observers. Odd-numbered subjects saw clusters that first appeared along the negative diagonal and then along the positive diagonal. Even-numbered subjects saw clusters that appeared positive to negative. We kept the presentation of diagonal orders constant within an observer to eliminate uncertainty of the presentation positions.

Observers reported whether the mean orientation of one cluster was tilted left or tilted right relative to the mean orientation of the other clusters by pressing the "F" or "J" key, respectively. Observers were instructed to respond as accurately as possible. Speed was not emphasized.

All three models assume an advantage in the repeated condition where observers see the display twice compared to the simultaneous condition where observers see the display only once. Subjects who did not meet this criterion were omitted from further analyses and replaced until a total of 12 subjects in each experiment were collected. One, two, three, and five subjects failed to show a repeated advantage in Experiments 1-4, respectively.[2]

Because of our sampling method, we filtered the small percentage of trials in which the perceptually correct response led to an "incorrect" feedback message. In Experiments 1-3, this meant that the mean orientation of a distractor cluster was tilted

---

[2] Including filtered subjects did not alter the results qualitatively.

either more rightward (or leftward) than the mean orientation of the target cluster. The cluster that appeared to be the target was in fact a distractor on these trials. A total of 1, 0, and 0 out of 3,456 experimental trials across all twelve observers in Experiment 1, 2, and 3, were filtered, respectively. In Experiment 4, trials in which the mean of the entire set of thirty-six items was not tilted in the intended direction were filtered. A total of 8 out of 3,456 experimental trials (.0023%) were omitted. The elimination of these trials did not alter the results qualitatively.

After filtering, the accuracy data for the simultaneous, sequential, and repeated conditions were transformed to arcsin values to normalize their distributions and the underlying assumptions of the repeated-measures ANOVA were confirmed. Assumptions of normality and sphericity were confirmed using a one-sample Kolmogorov-Smirnov test and Mauchly's test, respectively. When violations of sphericity were found, *p*-values were adjusted based on the Greenhouse-Geisser epsilon correction on degrees of freedom (Jennings & Wood, 1976). Two follow-up paired *t*-tests, one between the simultaneous and sequential conditions, and another between the sequential and repeated conditions, were used after significance of the final model was verified.

3.5.2   Results and discussion

Figure 3.2 shows mean percent correct as a function of display, collapsed across all observers. Error bars are within-subject standard errors (Cousineau, 2005; Moray, 2008). Notice that *Figure 3.2* has two line labels. One of these lines defines the "unlimited capacity" prediction while the other defines the "fixed capacity" prediction.

These lines can be thought of as boundary conditions. The simultaneous condition (where subjects see all four sets one time) provides a lower bound of processing performance whereas the repeated condition (where subjects see all four sets twice) provides an upper bound of performance. The "fixed capacity" and "unlimited capacity" labels define the theoretical model that is supported as a function of where performance in the sequential conditions falls (see Scharff et al., 2011a, Appendix, for details regarding predictions). Evidence of unlimited-capacity processing is concluded if the sequential condition falls on the line established by the simultaneous condition. In contrast, evidence of fixed-capacity processing is concluded if the sequential condition falls in line with the repeated condition. In Experiment 1, we found that sequential was equal to repeated performance and that there was a reliable decrement in the simultaneous condition. This pattern of results is consistent with a fixed-capacity model and inconsistent with an unlimited-capacity model.

Arcsin transformed values of mean percent correct were submitted to a one-way repeated-measures ANOVA with the simultaneous, sequential, and repeated display conditions as the within-subjects variable. The final model was significant, $F(1.16, 12.72)$ = 5.64, p = .030, $p\eta^2$ = .339, MSE = .007 (all Kolmogorov-Smirnov p > .766; Mauchly's p = .001; Greenhouse-Geisser ε = .579). As predicted by fixed-capacity processing, performance in the sequential condition (73% ± 2.05) was significantly greater than performance in the simultaneous condition (67% ± 1.21), $t(11)$ = 2.45, p = .032. Performance between the repeated (74% ± 1.11) and sequential conditions were equal, $t(11)$ = 0.09, p = .927. We conclude that establishing SSRs of mean orientation for

47

multiple ensembles depend on limited-capacity processes, some of which may even

involve a fixed-rate processing bottleneck (see Scharff et al., 2011a)
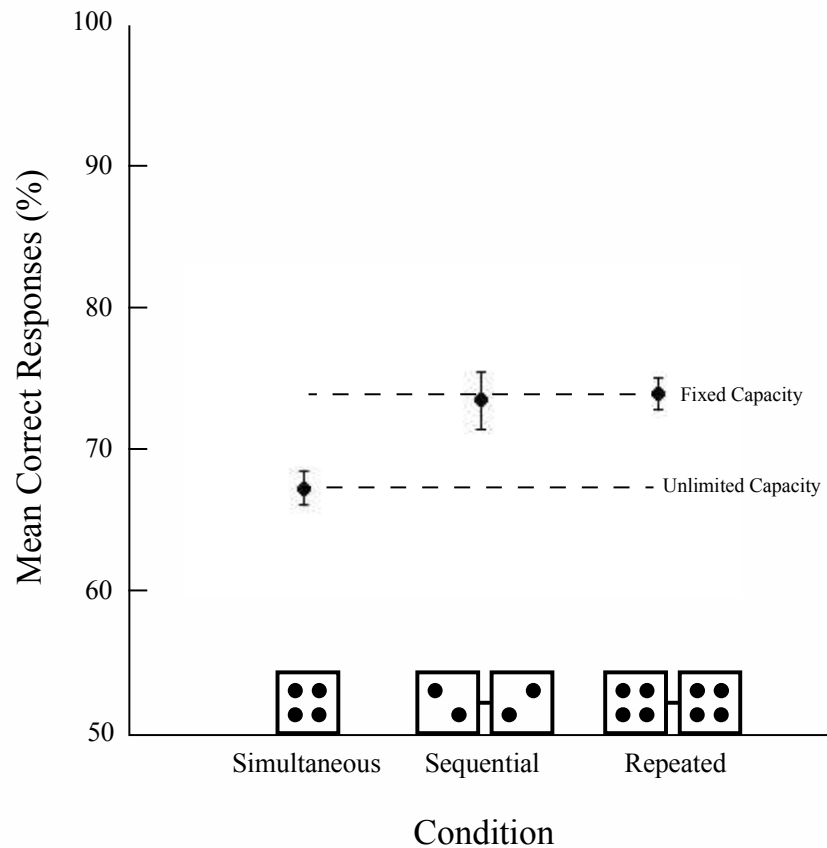
Figure 3.2. Mean correct responses (%) as a function of display collapsed across observers in Experiment 1. Performance in the sequential condition was better than performance in the simultaneous condition and equal to performance in the repeated condition. These results suggest that mean orientation SSRs for multiple sets engage fixed-capacity processes. Error bars are within-subject standard errors (Cousineau, 2005; Moray, 2008).

We would like to now offer a few alternative explanations. The simultaneous–

sequential method assumes that the simultaneous and sequential displays differ only with

respect to how many stimuli must be processed at a given time. They did necessarily

differ, however, in when the target appeared within the trial sequence. In the

simultaneous condition the target always appeared in the "first" frame because that was

the only frame, whereas in the sequential condition, the target appeared in either the first frame or the second frame. This difference might provide a disadvantage to the sequential condition if there are any memory differences across the two conditions. To assess this possibility, we compared performance in the sequential condition for trials on which the target appeared in the first and second frames. No reliable difference was found: 72% (first frame) vs. 75% (second frame), $F(1,11) = 0.63$, $p = .446$, p$\eta^2 = .054$, $MSE = .009$ (all Kolmogorov-Smirnov $p > .543$).

With our stimulus design, there are two potential strategies that can be used to bypass a calculation of mean orientation. First, responses may be based on the orientation information of individual Gabor patches rather than on mean orientation. Specifically, if the most extreme orientation in the display points leftward, for example, then observers may use this information as a shortcut to a "tilted left" response without ever calculating a summary of each cluster. We used distributions with large standard deviations (see *methods* section) in order to minimize this potential strategy. Because of the large target-distractor overlap, the most tilted item in any given display may have originated from a distractor set and therefore an incorrect response would be obtained to the extent that observers used this information as a basis for their response. Observers may still use this strategy even if it is unreliable, however. If they had, we maintain that the results of Experiment 1 would have been consistent with an unlimited-capacity model. A later experiment in this paper tests the capacity limitations of processing the individual orientations unique to each cluster. Specifically, in Experiment 3, each cluster is represented by a single Gabor patch and the target patch was usually the most tilted item in the display. Observers could therefore exploit the tilt direction of individual

orientations in these displays and base their response on the local item with the greatest tilt. We find evidence of unlimited capacity, which suggests that this strategy was not used in Experiment 1 since processing was limited.

Although using large standard deviations discouraged responses on the basis of local orientations, it is possible that the evidence of limited-capacity processing we observed is caused by having to establish an average without enough information. It may have been too difficult to extract the mean from orientation distributions with large variances using only nine items (e.g., Dakin, 2001). Summary extraction for multiple sets might proceed in parallel, unlimited capacity had the variance been smaller or the number of items per set larger. Unfortunately, it would be difficult to rule the use of local orientation cues as a potential strategy in this case since both would unfold without interference.

The second strategy is that the overall difference in the pattern of orientations across the target and distractor clusters may automatically direct attention to the target (see *Figure 3.1*). The Gabors within each distractor cluster will be, on average, composed of items that are tilted both left and right while the Gabors within the target clusters will be composed of orientations tilted in the same direction. The detection of pattern discontinuities is also an unlimited-capacity process (e.g., Huang, Pashler, & Junge, 2004). We conclude that both of these potential strategies would be of more concern had the data been consistent with unlimited-capacity processing. Given that it was not, it suggests that observers did not use such strategies.

3.5.2.1 Discussion of similar work on this topic

Experiment 1 also shares similarities with Halberda et al. (2006) who used a pre-post cueing paradigm to test the number of sets that could be enumerated simultaneously without interference. Observers saw multiple subsets of dots and estimated the number of dots in the cued set. When the relevant set was cued before the stimulus array (*pre-cue*), observers could use this information to focus on a single set and ignore the irrelevant sets. In contrast, when the relevant set was cued after the array was presented (*post-cue*), successful performance required the enumeration of all of the sets. Equal performance in the pre- and post-cue conditions in this design suggests parallel unlimited processing of the relevant information. Indeed, in the Halberda et al. (2006) study, performance was not reliably different between the pre- and post-cue conditions when two subsets of dots required enumeration (see also Emmanouil & Treisman, 2008; Im & Chong, 2014; but see Poltoratski & Xu, 2013 who obtained a pre-cue advantage for two subsets). Thus, evidence using a pre-cue/post-cue method has led to the conclusion of "unlimited-capacity" for SSRs for multiple sets of items, whereas evidence from the simultaneous–sequential method has led to the conclusion that establishing multiple sets depends on limited-capacity processes (Experiment 1). We suggest that this difference reflects a difference in what "capacity" is referring to. Specifically, the conditions of the Halberda et al. study were such that performance was limited by storage capacity, rather than online capacity. That is, processing was constrained by the number of sets that could be maintained in memory rather than the degree to which processing could be engaged independently by multiple stimuli. Indeed, Poltoratski and Xu (2013) and Im and Chong

(2014) used a design similar to Halberda et al. and found that averaging performance is limited by, and cannot be separated from, visual working memory capacity. In contrast, the simultaneous–sequential method can be dissociated from storage capacity limits; if stimulus presentation conditions are such that performance is limited by how much information can be extracted from the display (e.g., because stimuli are presented briefly), then limited-capacity processing predicts a difference between simultaneous versus sequential even for one versus two items (i.e., less than the 3-4 item limit). Two versus four has been used in order to minimize contamination from differences in eye movements across conditions and to minimize contamination from sensory effects like crowding, but the logic is identical. Therefore we conclude that the apparent difference in results between the pre-post cueing paradigm and the simultaneous–sequential method likely arise from the different forms of capacity to which these methods measure.

3.6   Experiment 2: Control experiment 1

The conclusion that establishing SSRs of mean orientation is limited capacity relies on demonstrating that some other aspect of the task or design, unrelated to averaging, was not driving the observed advantage in the sequential condition. There are several potential factors to rule out, such as crowding of the Gabors within a set (Banno & Saiki, 2012; Bouma, 1970), low target-distractor discriminability across sets, and the involvement of limited-capacity comparison processes. To test the possibility that one or more of these factors was the cause of limited performance, we conducted a control

experiment in which the task required all of the same processes except for actually calculating mean orientation.

The task in Experiment 2 was identical to that in Experiment 1; report the direction of average tilt (left or right) in the cluster with the non-vertical mean orientation. The orientations of Gabors within each cluster, however, were identical and all were set to the mean of their respective cluster from Experiment 1 (*Figure 3.3*). Because the mean of each group was provided directly, there was no need to compute an average orientation to do the task.

Multiple alternative explanations of the limited-capacity processing result that was obtained in Experiment 1 were tested using this design. First, the explanation that the crowding of items within each cluster impaired mean estimations (Banno & Saiki, 2012) more so in the simultaneous condition than in the sequential conditions can be ruled out as driving the observed limitation in Experiment 1 because the stimulus spacing in Experiment 2 was the same as in Experiment 1. Therefore the extent of crowding that would occur in Experiment 2 is at least physically equal to, and may even be perceptually greater than (Kooi et al., 1994), the crowding that occurred in Experiment 1. Second, target-distractor discriminability of the means is the same in this experiment as Experiment 1 because the mean values were identical across the two experiments. Finally, this experiment requires the same number of comparisons across clusters as Experiment 1. Despite these common aspects, we observed evidence of unlimited-capacity processing in Experiment 2 and limited-capacity processing in Experiment 1, suggesting that the source of the limitation in Experiment 1 was the need to calculate the mean orientation for each of the groups.

53

(A) Simultaneous

(B) Sequential
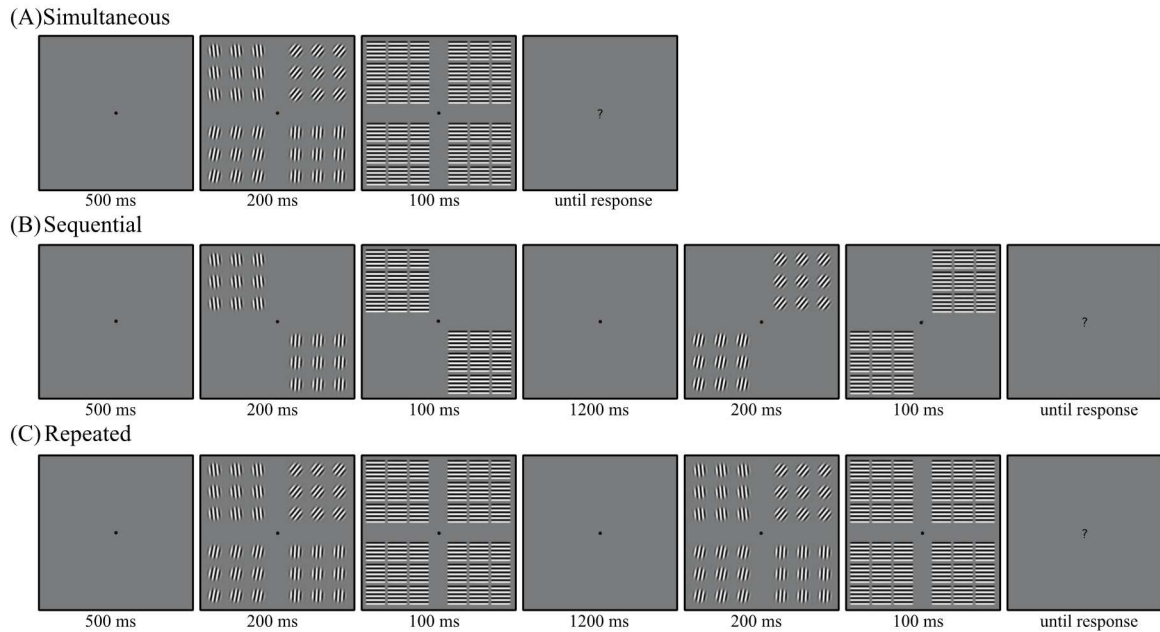
(C) Repeated

Figure 3.3. Trial events for the (A) simultaneous, (B) sequential, and (C) repeated conditions in Experiment 2. The mean orientation of each cluster was calculated after the orientations of Gabors within each cluster were sampled from their respective distributions. All Gabors within a given cluster was then adjusted according to that cluster's mean. Establishing summary representations are no longer necessary to perform the task. The target cluster is tilted right and presented in the upper right corner in this example.

3.6.2 Methods

All aspects of the method were identical to Experiment 1, with the exceptions noted below.

Twelve new undergraduate volunteers from the University of Iowa participated in exchange for course credit (2 male, 10 female, age range: 18 – 20 years, 11 right-handed).

The orientations of the Gabors within each of the four clusters were randomly chosen from the appropriate target or distractor distribution. The mean orientation for each cluster was then calculated and the orientations of all nine Gabors within a given

cluster were set to that cluster's mean prior to presentation (*Figure 3.3*). The orientations of the Gabors within each cluster were therefore identical.

As before, the default exposure duration for the simultaneous, sequential, and repeated conditions was 200 ms. The average adjusted exposure duration for all subjects after tracking remained at 200 ms.

3.6.3 Results and Discussion

Figure 3.4 shows the mean percent correct as a function of display collapsed across all observers. Equal performance between the simultaneous and sequential conditions was observed. There was also an advantage in the repeated condition. In contrast to Experiment 1, the pattern of data in Experiment 2 is consistent with an unlimited-capacity model and inconsistent with a limited-capacity model.

Arcsin transformed values were submitted to a one-way repeated-measures ANOVA with display as the within-subjects factor (all Kolmogorov-Smirnov $p > .907$; Mauchly's $p = .359$). The final model was significant, $F(2,22) = 17.76$, $p < .001$, $p\eta^2 = .618$, $MSE = .003$. As predicted by unlimited-capacity processing, accuracy was not reliably greater in the sequential condition (77% ± 1.11) than in the simultaneous condition (78% ± 1.13), $t(11) = 1.17$, $p = .269$. However, performance in the repeated condition (85% ± 0.92) was significantly higher than performance in the sequential condition, $t(11) = 4.82$, $p < .001$.
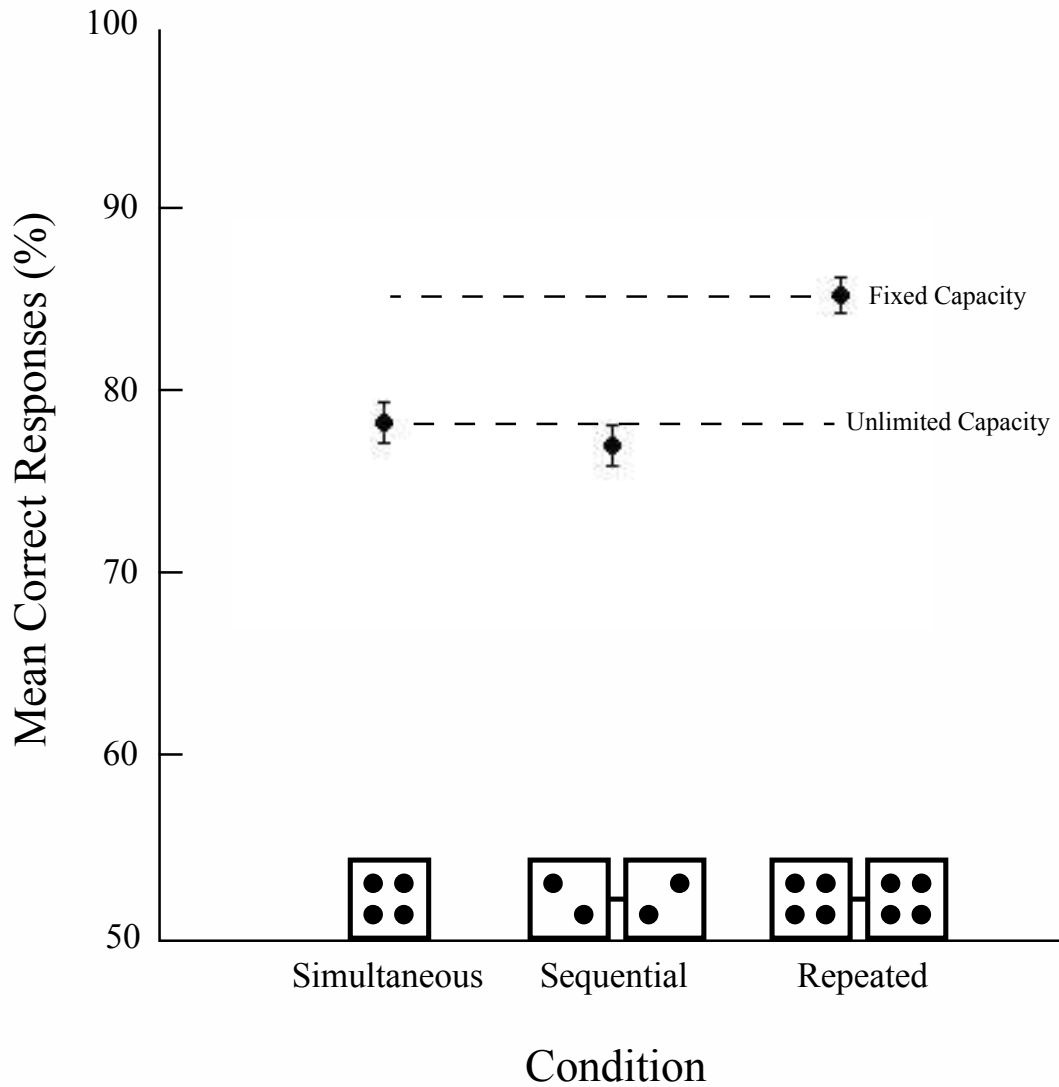
Figure 3.4. Mean correct responses (%) as a function of display collapsed across observers in Experiment 2. Performance was equal across the simultaneous and sequential conditions. There was also a reliable advantage in the repeated condition. Evidence consistent with unlimited-capacity processing was obtained when the task no longer required that subjects compute the average of each cluster. Error bars are within-subject standard errors (Cousineau, 2005; Moray, 2008).

We again compared performance within sequential trials when the target was

presented in the first frame versus the second frame. Performance across both frames

were statistically equal, 75% (first frame) vs. 79% (second frame), $F(1,11) = 2.55$, $p =$

.139, pη^2 = .188, *MSE* = .006 (all Kolmogorov-Smirnov *p* > .865). Targets presented closer in time to response were not remembered better.

Everything about Experiment 2 was the same as that of Experiment 1 except for the need to establish an SSR of mean orientation. Whereas Experiment 1 yielded evidence of limited-capacity processing, Experiment 2 yielded evidence of unlimited-capacity processing. We conclude that processing was limited in Experiment 1 specifically because it required the computation of mean orientation to do the task, and therefore that establishing SSRs of mean orientation involves limited-capacity processes.

3.7   Experiment 3: A second control experiment

In Experiment 2 the same orientation was repeated nine times within a given set. This redundancy may have had the unintended consequence of strengthening the represented average through probability summation. That is, it is possible that observers computed average orientations in Experiment 2, despite not having to do so in order to do the task. If they did, then the unlimited-capacity result might reflect an advantage for establishing SSRs on the basis of homogeneous sets compared to heterogeneous sets (Chong & Treisman, 2003; see also Utochkin & Tiurina, 2014), rather than reflecting them not doing the averaging process at all as we concluded. To test this possibility, we conducted a second control experiment in which a single Gabor patch was presented in lieu of the four 'clusters'. If the evidence of unlimited-capacity processing persists when we remove the repeating orientations, then we could rule out that the averaging of homogeneous sets was the sole cause of the results in Experiment 2.

3.7.2 Methods

All aspects of the method were identical to Experiment 2, with the exceptions noted below.

Twelve new undergraduate volunteers from the University of Iowa participated in exchange for course credit (1 male, 11 female, age range: 18 – 21 years, 11 right-handed).

The same displays presented in Experiment 2 were used except that only the center Gabor patch of each cluster was presented (*Figure 3.5*).
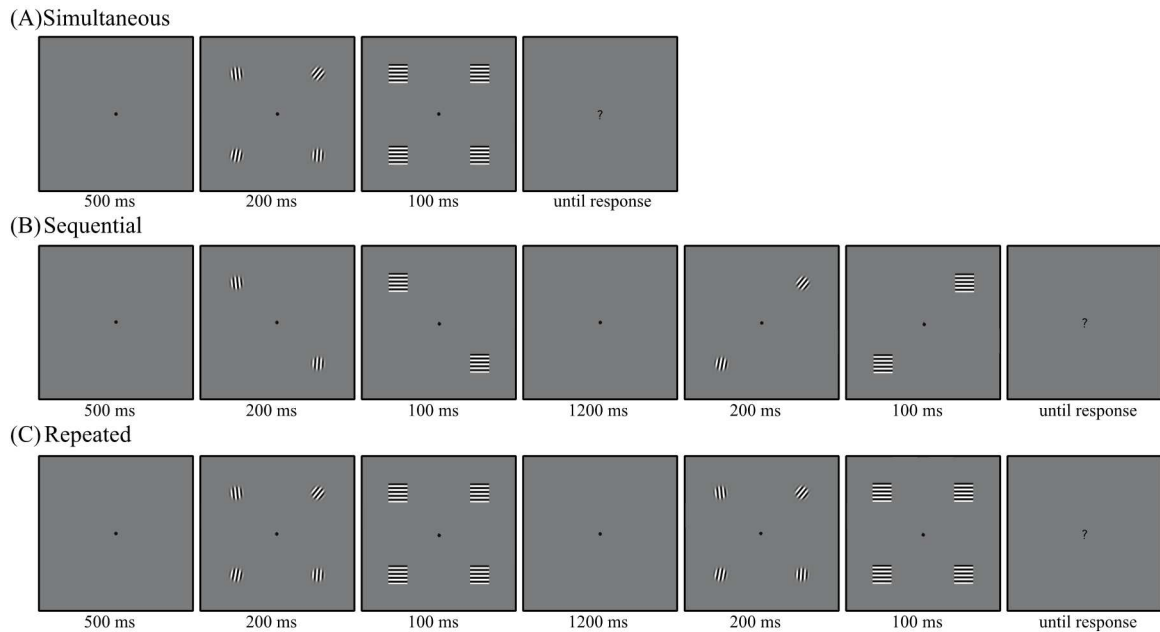
Figure 3.5. Trial events for the (A) simultaneous, (B) sequential, and (C) repeated conditions in Experiment 3. Observers were given the mean of each cluster, which was represented by the orientation of a single circle. The correct response is tilted right in this example.

As before, the default exposure duration for the simultaneous, sequential, and repeated conditions was 200 ms. The average adjusted exposure duration for all subjects after tracking was 180 ms.

3.7.3   Results and Discussion

Figure 3.6 shows the mean percent correct as a function of display collapsed across all observers. The data were again consistent with an unlimited-capacity model and inconsistent with a limited-capacity model.
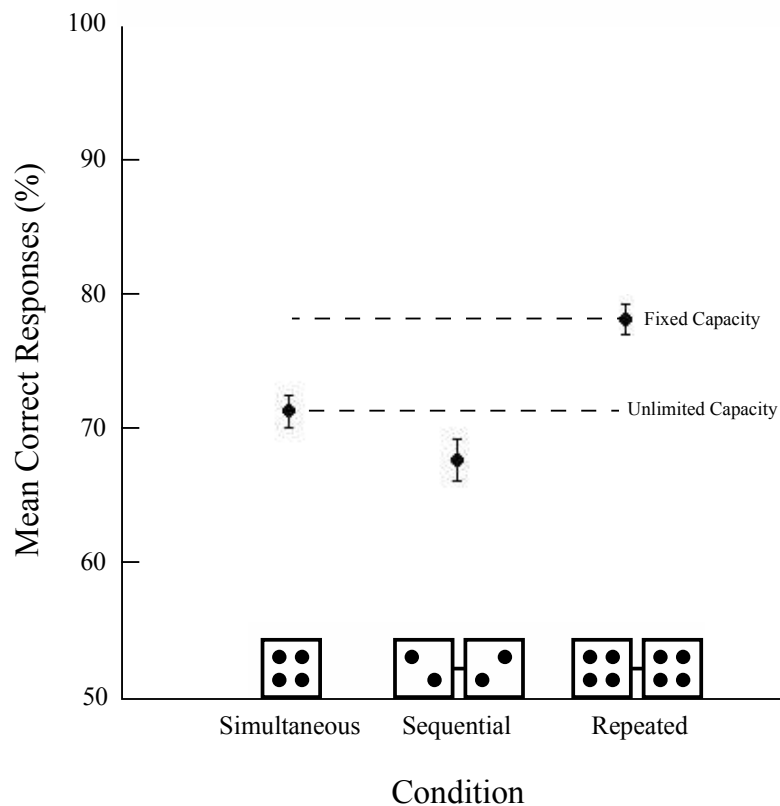
Figure 3.6. Mean correct responses (%) as a function of display collapsed across observers in Experiment 3. Performance was equal across the simultaneous and sequential conditions and there was also a reliable advantage in the repeated condition. These results are consistent with the unlimited-capacity model. Error bars are within-subject standard errors (Cousineau, 2005; Moray, 2008).

Arcsin transformed values were submitted to a one-way repeated-measures ANOVA with display as the within-subjects factor (all Kolmogorov-Smirnov $p > .408$; Mauchly's $p = .290$). The final model was significant, $F(2,22) = 18.06$, $p < .001$, $p\eta^2 = .621$, $MSE = .003$. As predicted by unlimited-capacity processing, accuracy was equal between the sequential ($68\% \pm 1.51$) and simultaneous ($71\% \pm 1.21$) conditions, $t(11) = 1.92$, $p = .081$. However, performance in the repeated condition ($78\% \pm 1.13$) was significantly higher than performance in the sequential condition, $t(11) = 5.65$, $p < .001$.

Performance within sequential trials when the target was presented in the first frame versus the second frame were statistically equal, 69% (first frame) vs. 66% (second frame), $F(1,11) = 1.12$, $p = .313$, $p\eta^2 = .092$, $MSE = .006$ (all Kolmogorov-Smirnov $p > .639$). There was no memory advantage for targets presented closer in time to response.

The results of this experiment provide further confidence in our original interpretation of the results of Experiment 1. That is, the evidence of limited-capacity processing found in that experiment can be attributed to the need to establish SSRs of mean orientation. When the task was the same, except that no average had to be computed, the results indicated unlimited-capacity processing. This was true in this experiment in which only a single item was presented in each cluster, and hence no average was needed, and in Experiment 2 in which every item in the cluster had the same orientation, and hence in principle no average was needed. The results from these three experiments combined strongly suggest that it is the averaging process that depends on limited-capacity processes.

3.8 Experiment 4: Establishing a single summary of mean orientation

We now turn to the question of limited capacity with regard to what? Relatively few studies have made the distinction between establishing summary representations across multiple sets of stimuli versus establishing a single summary representation across multiple items within a single set (Halberda, Sires, & Feigenson, 2006; Poltoratski & Xu, 2013). The conclusion offered from the preceding experiments that establishing SSRs for mean orientation is limited capacity is in regard to multiple sets of multiple items. That is, the evidence so far indicates that people cannot simultaneously establish SSRs of mean orientation for multiple ensembles of stimuli without mutual interference. It is a separate question whether SSRs for multiple items within an ensemble can be established independently of the number of items within the ensemble. This is an important distinction to make because conclusions drawn from multi-set tasks (e.g., Banno & Saiki, 2012; Oriet & Brand, 2013) do not generalize to single-set tasks (e.g., Ariely, 2001; Robitaille & Harris, 2011). This may be because, as we recently showed for mean size (Attarha et al., 2014b), establishing SSRs for a given attribute may be limited with regard to multiple ensembles, but unlimited with regard to items within a single ensemble. We address this contrast with regard to orientation in Experiment 4.

3.8.2 Methods

All aspects of the method were identical to Experiment 1, with the exceptions noted below.

Twelve new undergraduate volunteers from the University of Iowa participated in exchange for course credit (0 male, 12 female, age range: 18 – 22 years, 10 right-handed).

To create a single cluster, the four clusters of Gabor patches from Experiment 1 were placed on an evenly-spaced grid centered at fixation (*Figure 3.7*). Each patch was separated horizontally and vertically by 2.33° center-to-center. The size of the whole display was 13.91° × 13.91°.
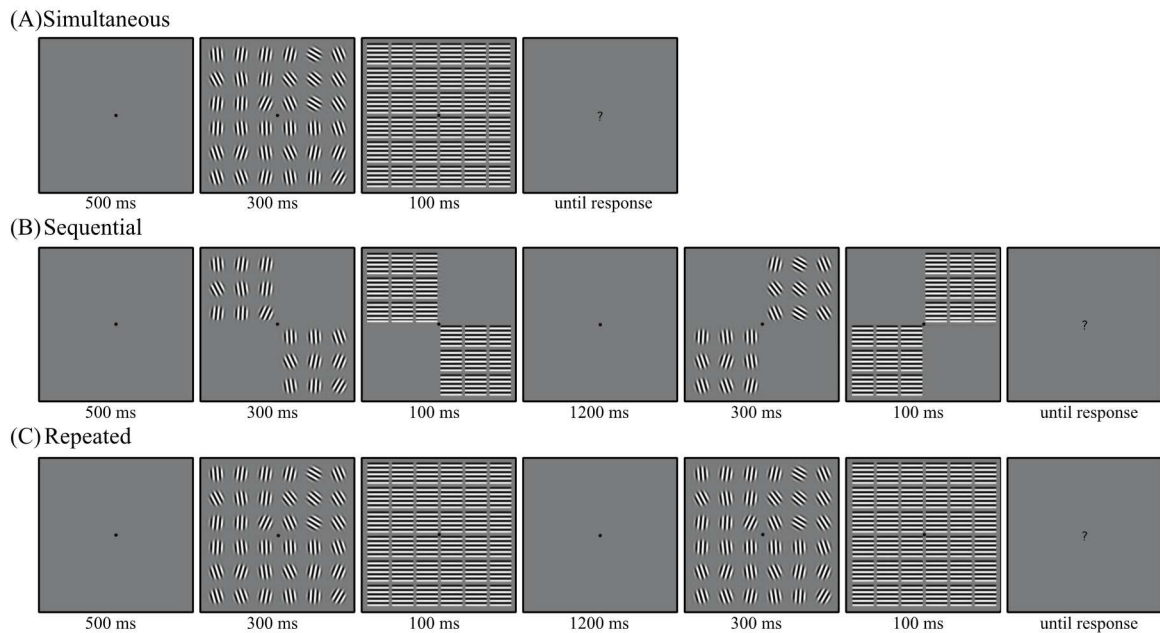
Figure 3.7. Trial events for the (A) simultaneous, (B) sequential, and (C) repeated conditions in Experiment 4. The four clusters from Experiment 1 were presented on an equally spaced grid to produce a single cluster with 36 items. Observers reported whether the mean orientation of the entire cluster was tilted left or right relative to vertical. The correct answer in this example is tilted left.

A pilot of this experiment demonstrated that subjects could not perform the task above chance-levels at a viewing duration of 200 ms. The default exposure duration for

the simultaneous, sequential, and repeated conditions was therefore set to 300 ms. The average adjusted exposure duration for all subjects was 310 ms.

The task was to report whether the average orientation over the entire set of thirty six items was tilted left ("F" key) or right ("J" key) relative to vertical.

3.8.3   Results and Discussion

Figure 3.8 shows the mean percent correct as a function of condition collapsed across observers. The data were consistent with an unlimited-capacity model and inconsistent with a limited-capacity model.

Arcsin transformed values were submitted to a one-way repeated-measures ANOVA with condition as the within-subjects factor (all Kolmogorov-Smirnov $p > .960$; Mauchly's $p = .086$, Greenhouse-Geisser epsilon = .721). The final model was significant, $F(1.44,15.85) = 9.43$, $p = .004$, $p\eta^2 = .462$, $MSE = .003$. As predicted by unlimited-capacity processing, accuracy was not reliably greater in the sequential condition (65% ± 1.71) than in the simultaneous condition (66% ± 1.00), $t(11) = 0.57$, $p = .582$. However, performance in the sequential condition was significantly lower than performance in the repeated condition (73% ± 1.21), $t(11) = 3.39$, $p = .006$.

Performance across both frames in the sequential condition were statistically equal, 65% (first frame) vs. 65% (second frame), $F(1,11) = 0.01$, $p = .937$, $p\eta^2 = .001$, $MSE = .006$ (all Kolmogorov-Smirnov $p > .687$), suggesting that targets presented first did not suffer from more memory loss than targets presented closer in time to response.
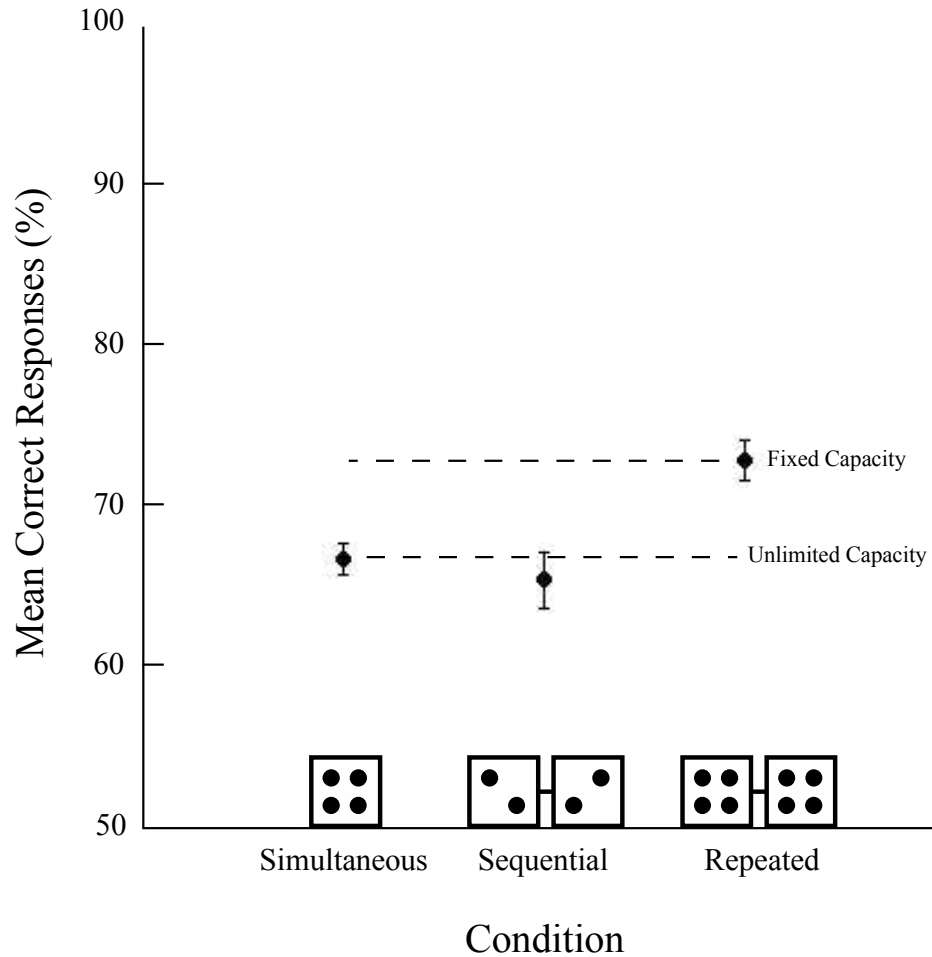
Figure 3.8. Mean correct responses (%) as a function of display collapsed across observers in Experiment 4. Evidence consistent with unlimited capacity was obtained when summary statistics were computed for a single set. Error bars are within-subject standard errors (Cousineau, 2005; Moray, 2008).

In summary, although establishing summary representations of mean orientation for multiple sets depended on limited-capacity processes (Experiment 1), the results of Experiment 4 indicate that establishing a single summary representation of mean orientation, across multiple items, can unfold entirely through unlimited-capacity processes. This finding is consistent with the results of Halberda et al. (2006) who found that the enumeration of a single summary proceeds without cost (see also Chong & Treisman, 2005b).

64

3.9   General Discussion

The visual system has been likened to a statistician that is capable of summarizing the features of similar items into efficient representations that guide behavior (e.g., Balas, Nakano, & Rosenholtz, 2010; Brady & Alvarez, 2011; Chong et al., 2008; Im & Chong, 2009; Joo et al., 2009; Rosenholtz, 2011; Rosenholtz et al., 2012). These representations are proposed to involve mechanisms that precede the limited bottleneck (Chong & Treisman, 2005a, p. 899; see also Alvarez, 2011; Chong & Treisman, 2003; 2005b; Oriet & Brand, 2013), which therefore implies that they are established through unlimited-capacity processes. We used the simultaneous–sequential method to test the capacity limitations of forming multiple SSRs of mean orientation, which is one of the main summaries for which the discussion of parallel processing is based. Performance was higher when fewer numbers of summaries had to be processed at a given time. The advantage for sequential over simultaneous presentation is consistent with a limited-capacity model and inconsistent with an unlimited-capacity model. Summaries of multiple ensembles may not be summarized independently, even for low-level features such as orientation. In contrast, when the same thirty-six items were grouped into a single cluster, the results were consistent with the opposite processing extreme, suggesting that averaging unfolds, without interference, regardless of the number of items that compose a single set (see also Halberda et al., 2006).

The same conclusion was reached in the case of mean size summaries. Attarha, Moore, and Vecera (2014b) used the simultaneous–sequential method and found that mean size summaries were highly limited in processing capacity. In that study, four sets

65

of discs with various diameters were randomly sampled from their corresponding target or distractor distributions. The task was to report whether the mean size of one of the sets was larger or smaller than the three remaining distractor sets. Performance in the sequential condition was better than the simultaneous condition and equal to performance in the repeated condition, suggesting that size summaries are mediated by a fixed-rate bottleneck.

To the extent that the two most studied summary representations – mean size and mean orientation – are not unlimited-capacity, it decreases confidence in the view that SSRs drive a global sense of visual completeness in the periphery. A coarse representation of summaries would need to be established in multiple regions of the visual field, rather than only a single region, in order to meet this function.

4.5   Overview

The simultaneous–sequential method was used to test the processing capacity of statistical summary representations both within and between feature dimensions. Sixteen gratings varied with regard to their size and orientation. In Experiment 1, the gratings were equally divided into four separate smaller sets, one of which with a mean size that was larger or smaller than the other three sets, and one of which with a mean orientation that was tilted more leftward or rightward. The task was to report the mean size and orientation of the oddball sets. This therefore required four summary representations for size and another four for orientation. The sets were presented at the same time in the simultaneous condition or across two temporal frames in the sequential condition. Experiment 1 showed evidence of a sequential advantage, suggesting that the system may be limited with respect to establishing multiple within-feature summaries. Experiment 2 eliminates the possibility that some aspect of the task, other than averaging, was contributing to this observed limitation. In Experiment 3, the same sixteen gratings appeared as one large superset, and therefore the task only required one summary representation for size and another one for orientation. Equal simultaneous-sequential performance indicated that between-feature summaries are capacity free. In summary, the results indicate that within-feature summaries involve limited-capacity processes but between-feature summaries do not. These findings challenge the view that within-feature summaries drive a global sense of visual continuity across areas of the peripheral visual

field, and suggest a shift in focus to seeking an understanding of how between-feature summaries in one area of the environment control behavior.

The experiments described in this chapter are under review: Attarha, M. & Moore, C.M. The perceptual processing capacity of summary statistics between and within feature dimensions. *Journal of Vision.*

4.6   Introduction

Summary statistics are proposed to "…precede the limited capacity bottleneck…" (Chong & Treisman, 2005a, p. 899; see also Alvarez, 2011; Alvarez & Oliva, 2008; Ariely, 2001; Brady & Alvarez, 2011; Chong & Treisman, 2003, 2005b; Dakin & Watt, 1997; Demeyere et al., 2008; Oriet & Brand, 2013; Rosenholtz, 2011; Robitaille & Harris, 2011). This view predicts that summary representations are established through unlimited-capacity processes, which is to say that they unfold independently (i.e., without interference) of the number of items to be processed.

Because collections of objects in the environment are most often comprised of combinations of multiple different feature properties, the view that summary representations play a critical role in abstracting the vast amount of information in the visual world depends partially on demonstrating that summaries can be established independently between different feature dimensions. Stated another way, accurate scene perception would suffer if behavior could only be guided by a single feature representation at any given time. Rather, it is necessary that the system establish all or most feature representations that define a particular collection of items.

Emmanouil and Treisman (2008) used a pre-post cueing paradigm to determine whether statistical averages could be generated for multiple dimensions without interference. Observers saw two sets of circles, separated on the left and right sides of the display. The circles varied in both size and the speed at which they moved. On each trial, observers were cued to perform one of two tasks: report which set (left or right) had the larger mean size or the larger mean speed. When the cued dimension was *pre-cued*, occurring prior to stimulus onset, observers could average over the relevant feature while the displays were present and ignore the non-cued feature. Performance was based on the statistical extraction of only one feature in this case. In contrast, when the cued dimension was *post-cued*, occurring after stimulus offset, observers had to average over both feature dimensions in order to successfully perform the task because they could not know which of the two they would have to report. According to the logic of this method, if statistical extraction for both dimensions unfolds in parallel without interference, then performance should be equal between the pre- and post-cue conditions. It should be possible to average two features just as well as one. Alternatively, if averaging one dimension interferes with averaging the other, then performance should be better in the pre-cue than the post-cue condition. The results were consistent with this latter alternative; performance was better when the to-be-reported dimension was pre-cued than when it was post-cued.

Although Emmanouil and Treisman's (2008) findings seem to indicate a cost for establishing summary representations for two different features simultaneously, it is possible that the cost derived from having to establish multiple summaries *within* a given dimension. This is because in the post-cued conditions, observers had to summarize size

for both sets and speed for both sets. Two within-dimension summaries were therefore required for the size task and another two for speed. It is unclear, therefore, whether the pre-cue advantage reflects limited within-dimension averaging, limited between-dimension averaging, or both. Indeed, using the pre-post cue method, Poltoratski and Xu (2013) found a performance decrement in the post-cue condition for two within-feature summaries, indicating that selection of the relevant set beforehand could improve performance (see also Brand, Oriet, & Tottenham, 2012). Results from our lab using a different method were also consistent with the view that forming multiple within-dimension summaries causes interference. Specifically, we used an extended version of the simultaneous-sequential method (Shiffrin & Gardner, 1972; Scharff et al., 2011) to test the perceptual processing capacity of mean size and mean orientation summaries. Those studies revealed that no more than a single summary could be established independently within either dimension (Attarha & Moore, in press; Attarha, Moore, & Vecera, 2014). We use that method in the current study.

Summary statistic representations may be hierarchically established within separate pathways of the visual system (Haberman & Whitney, 2009; Haberman & Whitney, 2011; Whitney et al., 2014). To paraphrase Whitney et al., (2014), summaries of basic visual features – such as brightness (Bauer, 2009) and orientation (Dakin, 2001) – may be generated by mechanisms in early visual stages that pool the output from various feature-selective cells (Suzuki, 2005; Whitney et al., 2014). On the other hand, more complex summaries that require the integration of multiple component feature populations – such as size (Ariely, 2001) and motion (Watamaniuk, Sekular, & Williams, 1989) – may be generated further along the ventral or dorsal pathways, or even after the

70

convergence of these streams as the case may be for summaries based on biological motion (Sweeny, Haroz, & Whitney, 2013).

Based on the hierarchical model of summary formation, any interference for establishing multiple between-feature summaries should be reduced to the extent that those summaries are generated in non-overlapping visual stages. We used the extended version of the simultaneous-sequential method to test this prediction for mean orientation and mean size in Experiment 3, after demonstrating significant costs in establishing multiple within-feature summaries in Experiment 1. To preview the results, we find that within-feature summaries engage limited-capacity processes and that between-feature summaries engage unlimited-capacity processes. These results are inconsistent with the conclusion that the visual system cannot generate summary representations for multiple different features without cost (Emmanouil and Treisman's, 2008).

4.7   Experiment 1: Within-feature summaries of mean orientation and size

4.7.2   Methods

A power analysis based on a pilot run of the experiment with 2 subjects indicated that at least four observers were needed to achieve at least 80% power (N*; Cohen, 1988). We increased this number to six in order to be consistent with the number of observers needed to satisfy the full sequence of counterbalanced conditions in subsequent experiments. All volunteers were from University of Iowa's psychology department (4 male, 2 female, age range: 18 – 31 years, 0 left-handed). The experiment was conducted

in accordance with the University of Iowa Internal Review Board (IRB) approved policies and procedures.

Stimuli were displayed on a cathode ray tube monitor (19-inch ViewSonic G90fB) controlled by a Macintosh Pro (Mac OS X) with a 512MB NVIDIA GeForce 8800 GT graphics card (1024 by 768 pixels, viewing distance of 61.5 cm, refresh rate of 100 Hz). Stimuli were generated using the Psychophysics Toolbox Version 3.0.11 (Brainard, 1997; Pelli, 1997) for MATLAB (Version 8.2, Mathworks, MA). Observers sat in a height-adjustable chair and used an adjustable chin rest to maintain a constant viewing distance from the monitor. The room was brightly lit to enhance visibility of the response keys.

Sixteen sinusoidal gratings that varied in both orientation and size were equally divided into four sets and presented on a neutral gray background (37.14 cd/m^2) (*Figure 4.1*). The gratings had a spatial frequency of 4 cycles and were presented at the maximum contrast that could be produced by the monitor (50.06 cd/m^2). On every trial, the orientations and diameters of items within each set were determined using independent sampling procedures. The orientations of the gratings within three randomly-selected sets were randomly chosen from a Gaussian distractor distribution ($\mu = 0°$; $\sigma = 8°$), while the orientations of items within the remaining target set were randomly chosen equally from either a Gaussian tilted-left distribution ($\mu = -15°$; $\sigma = 8°$) or a Gaussian tilted-right distribution ($\mu = 15°$; $\sigma = 8°$). Vertical was 0°. In addition, the diameters of gratings within three randomly-selected sets were randomly chosen from a Gaussian distractor distribution ($\mu = 1.86°$; $\sigma = 0.28°$), while the diameters of gratings within the remaining target set were equally chosen from either a Gaussian small-target distribution ($\mu = 1.40°$;

72

σ = 0.28°) or a Gaussian large-target distribution (μ = 2.33°; σ = 0.28°).

Each of the four sets were centered on a corner of an imaginary square approximately 5.59° from fixation. The center of the grating closest to fixation was 3.26° away, while the center of the grating furthest from fixation was 7.91° away. A distance of 7.91° separated the sets horizontally and vertically, center-to-center.
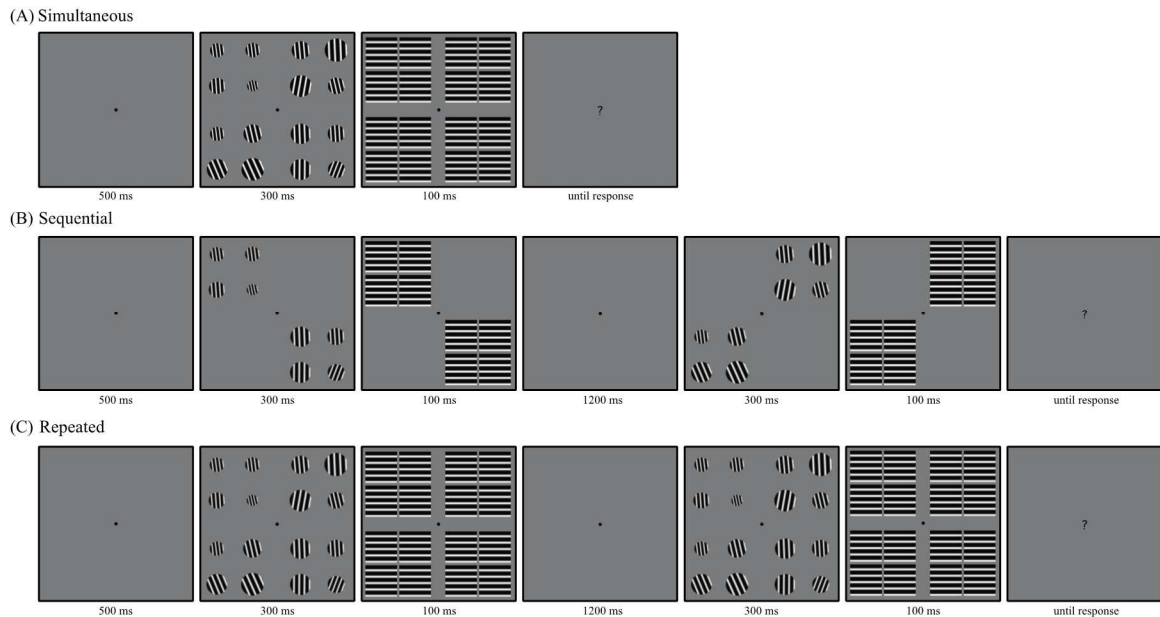
Figure 4.1. Trial events for the (A) simultaneous, (B) sequential, and (C) repeated conditions in Experiment 1. Observers saw four sets of gratings in which the items of each set varied in their orientation and size. In the case of orientation, the mean of the target set was tilted either left or right relative to the other three roughly-vertical distractor sets. In the case of size, the mean of the target set was either smaller or larger than the other three similarly-sized distractor sets. Observers were asked to establish a representation of the mean orientation and mean size for each set. Observers reported the tilt direction (left or right) and size (large or small) of the oddball sets. The correct response is "left and small" in this example.

Observers completed one 45-minute session. The session began with three practice blocks of 10 randomly-selected trials, each of which presented the critical displays at increasingly shorter exposure durations: 1000 ms, 300 ms, and 100 ms, respectively. The practice block was followed by 6 experimental blocks of 48 trials each

(96 observations per display type, 288 experimental observations per subject). Practice trials were excluded from all analyses.

All trials began with a black, centrally located fixation dot for 500 ms (3 cd/m²; 2 pixel diameter). Observers were instructed to maintain central fixation throughout the experiment. In the simultaneous condition, the fixation display was followed by the four sets of gratings (*Figure 4.1*). Each grating was subsequently masked by a square-shaped grating patch that was oriented horizontally at 90° for 100 ms (3.07° × 3.07°). A blank screen with a question mark ("?") at fixation followed the mask display and remained on the screen until a response was made (*Figure 4.1A*). In the sequential condition, fixation was followed by two sets of gratings presented along either the positive or negative diagonal, masks for 100 ms, a blank ISI of 1,200 ms, the other two sets of gratings presented along the opposite diagonal, masks again for 100 ms, and a blank screen with a question mark until response (*Figure 4.1B*). The repeated condition was the same as the sequential condition except that all four sets appeared in both of the two displays (*Figure 4.1C*). Written feedback ("correct" / "incorrect") was given at fixation following each response for 500 ms. The next trial automatically began 1,000 ms after the feedback display.

The initial exposure duration of the critical displays for the first block of the main experiment was set to the duration of the practice block that yielded above chance performance. The average initial duration across observers was 200 ms (see Attarha & Moore, in press; see also Whiting & Oriet, 2011). In addition, a coarse tracking procedure altered the exposure duration throughout the main experiment, block-by-block, on the basis of performance in the simultaneous condition only. If performance in the

74

simultaneous condition was within 10% of perfect performance on a given block, then the exposure duration for the simultaneous, sequential, and repeated conditions was decreased by 10 ms on the next block. Moreover, if performance was only 10% above chance (or lower) in the simultaneous condition, then the exposure duration in all three conditions increased by 10 ms. Chance performance was 25% in this 4 alternative forced-choice task. The average adjusted exposure duration across all subjects was 240 ms.

The full factorial combination of *display type* (simultaneous, sequential, repeated), *orientation target type* (left, right), and *size target type* (large, small) were randomly mixed within blocks of trials and appeared equally often. The target positions for the orientation and size target sets were sampled randomly from the following four possible positions: upper-left, upper-right, lower-left, lower-right. Which of the two diagonally opposite positions were presented first in the sequential display was constant for a given observer but varied across observers. Odd-numbered subjects saw sets of gratings that first appeared along the negative diagonal and then along the positive diagonal. Even-numbered subjects saw sets of gratings that appeared positive to negative. We kept the presentation of diagonal orders constant within an observer to eliminate uncertainty of the presentation positions.

Observers performed a dual task in which they reported the tilt direction (leftward or rightward) and size (larger or smaller) of the oddball sets. Observers pressed the "1", "4", "3", and "6" keys on the number pad of a standard keyboard using their index and middle fingers depending on whether the targets were "left and small", "left and large", "right and small", or "right and large", respectively. Observers were instructed to respond as accurately as possible. Speed was not emphasized.

75

We filtered the small percentage of trials in which the stimulus response led to an "incorrect" feedback message. In Experiments 1-2, this meant that either the mean orientation of a distractor set was, by chance, tilted either more leftward (or rightward) than the mean orientation of the target set, or that the mean size of a distractor set was smaller or larger than the target set. The set that appeared to be the target was in fact a distractor on these trials. A total of 10 and 18 out of 1,728 experimental trials across all six observers were filtered on this basis in Experiments 1 and 2, respectively. In Experiments 3A-C, we filtered trials in which (A) the mean orientation of the entire set of sixteen items was not tilted in the intended direction relative to vertical (B) the mean diameter of the entire set was not smaller or larger than the mean of the distractor distribution, or (C) either the mean orientation or mean size were the incorrect tilt or size. A total of 44, 51, and 91 out of 1,728 experimental trials across all six observers in Experiments 3A-C were filtered, respectively. The elimination of these trials did not change the pattern of results.

After filtering, the accuracy data were transformed to arcsin values to normalize their distributions and the underlying assumptions of the repeated-measures ANOVA were confirmed. Assumptions of normality and sphericity were confirmed using the Shapiro-Wilk test (Shapiro-Wilk, 1965) and Mauchly's test (Mauchly, 1940), respectively. When violations of sphericity were found, p-values were adjusted based on the Greenhouse-Geisser epsilon correction on degrees of freedom (Jennings & Wood, 1976). Two follow-up paired t-tests, one between the simultaneous and sequential conditions, and another between the sequential and repeated conditions, were used after significance of the final model was verified. An alpha level of .05 was used to determine

significance for all statistical tests.

4.7.3 Results and discussion
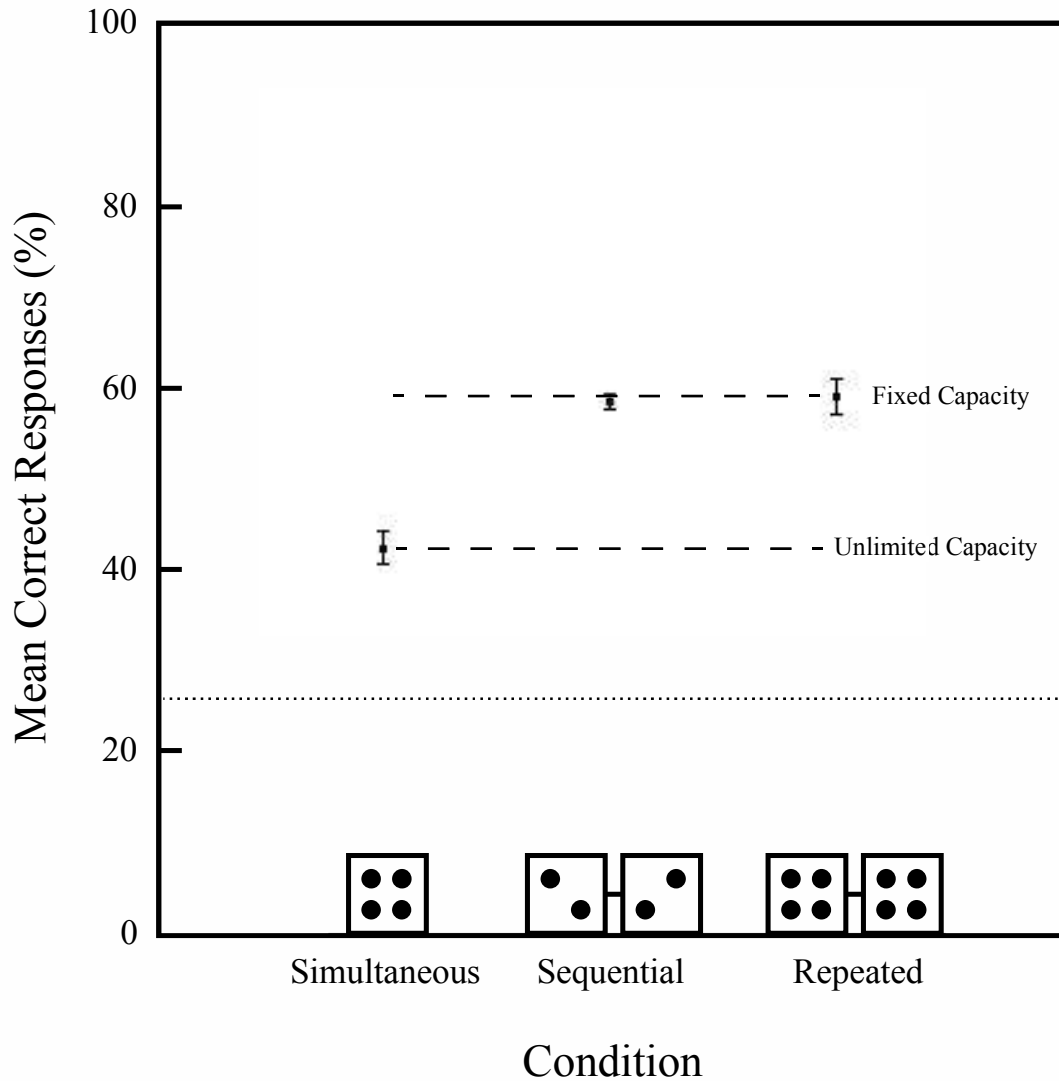
Figure 4.2. Mean correct responses (%) as a function of display collapsed across observers in Experiment 1. Consistent with the fixed-capacity model, performance in the sequential condition was better than performance in the simultaneous condition and equal to performance in the repeated condition. These results suggest that generating summaries for two features is mediated by a fixed-rate bottleneck if those summaries appear in different sets. Error bars are within-subject standard errors (Cousineau, 2005; Moray, 2008). The dotted line indicates chance performance.

Figure 4.2 shows mean percent correct as a function of display, collapsed across all observers. Error bars are within-subject standard errors (Cousineau, 2005; Moray, 2008). Dashed lines in *Figure 4.2* define both the unlimited- and fixed-capacity predictions. According to the logic of the simultaneous–sequential method, if performance in the sequential condition falls on the line determined by the simultaneous condition, then the unlimited-capacity model is supported. In contrast, if performance in the sequential condition falls on the line determined by the repeated condition, then the fixed-capacity model is supported. More formal accounts of specific versions of these models are offered in a previous paper (Scharff et al., 2011, Appendix).

In Experiment 1, we found that performance in the sequential condition was statistically equal to that in the repeated condition and that performance was reliably worse in the simultaneous condition. This pattern of results is inconsistent with an unlimited-capacity model and consistent with a fixed-capacity model. Arcsin transformed values of mean percent correct were submitted to a one-way repeated-measures ANOVA with display type as the within-subjects factor. The final model was significant, $F(2, 10)$ = 32.05, $p < .001$, $p\eta^2 = .865$, $MSE = .002$ (all Shapiro-Wilk $p > .089$; Mauchly's $p = .234$). As predicted by fixed-capacity processing, performance in the sequential condition ($58\% \pm 0.83$) was significantly greater than performance in the simultaneous condition ($42\% \pm 1.74$), $t(5) = 9.85$, $p < .001$. Performance between the repeated ($59\% \pm 2.03$) and sequential conditions were equal, $t(5) = 0.37$, $p = .727$. We conclude that establishing multiple within-feature summaries undergoes quite a bit of interference and is therefore highly limited in processing capacity (see Scharff et al., 2011).

It is worthwhile to note that the simultaneous and sequential conditions differ in two critical ways. The first is with respect to the number of sets that must undergo statistical extraction at any given time (two or four in this particular experiment). However, a second difference between these conditions is how close in time the target appeared before observers were allowed to enter their response. In the simultaneous condition, the target always appeared in the frame immediately preceding response, while in the sequential condition, the target could appear in either the first or second frame. A memory disadvantage for first-frame targets in the sequential condition may have biased performance. We tested this possibility by comparing accuracy between both frames. Performance was statistically equal regardless of whether targets appeared first (58%) or second (62%), $t(5) = 0.70$, $p = .516$ (Shapiro-Wilk $p = .489$). These data suggest that targets presented further in time from response did not suffer from greater memory loss.

A potential limitation of the above results is that averaging performance decreases when set size is small and when the variance between items is large (Marchant et al., 2013; Robitaille & Harris, 2011). For example, Dakin (2001) showed that averaging thresholds increase when the standard deviation of the distribution was greater than 8, especially when the number of items per set was low. There may therefore be concern that Experiment 1 supports a fixed-capacity model only because means had to be extracted from only four, heterogeneous items. Under this view, evidence of unlimited capacity may have been obtained had each set been composed by a larger number of items. In previous work we responded to this concern by increasing the number of items per set from 4 to 9. Like the 4-item experiment, the 9-item experiment yielded evidence of limited-capacity processing (Attarha & Moore, 2014, Attarha & Moore, in press).

Assuming that 9 items/set is sufficient for statistical extraction, we conclude that using a larger set size or a smaller variance would not have eliminated the observed limitation in Experiment 1.

4.8   Experiment 2: Control experiment

In Experiment 1, we conclude that generating multiple summaries within two different feature dimensions, in this case mean orientation and size, produces significant interference. However, successful completion of the task in Experiment 1 required more than just statistical averaging. The involvement of other mechanisms with limited capacity, such as a limited comparison process, might have contributed to the observed advantage in the sequential condition. That is, the comparison of multiple summaries may undergo less interference in the sequential condition, when the representations from only two sets require comparison at any given time, than in the simultaneous condition, when all four sets require comparison at once. Under this view, statistical extraction itself would be unlimited but appear limited experimentally due to the need to compare sets. We tested this possibility in Experiment 2. Observers were required to perform the task from Experiment 1 but now without averaging. Specifically, all gratings within a given set were identical and reflected the mean orientation and mean size of their respective set. Observers could exploit this redundancy and compare individual gratings within each set in order to circumvent the averaging process. If the limited-capacity results from Experiment 1 are due to averaging and nothing else, then eliminating the need to generate

averages should support an unlimited-capacity model. This follows because all aspects of the task, including the number of comparisons between sets, remain the same.

4.8.2 Methods

All aspects of the method were identical to Experiment 1, with the exceptions noted below.

Six new undergraduate volunteers from the University of Iowa participated in exchange for course credit (3 male, 3 female, age range: 18 – 23 years, 1 left-handed).

After the orientations and sizes of the gratings within each of the four sets were randomly chosen from their appropriate target or distractor distributions, the means of both features were calculated and every item within a given set was adjusted to the mean of their respective set prior to presentation (*Figure 4.3*). The gratings within each set were therefore the exact same orientation and size.

As before, the exposure duration of the critical displays on block one of the main experiment was based on the duration of the practice block that yielded above chance performance. The average initial duration for all subjects was 300 ms. The average adjusted exposure duration after tracking was 310 ms.

Figure 4.3. Trial events for the (A) simultaneous, (B) sequential, and (C) repeated conditions in Experiment 3. The orientations and sizes of items within each set were adjusted according to the mean of their respective set and were therefore identical. Since the mean of the sets was provided directly, summary statistics are no longer necessary to perform the task. The task was otherwise the same to that of Experiment 1. The correct response is "right and small" in this example.

4.8.3 Results and discussion

Figure 4.4 shows the mean percent correct as a function of display collapsed across all observers. In contrast to Experiment 1, the pattern of results in Experiment 2 is consistent with an unlimited-capacity model and inconsistent with a fixed-capacity model.

Arcsin transformed values were submitted to a one-way repeated-measures ANOVA with display as the within-subjects factor. The final model was significant, $F(2,10) = 17.19$, $p = .001$, $p\eta^2 = .776$, $MSE = .002$ (all Shapiro-Wilk $p > .294$; Mauchly's $p = .134$). As predicted by unlimited-capacity processing, accuracy was not reliably

greater in the sequential condition (62% ± 1.88) than in the simultaneous condition (63% ± 1.88), $t(5) = 0.19$, $p = .861$. However, performance in the repeated condition (74% ± 0.77) was significantly higher than performance in the sequential condition, $t(5) = 6.54$, $p = .001$.
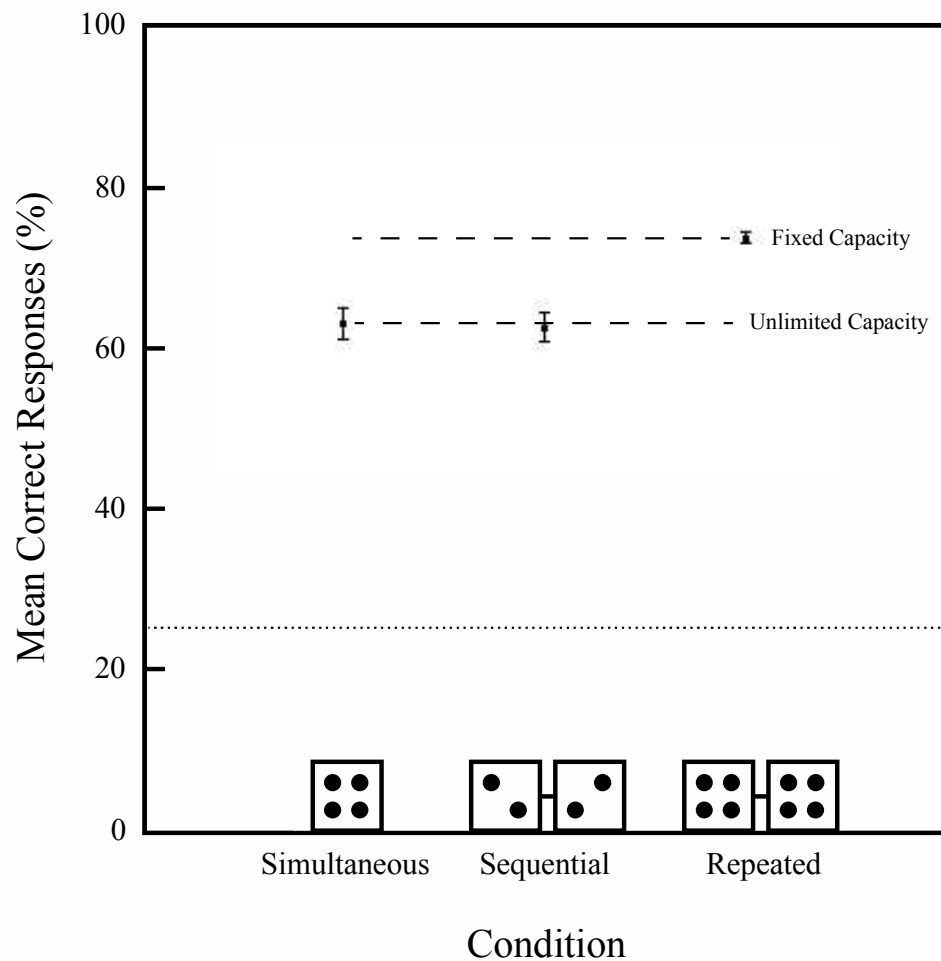
Figure 4.4. Mean correct responses (%) as a function of display collapsed across observers in Experiment 3. Consistent with the unlimited-capacity model, performance in the sequential condition was equal to the simultaneous condition and reliably worse than performance in the repeated condition. Error bars are within-subject standard errors (Cousineau, 2005; Moray, 2008). The dotted line indicates chance performance.

We again compared performance within sequential trials for when the target was presented in the first frame versus the second frame. Performance across both frames were statistically equal, 67% (first frame) vs. 66% (second frame); $t(5) = 0.28$, $p = .794$ (Shapiro-Wilk $p = .153$). Targets presented closer in time to response were not remembered better.

The critical difference between Experiments 1 and 2 is whether the task can be performed with or without the generation of summary representations. We find evidence of fixed-capacity processing when summaries are required and evidence of unlimited-capacity processing when summaries are not. These results increase confidence in the conclusion that limited averaging processes produced the observed capacity limitation in Experiment 1.

In addition to ruling out the possibility that the pattern of results in the first experiment was caused by a limited comparison process, Experiment 2 also controls for the influence of at least two other additional factors that may have limited performance. First, crowding influences judgments of mean computation (Banno & Saiki, 2012), and given the range of sizes and degree of separation between gratings, the items within a set violated Bouma's half-eccentricity principle (Bouma, 1970). It is possible that crowding was stronger in the simultaneous condition when all four sets appeared simultaneously than in the sequential condition when only two sets appeared at a given time. Fewer instances of inter-item crowding may have led to the observed advantage in the sequential condition. However, the stimulus spacing between Experiments 1 and 2 were identical, suggesting that the limitation in Experiment 1 would have been replicated in Experiment 2 had the sole cause been crowding. Second, it may be argued that the mean

difference between the target and distractor sets may have been too small a discrimination. Notice, though, that the mean values for the target and distractor sets were identical across both experiments and yet we obtained evidence of fixed-capacity processing in one case and unlimited-capacity processing in the other. We conclude that the source of the limitation in Experiment 1 was due to generating multiple mean representations across multiple sets.

4.9   Experiments 3A-C: Between-feature summaries of mean orientation and size

In previous studies, we have shown that capacity limitations differ with respect to whether summaries are generated over multiple sets of items, or over many items within a single set. It appears that multi-set summaries, which require multiple within-feature representations, undergo mutual interference whereas single-set summaries, no matter how large the set, unfold independently (Attarha & Moore, in press; Attarha, Moore, & Vecera, 2014; see also Poltoratski & Xu, 2013). These studies tested mean orientation and mean size summary statistics alone. In Experiments 3A-C of the current paper, we test whether the system can independently generate summaries between dimensions over items of a single set.

We used identical stimuli for Experiments A, B, and C, but we altered the task instructions for each experiment (see *Figure 4.5*). There were three tasks total. In the Report Orientation task, observers reported whether the mean orientation of the entire set was tilted left or right relative to vertical. In the Report Size task, observers reported whether the mean size of the entire set was larger or smaller that the size of a probe circle

that was set to the mean diameter of the distractor distribution. (The probe was presented only during practice trials and did not appear during the main experiment.) Finally, in the Report Orientation and Size task, observers reported both the tilt direction and size of the whole set.

If simultaneously forming a single summary of both orientation and size is a limiting factor of statistical extraction, then a limited-capacity model should be supported. This result would be consistent with the results of Emmanouil and Treisman (2008). However, unlike the task in Emmanouil and Treisman, the current task only requires a representation of one summary per dimension and therefore cannot be limited by having to establish multiple summaries within both dimensions. With this change, it is possible that we will find evidence of concurrent summary processing between dimensions. An unlimited-capacity model should be supported in this case.

4.9.2 Methods

All aspects of the method were identical to Experiment 1, with the exceptions noted below.

Six new volunteers from University of Iowa's psychology department participated in three sessions performed on separate days (3 male, 3 female, age range: 21 – 32 years, 0 left-handed).

The items from Experiment 1 were placed on an evenly-spaced grid centered at fixation (*Figure 4.5*). The gratings and masks were separated horizontally and vertically by 3.26° center-to-center. The size of the whole display was approximately 12° × 12°.
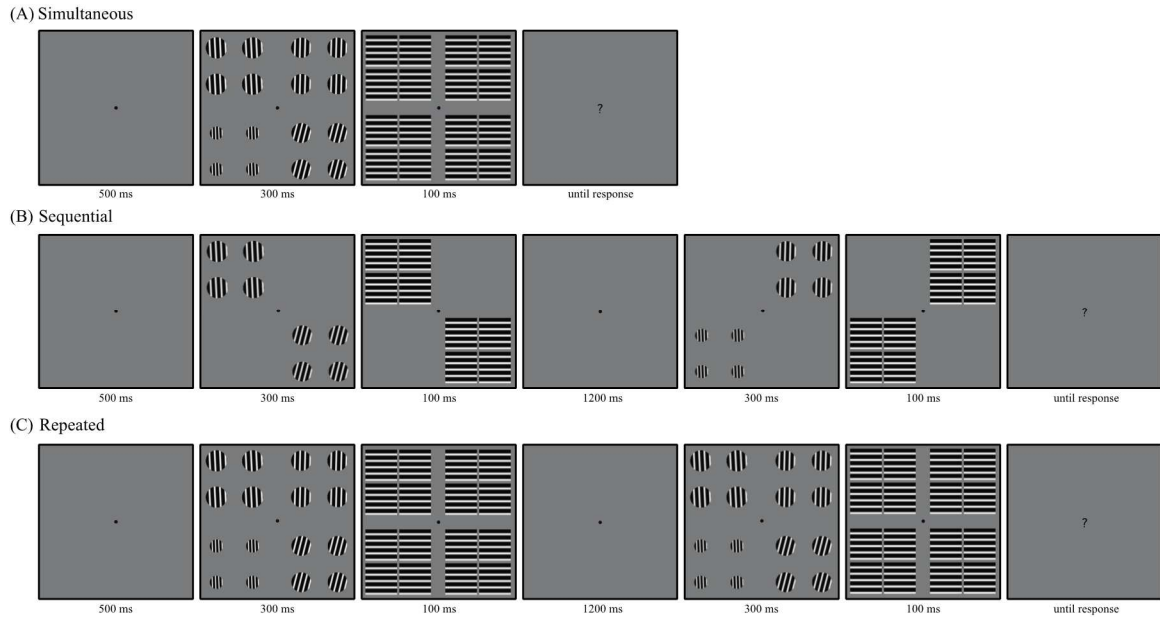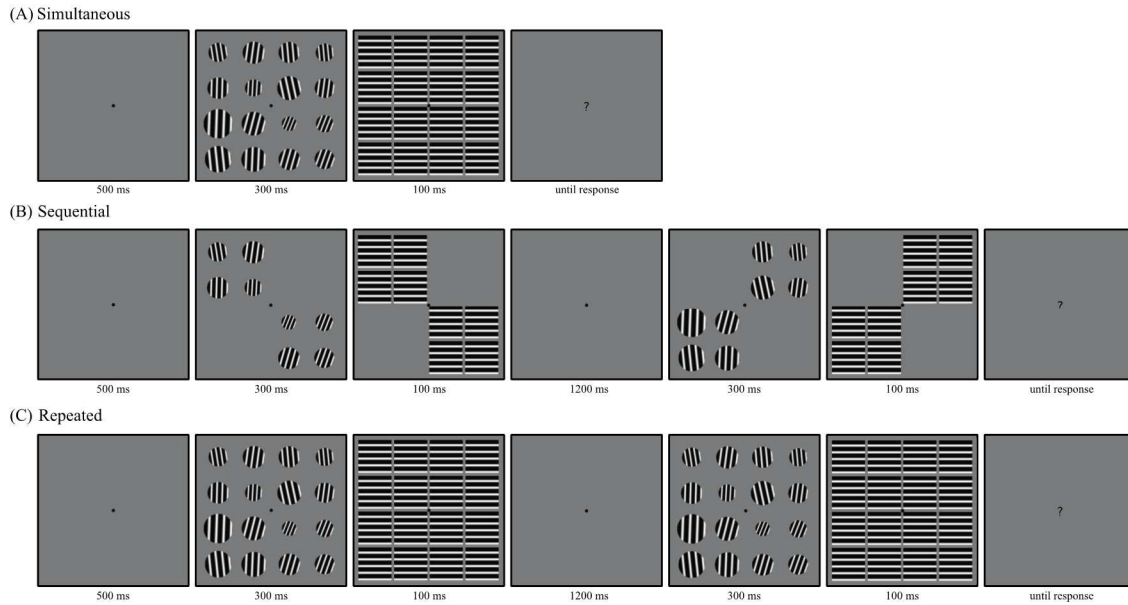
Figure 4.5. Trial events for the (A) simultaneous, (B) sequential, and (C) repeated conditions in Experiment 2. The items from Experiment 1 were re-spaced to produce a single set of 16 items. Observers participated in three experimental sessions using these displays, each of which had a different task. In the Report Orientation task, observers were told to ignore size and report whether the mean orientation of the entire set is tilted left or right relative to vertical. The correct answer is "right" in this example. In the Report Size task, observers ignored orientation and reported whether the mean size of the entire set was larger or smaller than the size of a probe circle (not shown) that was set to the mean diameter of the distractor distribution. The probe circle was only presented on practice trials. In the Report Orientation & Size task, observers reported both features. The correct answer is "right and large".

      The same observers participated in three experimental sessions, one for each of the following task types: Report Orientation, Report Size, and Report Orientation & Size. As before, each session began with three practice blocks of 10 trials each. A slight modification was made to all practice trials in which an estimation of mean size was required. After each of these trials, a black probe disc, adjusted to the mean diameter of the distractor distribution, appeared on the response screen at central fixation (3 cd/m^2; 1.86°). The probe disc was omitted from the main study to keep the trial events consistent across experiments. Each session lasted approximately 45-minutes and was performed on

separate days in complete counterbalanced order.

The average initial exposure durations for the Report Orientation and Report Size experimental sessions were both 100 ms while the initial duration for the Report Orientation & Size task was 200 ms. The average adjusted exposure durations for these sessions was 60 ms, 90 ms, and 230 ms, respectively.

In the Report Orientation session, observers determined whether the mean orientation over the entire set of sixteen items was tilted left ("1" key) or right ("6" key) from vertical. In the Report Size session, observers reported whether the mean diameter of the set was larger ("4" key) or smaller ("3" key) than the size of the probe circle that was presented on the practice trials. In the Report Orientation & Size session, observers reported both orientation and size using the same response-key mapping described in the task section of Experiment 1.

4.9.3 Results and discussion

Figure 4.6 shows the mean percent correct as a function of condition collapsed across observers. Across all three experimental sessions – report orientation (*Figure 4.6A*), report size (*Figure 4.6B*), and report both (*Figure 4.6C*) – the data were consistent with an unlimited-capacity model and inconsistent with a limited-capacity model.
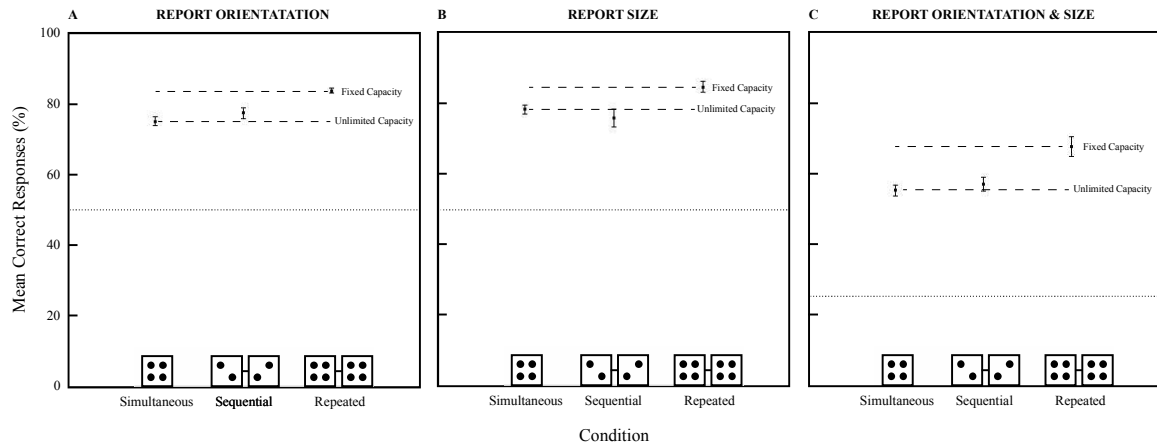
Figure 4.6. Mean correct responses (%) as a function of display collapsed across observers in Experiment 2A-C. Across all three task types -- report orientation, report size, report both orientation and size -- performance was equal across the simultaneous and sequential conditions and there was a reliable advantage in the repeated condition. These results are consistent with the unlimited-capacity model. Error bars are within-subject standard errors (Cousineau, 2005; Moray, 2008). Dotted lines indicate chance performance.

Arcsin transformed values were submitted to a one-way repeated-measures ANOVA with condition as the within-subjects factor. The final model was significant for all three task types (Orientation: $F(2,10) = 12.34$, $p = .002$, pη^2 = .712, *MSE* = .002, all Shapiro-Wilk $p > .053$, Mauchly's $p = .201$; Size: $F(2,10) = 6.38$, $p = .016$, pη^2 = .561, *MSE* = .003, all Shapiro-Wilk $p > .073$, Mauchly's $p = .188$; Both: $F(2,10) = 9.34$, $p = .005$, pη^2 = .651, *MSE* = .003, all Shapiro-Wilk $p > .220$, Mauchly's $p = .199$). As predicted by unlimited-capacity processing, accuracy was not reliably greater in the sequential condition (Orientation: 77% ± 1.50; Size: 75% ± 2.51; Both: 57% ± 1.95) than in the simultaneous condition (Orientation: 75% ± 1.34, $t(5) = 1.01$, $p = .359$; Size: 77% ± 1.37, $t(5) = 0.84$, $p = .440$; Both: 55% ± 1.53, $t(5) = 0.97$, $p = .376$). However, performance in the sequential condition was significantly lower than performance in the repeated condition (Orientation: 83% ± 0.65, $t(5) = 3.82$, $p = .012$; Size: 84% ± 1.60, $t(5) = 2.77$, $p = .039$; Both: 68% ± 2.82, $t(5) = 2.81$, $p = .037$). We conclude that the

89

establishment of multiple between-feature summary representations depends entirely on parallel, unlimited-capacity processes.

In order to test whether targets presented in the second frame of the sequential condition had an advantage over targets presented in the first frame, performance across both frames were compared for each of the three sessions. Performance across the sequential frames was statistically equal in both the Report Orientation task [71% (first frame) vs. 82% (second frame), $t(5)=2.20$, $p = .079$ (Shapiro-Wilk $p = .687$); see also Attarha & Moore, in press] and the Report Orientation and Size task [56% (first frame) vs. 61% (second frame), $t(5)=0.85$, $p = .435$ (Shapiro-Wilk $p = .700$)]. However, targets presented closer in time to response were remembered better than targets that appeared first in the Report Size task [70% (first frame) vs. 81% (second frame), $t(5)=2.79$, $p = .038$ (Shapiro-Wilk $p = .635$)]. This finding suggests that, in the case of mean size, memory differences may have contributed to lower performance in sequential condition (but see Attarha, Moore, & Vecera, 2014).

We'd like to now turn to discussing a few alternative explanations. In our displays, each set consisted of multiple items. Our goal from the outset was the ensure that observers were establishing a representation of the mean that incorporated all (or most) of these items rather than engaging in an alternative strategy in which they simply based their response on information contained within the most distinct local item. To this end, we used distributions with a large degree of overlap. The target and distractor distributions had a mean separation of 15° and a standard deviation of 8°. As a result, the most distinct item on any given trial may have originated from a distractor set, rather than a target set. Observers would thus obtain an incorrect response if their response were

90

based on the identity of the outlier. This would render a strategy based on individual items, rather than on the set of items, unreliable. Furthermore, the results of Experiment 3 provide evidence against this account of the results. In this third experiment, the same 16 items from Experiment 1 were presented in a single set (instead of in four separate sets). The task was otherwise the same. If the observed limited-capacity results in Experiment 1 were caused by how efficiently observers could process individual items, then that limitation should persist in Experiment 3. This follows because the items – specifically the degree of target-distractor heterogeneity and the assumed local target item – are identical across both experiments. Instead, we find evidence consistent with an unlimited-capacity model.

Another alternative to the formation of summary representations, in the context of orientation-averaging task in particular, would be to use the overall difference in the pattern of orientations across sets to direct attention to the most likely target set. Over the course of the experiment, the items belonging to a distractor set will typically consist of items tilted to the left and right of vertical whereas the items composing the target set will typically slant in the same direction (see *Figure 4.1*). Observers may arrive at the correct answer by exploiting these pattern discontinuities. However, it is worthwhile to note that Huang, Pashler, and Junge (2004) have shown that this sort of pattern detection engages only unlimited-capacity processes. Given that we obtained evidence consistent with a fixed-capacity model – the opposite processing extreme reported by Huang et al. (2004) – we conclude that observers did not use this strategy in Experiment 1. In addition, performance levels in the size-only and orientation-only tasks were quite similar, even though such pattern discontinuities do not exist in the size task (Experiments A-B). This

91

finding increases confidence in the view that one task type did not benefit from some strategy that was unavailable in the other task.

Considering the issues mentioned above, it seems unlikely that the evidence of limited statistical extraction for multiple within-dimension summaries is attributed to pattern detection or to the limited processing of local items.

4.10 General discussion

It has been proposed that in order to provide a sense of visual completeness in the periphery, the visual system is equipped with specialized mechanisms that represent statistical properties of groups of like items (Ariely, 2001; Balas, Nakano, & Rosenholtz, 2010; Chong & Treisman, 2003; 2005a; 2005b; Im & Chong, 2009; Peterson & Beach, 1967; Pollard, 1984; Rosenholtz, 2011). These summary processes are thought to unfold across the visual field very early in the stream of visual processing via parallel, unlimited-capacity processes. Once established, these representations purportedly serve as a foundation for the operation of more complex processes.

Since it is a general rule that multiple different features define the objects available in the world, a useful summary representation would require that multiple between-feature summaries be established without limitation, at least for the features that define a particular collection of items. Emmanouil and Treisman (2008) found evidence against this hypothesis. In their study they reported a cost to averaging over two feature dimensions at the same time. In the current study, we suggest that their observed limitation was not due to the between-feature summary representations but rather to the

within-feature representations that were needed to perform their specific task. We used an extended version of the simultaneous-sequential method to reexamine the perceptual processing capacity of establishing multiple between- and within-feature summaries of mean orientation and size. The results indicate that multiple within-dimension summary representations are mediated through at least some limited-capacity processes (Experiment 1) whereas between-dimension summaries are mediated through only unlimited-capacity processes (Experiment 3). Notice that the stimuli across these two experiments were nearly identical and yet we obtained evidence of both processing extremes: maximally limited processing in the first experiment and maximally unlimited processing in the third. These findings contrast those reported by Emmanouil and Treisman (2008).

By demonstrating that the extraction of within-feature summaries involves limited-capacity processes and that between-feature summaries do not, we hope to, first, challenge the current dominant view that within-feature summaries drive a global sense of visual continuity in separate areas of the peripheral visual field, and to, second, encourage a shift in focus to understanding the functional role that between-feature summaries play in the control of behavior.

5.5   Overview

We used a correlated flankers task to test whether summary statistic representations require attention. Sets of oriented Gabor patches flanked a central target, and the task on each trial was to report the target's identity. The flankers were always to-be-ignored. Unbeknownst to observers, however, the mean orientation of the flanker sets was either positively or negatively correlated with the correct response on the unrelated central task. No flanker effect was observed when the correlation relied on establishing a mean of the flankers. In contrast, a reliable flanker effect was found using the same stimuli when the means were physically displayed. Observers did not need to generate summary-based representations in this latter case. These findings are inconsistent with the current dominant view that proposes that summary representations reflect an obligatory aspect of visual perception that occurs prior to the deployment of selective attention.

5.6   Introduction

Models of vision propose that a great deal of the information in the environment is processed *preattentively*, or prior to the deployment of selective attention (e.g., Bravo & Blake, 1990; Duncan & Humphreys, 1989; Treisman, 1985). For example, surface structure (e.g., Attarha, Moore, Scharff, & Palmer, 2014; Mattingley, Davis, & Driver, 1997) as well as perceptual groups defined on the basis of orientation, contrast, or

proximity, (e.g., Moore & Egeth, 1997; Rensink & Enns 1995) may serve as a first pass analysis of the scene, the output of which may be subsequently available for the allocation of attention and further processing (e.g., Julesz 1984; Treisman, 1988). For the last decade, *summary statistics*, or the ability to establish representations of the statistical regularities shared by multiple individual items (Alvarez, 2011), have been deemed as another preattentive process (Alvarez & Oliva, 2009; Chong & Treisman, 2003; Joo, Shin, Chong, & Blake, 2009). Summary statistics are proposed to operate after the system perceptually parses the scene into functional units and before attention is applied to items of behavioral relevance (Chong & Treisman, 2005a). Once established, summary representations may serve as the basis of more elaborate processing, essentially underlying phenomena that range from crowding to gist perception (Balas, Nakano, & Rosenholtz, 2010; Haberman & Whitney, 2011). The relationship of summary statistics to preattentive processing is therefore necessary to understanding the nature of the input to the sorts of processes that allow for efficient daily interaction.

If summary statistics are indeed analyzed without attention, then they should fulfill the various attributes that are used to define preattentive processing. Generally speaking, a preattentive process should unfold quickly, automatically, and in spatially parallel regions of the visual field without any limitations in processing capacity (e.g., Healy, Booth, & Enns, 1996; Neisser, 1967). The evidence for the fast speed of statistical processing shows that estimates of mean size for a group of circles can be reported in as little as 50 ms, and that such estimates do not become more precise with longer exposures (Chong & Treisman, 2003). Evidence consistent with the view that summary statistics unfold with parallel, unlimited-capacity comes from tasks that demonstrate equal

averaging performance between small and large sets (Ariely, 2001; Chong & Treisman, 2003; Chong & Treisman, 2005a), suggesting the possibility that the quality or number of sets that can be summarized does not depend on the number of sets presented (but see Attarha, Moore, & Vecera, 2014; Attarha & Moore, 2015).

Finally, evidence for statistical automaticity comes from studies showing that summary statistics are computed despite a remarkable inability to report the individual constituents used to establish the summary representation. In one of the most compelling demonstrations of automatic statistical integration, the orientation of a centrally located tilted grating, crowded out of explicit perception by a ring of flanking horizontal gratings, influenced the perceived orientation of the whole set (Parkes, Lund, Angelucci, Solomon, & Morgan, 2001). The authors concluded that averaging was "compulsory" (p. 739) because integration of the central and surrounding orientation information occurred even though observers could not individuate the orientation of the central patch. In another demonstration of automatic averaging, summary statistics appeared immune to top-down goals that were configured to avoid averaging certain items in the scene together. Specifically, Oriet and Brand (2013) used bilateral displays wherein interspersed sets of vertical and horizontal lines of varying length appeared on both halves of the screen. Observers reported which side of the screen (left or right) had the larger mean length for the relevant orientation (e.g., vertical). In the critical condition, the lines belonging to the ignored orientation (horizontal in this example) were adjusted in such a way that would bias length judgments for the attended orientation if involuntarily integrated. The authors found that information from the irrelevant set did in fact bias length judgments in a positive direction for the attended set, despite explicit instructions to ignore those items,

and despite a detriment to accuracy performance. The authors concluded that, "averaging appears to precede the deployment of selective attention" (p. 8).

Because the observed properties of summary statistics appear consistent with the attributes of preattentive processing, they are thought to unfold without attention. As a further test of this claim, Alvarez and Oliva (2008) used a multiple-object tracking task (Pylyshyn & Storm, 1988) in which subjects followed the spatial position of 4 of 8 items total. Therefore, 4 of the items were attended while the other 4 were not. The task was to identify the individual and centroid positions of both the attended and to-be-ignored items. Although subjects could not localize any of the four individual items that were to-be-ignored, they could nonetheless localize the centroid of multiple to-be-ignored items above chance levels. These studies have been taken as evidence that information can be abstracted and summarized even when attention is allocated elsewhere. The interpretation is that summary statistics unfold without attention (see also Joo, Shin, Chong, & Blake, 2009).

However, like those before us, we argue that several task guidelines must be met in order to determine whether summary statistics require attention (e.g., Miller, 1987). The task should ensure that (1) the to-be-attended stimuli do not involve a summarizing task, and that (2) the to-be-ignored stimuli need not be summarized for successful performance. Furthermore, the task should ensure that the to-be-attended and to-be-ignored stimuli are (3) spatially separated and (4) distinct in terms of their identities. Together, these guidelines minimize the probability that the irrelevant stimuli will attract attentive processing.

Any task failing to meet these guidelines would provide equivocal evidence regarding the attentional demands of summary statistics. For example, in Alvarez and Oliva (2008), observers were required to report the perceived center of the to-be-ignored items throughout the experimental session. Subjects would therefore need to devote at least some resources to these items in order to adequately perform the task, suggesting that they were not purely unattended (e.g., Mack et al., 1992; Mack & Rock, 1998). In other studies, such as Oriet and Brand (2013), the locations of the to-be-ignored items were spatially mixed with the locations of the attended items, perhaps causing a failure to select only the relevant information, or even producing judgment biases via perceptually based context effects (e.g., Im & Chong, 2009). As a consequence, irrelevant stimuli were certainly inside the focus of attention and could not be considered unattended. It is therefore possible that summary extraction was likely being tested under conditions of divided attention in these studies. On the basis of these task violations, we argue that it remains an open question as to whether summary representations are processed preattentively and we turn to another task, called the correlated flankers task, which fulfills the guidelines specified above.

5.6.1 Correlated flankers task and logic

The correlated flankers method can be used to test whether a given process unfolds without attention (Miller, 1987). In this method, observers perform a demanding central task in which they use specific responses to report the identity of a central target. To-be-ignored flankers appear on either side of the target, and unbeknownst to the observer,

certain flankers regularly co-occur with certain responses on the target task. The measure of interest is response time on *congruent* trials (wherein flankers are positively correlated with the correct response on the central task) relative to *incongruent* trials (flankers are positively correlated with the incorrect response). The logic of this method is that to-be-ignored stimuli must be processed automatically if they can bias performance on the demanding central task. Specifically, if the flankers are automatically processed without attention, then they should influence target processing and response time in the congruent condition should be faster than response time in the incongruent condition. Alternatively, if the flankers require attention, then no effect on target processing should be observed. Here, performance between the congruent and incongruent conditions should be equal.

We now revisit the aforementioned task guidelines on preattentive processing in the context of the correlated flankers task. In line with the first guideline, the central target task is unrelated to statistical averaging. Second, no aspects of the to-be-ignored flanker stimuli need to be reported. Instead, claims of attentional demands must be examined indirectly though the influence that to-be-ignored stimuli have on responses that require focused attention (i.e., processing of the flankers is measured indirectly through conditioned responses). Third, the targets and flankers are consistently mapped to spatially separate locations; the flankers are never presented at fixation and the target is never presented in the periphery. This maximizes the efficiency of attentional selection to the target location and minimizes the processing of the peripheral portions of the display. Fourth, the target and flanker stimuli do not overlap. They differ entirely in terms of their features and stimulus categories. This aspect of the design prevents the priming of flanker

stimuli through the processing of the target stimuli because the identities of the flankers and targets are distinct.

In the current study, we use the correlated flankers task to test whether summary representations of mean orientation are established preattentively. In Experiment 1, the flanking sets are composed of multiple Gabor patches with various orientations. Statistical averaging over the many different orientations is therefore necessary to establish the flanker-response correlation. In Experiment 2, the flanking sets reflect the mean after sampling. In this latter experiment, no averaging is necessary. The view that summaries are processed without attention predicts a significant correlated flanker effect in Experiment 1 because the effect relies on the ability of the observer to involuntarily generate a mean representation for seemingly task-irrelevant stimuli that appear in the periphery. In contrast, the view that summaries require attention predicts a significant flanker effect in Experiment 2 because the correlation is based on the displayed mean, rather than on an observer-generated mean.

5.7   Experiment 1: Correlated flankers of mean orientation summaries

5.7.2   Methods

24 undergraduate volunteers from the University of Iowa participated in exchange for course credit (15 female, age range: 19 – 22 years, 22 right-handed). This sample size was based on Mordkoff and Halterman (2008) who used target stimuli identical to that of the current study. All observers reported normal visual acuity and color vision.

Stimuli were displayed on a cathode ray tube monitor (19-inch ViewSonic G90fB) controlled by a Macintosh Pro (Mac OS X) with a 512MB NVIDIA GeForce 8800 GT graphics card (1024 by 768 pixels, viewing distance of 61.5 cm, refresh rate of 100 Hz). All stimuli appeared on a neutral gray background (37.14 cd/m^2) and were generated using the Psychophysics Toolbox Version 3.0.11 (Brainard, 1997; Pelli, 1997) for MATLAB (Version 8.2, Mathworks, MA). Observers sat in a height-adjustable chair and used an adjustable chin rest to maintain a constant viewing distance from the monitor.

The critical display on each trial contained a single target that appeared at fixation and a set of flankers that appeared on either side of the target (*Figure 1B, middle frame*). Targets were color-shape conjunctions defined by the colors red (34 cd/m^2) or green (28 cd/m^2) and the shapes square or diamond (2.39° × 2.39°). The combination of these colors and shapes produced four possible target conjunctions: red square, green diamond, red diamond, green square (*Figure 1A, top row*).
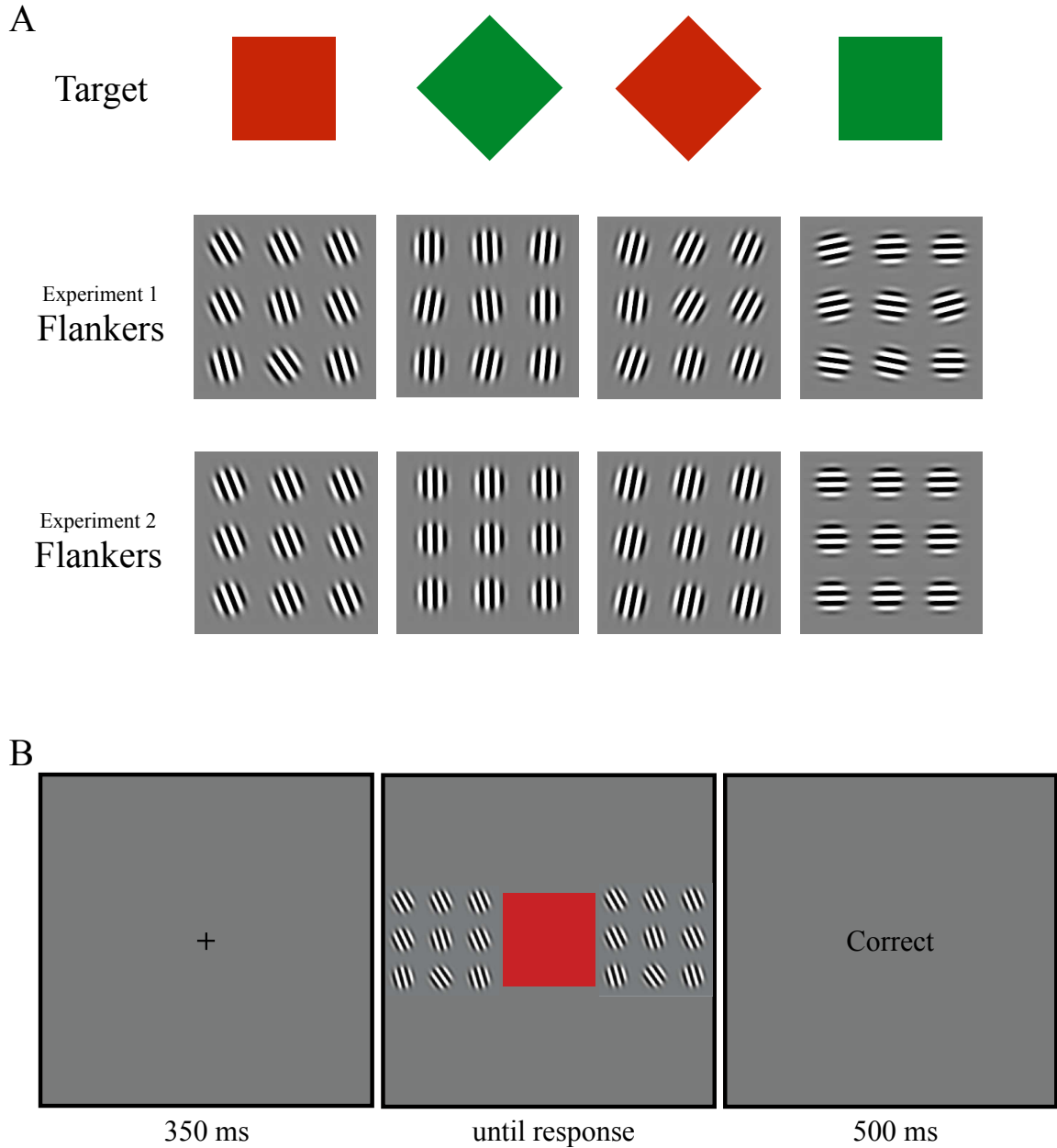
Figure 5.1. (A) The four possible targets (top row; red square, green diamond, red diamond, green square) as well as an example flanker from each of the four broad classes of orientations in Experiment 1 (middle row; tilted left, vertical, tilted right, horizontal) and Experiment 2 (bottom row). (B) Sequence of events for each trial. After fixation, a target item was flanked by identical sets of Gabor patches. Observers pressed the "F" key if the target was either a red square or a green diamond and the "J" key if the target was either a red diamond or a green square. The flankers were never mentioned. The correct response is "F" in this example.

Two targets were assigned to each of two response keys. On each trial, the task was to identify the central target by pressing the "F" key with the left index finger if the target was either a red square or a green diamond and the "J" key with the right index finger if the target was either a red diamond or a green square. Both speed and accuracy were emphasized.

A set of nine Gabor patches flanked the target (Gabor, 1946) (patch diameter: ~1.58°; set diameter: ~6.98° × 6.98°). Gabors had a spatial frequency of 3 cycles per degree and were windowed by a symmetric Gaussian envelope with a spatial constant of 7 pixels at maximum contrast (50.06 cd/m^2). The center of both sets was located 6.51° from the center of the target. A horizontal and vertical distance of 2.52° separated the Gabors, center-to-center.

The orientations within each flanker set varied in Experiment 1 (*Figure 1A, middle row*). On every trial, the orientations were randomly chosen from one of four Gaussian distributions that corresponded to the following general categories: tilted left (μ = -45°; σ = 15°), tilted right (μ = 45°; σ = 15°), horizontal (μ = 0°; σ = 15°), vertical (μ = 90°; σ = 15°). The flanker sets were identical on either side of the target. Similar to the four targets, the four flankers were also correlated with the "F" and "J" response keys. For example, for a given observer, *tilted left* and *vertical* might be positively correlated with "F" and *tilted right* and *horizontal* might be positively correlated with "J". The flankers were never mentioned in the task instructions.

Observers completed one 60-minute session of 20 total blocks. There were 34 trials per block: 2 randomly selected warm-up trials followed by the 32 planned trials (see *Table 1*). The session began with 4 practice blocks followed by 16 experimental blocks.

There were a total of 512 experimental observations per subject: 256 for the congruent and incongruent condition each. Only the data from the experimental blocks were retained for analysis.

Trials began with a centrally located black fixation cross (0.8° × 0.8°) for 350 ms, after which the target display was presented until a response was made. Written feedback ("correct" / "incorrect") was given at fixation following each response for 500 ms. The next trial automatically began 1,500 ms after feedback.

Upon finishing all trials, observers were asked to complete an unanticipated three-question survey (*Appendix A*). Question 1 was a recall measure that determined whether subjects noticed any correlation between the mean orientation of the flanking set and the available responses [yes or no]. A high proportion of "yes" responses would indicate that observers were explicitly aware of the flanker-response relationship (see Miller, 1987). Question 2 required a forced choice decision regarding the correlation (see Schmidt & Dark, 1998). If observers scored above chance, then it would suggest that an implicit association of this relationship had been established. Finally, Question 3 was a confidence measure in which observers used a Likert-type scale to rate the certainty of their forced-choice responses (1 = not at all sure; 2 = somewhat sure; 3 = very sure). High confidence was expected to the extent that observers believed that they were answering correctly.

We use the same design reported in Mordkoff and Halterman (2008) and describe those details here using the same terminology and design structure as that found in the original report (see also Miller, 1987). Half of the total number of trials was designated *inducing trials* while the other half was designated *test trials* (see *Table 1* for an

104

example). The purpose of inducing trials was to establish conditioned associations between certain flanking orientations and certain response keys. This was achieved by forcing some of the combinations between the four target types and the four flanker types to only ever appear together and other combinations to never appear at all. Test trials, in contrast, kept constant the combinations between the four flankers and four targets and were therefore retained for analysis. The purpose of test trials was to measure the effect of conditioned associations on response time by comparing trials in which certain flanking orientations were positively correlated with the correct response on the central task (*congruent trials*; denoted with an "+" in *table 1*) to trials in which flanking orientations were negatively correlated with the correct response on the central task (*incongruent trials*; denoted with an "-" in *table* 1).
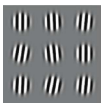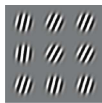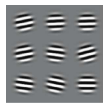
Number of trials per block with each flanker orientation

Target	Correct Response					Trial Type
🟥	F	4	4	0	0	Inducing
🔷	F	2+	2+	2−	2−	Test
🔻	J	2−	2−	2+	2+	Test
🟩	J	0	0	4	4	Inducing

Table 1. Experimental design. The four targets and four flankers appeared equally often, but the frequencies of the possible combinations were unbalanced. Inducing trials established a correlation between the flanking orientations and the response keys whereas test trials measured the influence of these correlations on performance. The "+" signs indicate trials in which the flanking orientation was positively associated with the correct response. The "-" signs indicate that the flanking orientation was negatively associated with the correct response. Which flanker orientations were associated with which response, as well as which targets and flankers defined inducing versus test trials, were randomly determined for each observer. The specific stimuli presented below serve as an example.

The full factorial combination of *target type* (red square, green diamond, red diamond, green square) and flanker-response *congruency* (congruent, incongruent) were randomly mixed within blocks of trials and appeared equally often. Which two categories of orientations were associated with which two response keys, and which two were designated to inducing trials or test trials, were constant for a given observer but varied across observers. The same was true in terms of whether targets for the inducing and test trials were defined on the basis of color or shape. Inducing and test trial types were split by color for odd-numbered subjects and by shape for even-numbered subjects.

Response time data were log-transformed and accuracy data were arcsin transformed to normalize their distributions. The underlying assumptions of all statistical tests were confirmed. Assumptions of normality and sphericity were confirmed using the Shapiro-Wilk test and Mauchly's test, respectively. When violations of sphericity were found, *p*-values were adjusted based on the Greenhouse-Geisser epsilon correction on degrees of freedom (Jennings & Wood, 1976). Assumptions of equal variances for between-experiment comparisons were confirmed using Levene's test. Paired *t*-tests between the congruent and incongruent conditions were used to determine significance.

5.7.3 Results and discussion

Figure 2A shows mean response time (ms) as a function of congruency for Experiment 1. Error bars are within-subject standard errors (Cousineau, 2005; Moray, 2008). No flanker effect was observed between the congruent (755 ms) and incongruent conditions (740 ms), [mean difference and standard error (SE): -15 ± 11 ms; $t(23) = 0.98$,

106

$p = .339$; Shapiro-Wilk $p = .923$]. Percent correct responses was high and statistically equal for the congruent (89.10%) and incongruent (89.49%) conditions, [0.39 ± 0.47 ms; $t(23) = 0.51$, $p = .615$; Shapiro-Wilk $p = .974$]. This pattern of results is inconsistent with the view that summaries are generated preattentively.
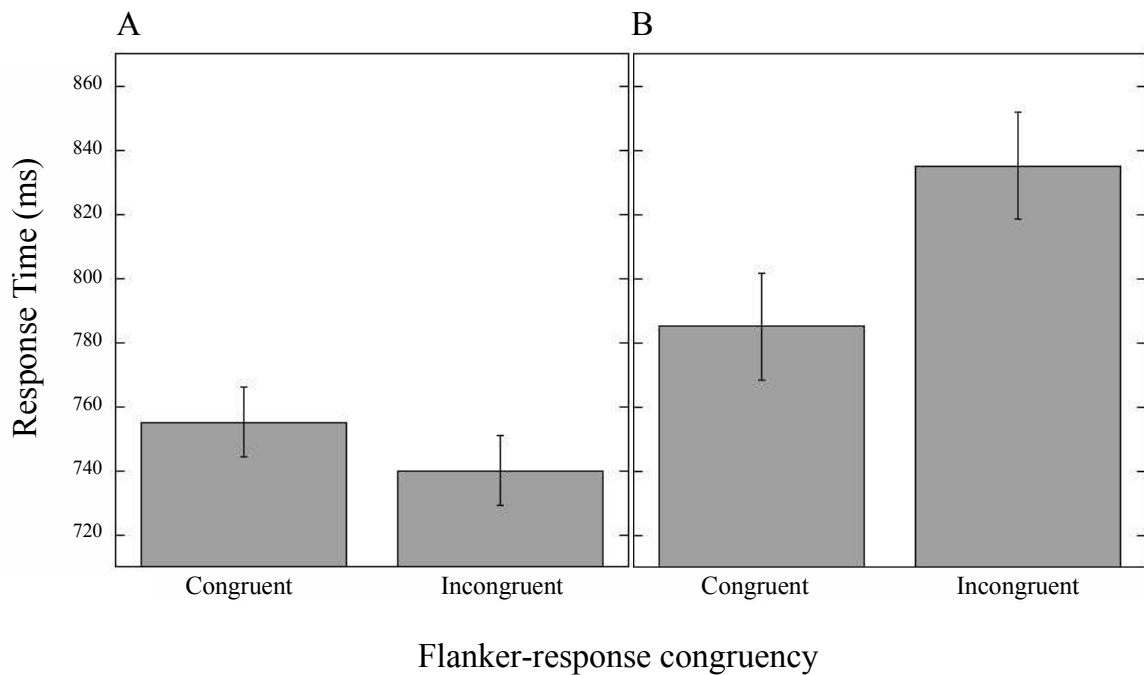
Figure 5.2. Mean response time (ms) as a function of the congruent and incongruent conditions collapsed across observers. (A) When averaging was required in Experiment 1, no flanker effect was observed suggesting that statistical averages may not be generated automatically. (B) In contrast, when the means of the flankers were provided in Experiment 2, a large flanker effect was found suggesting that such representations of the average could have sped performance in the congruent condition in Experiment 1 to the extent that they were established at all. Error bars are within-subject standard errors (Cousineau, 2005; Moray, 2008).

We now turn to the questionnaire. The proportion of "yes" responses to Question 1 was low (2/24 = 8%) suggesting that observers were not explicitly aware of the correlation between flanker orientation and response. When those who made yes responses were asked to describe what they noticed regarding the flankers, one said, "they changed direction", and the other said, "they're changing". Accuracy on the forced-

choice question was no better than what would be expected by 50% chance (mean and SEM: 45.83 ± 7.32; $t(23) = 0.57$, $p = .575$). These data indicate that observers were not implicitly aware of the flanker-response relationship. In addition, many observers reported to have guessed their answers to Question 2, which is supported by the low mean confidence rating of 1.16. We correlated accuracy on the forced-choice responses (Question 2 on the questionnaire) with confidence ratings (Question 3) and found a significant negative correlation. Observers who performed the worst were the most confident and observers who performed the best were the least confident (Spearman $r = -.431$, $p = .035$, Shapiro-Wilk $= p > .454$; $t(22) = 2.24$, SE $= .174$)

We conclude that the failure to obtain a correlated flanker effect of summary-based flankers on target processing suggests that summary representations cannot be established in the absence of attention. These results challenge the view that statistical extraction unfolds preattentively.

5.8   Experiment 2: Control experiment

Notice that the flankers in the current study were chosen from four broad orientation categories. Throughout the course of the experiment, the orientations for the tilted right category, for example, may have been centered on 50° for one trial and 40° the next. The specific orientation displayed for a given category was never constant and thus the effect depended on the association between response and a range of oriented values. The conclusion that summary statistics require attention, therefore, relies on the

108

assumption that variably mapped flankers can establish conditioned responses. We tested this assumption in Experiment 2.

The task in Experiment 2 was identical to that of Experiment 1. Observers reported the identity of the central target. However, the mean orientation of the flanking set was provided directly after the orientations were sampled from their respective distribution (*Figure 1A, bottom row*). The displayed mean for any given category therefore changed throughout the experiment just as the represented mean would have in Experiment 1. Because the mean of the flanking set was physically displayed for observers, the task could be performed without generating a summary representation of the average.

5.8.2   Methods

All aspects of the method were identical to Experiment 1, with the exceptions noted below.

A new group of 24 undergraduate volunteers from the University of Iowa participated in exchange for course credit (14 female, age range: 18 – 23 years, 23 right-handed).

As in Experiment 1, the orientations of Gabors within the flanking sets were randomly chosen from one of four Gaussian distributions trial-by-trial. The difference was that the orientations of the displayed patches were adjusted to the mean of the sampled orientations (*Figure 1A, bottom row*). The orientations of the Gabors within the flanking set were therefore identical.

109

Procedure. The four flankers shown in Question 2 of the follow-up survey (*Appendix A*) were replaced with pictures of the flankers from *Figure 1A, bottom row*.

5.8.3 Results and discussion

Figure 2B shows mean response time (ms) as a function of congruency for Experiment 2. Error bars are within-subject standard errors (Cousineau, 2005; Moray, 2008). A reliable flanker effect was found between the congruent (785 ms) and incongruent conditions (835 ms), [mean difference and SE: 50 ± 16 ms; $t(23) = 2.37$, $p = .027$; Shapiro-Wilk $p = .826$]. There was no evidence of a speed accuracy tradeoff as percent correct responses were high and statistically equal for the congruent (87.63%) and incongruent (88.28%) conditions, [0.65 ± 0.53 ms; $t(23) = 1.15$, $p = .258$; Shapiro-Wilk $p = .981$].

We also compared mean response times between the congruent and incongruent conditions in Experiments 1 and 2 using a mixed ANOVA with congruency as the within-subjects factor and experiment as the between-subjects factor. A significant interaction was found, [$F(1, 46) = 5.34$, $p = .025$, $p\eta^2 = .104$; Shapiro-Wilk $p = .800$; Levene's test of equality of variances, $F(1, 46) = 1.14$, $p = .292$]. This reliable interaction indicates that we obtained a different pattern of results between our two experiments.

Questionnaire. Survey responses in Experiment 2 were similar to that of Experiment 1. As before, the proportion of "yes" responses to Question 1 was low (3/24 = 13%). The three observers who reported a yes response made the following statements regarding the flankers: "It affects it but I don't know how", "They moved", and "They

110

changed with different shapes but I don't know the specifics". Accuracy on the forced-choice question was no better than 50% chance [mean and SEM: 47.92 ± 6.37; $t(23) = 0.33$, $p = .747$] and mean confidence was reported at a low rating of 1.46. Accuracy on the forced-choice question and responses on the confidence scale were not correlated, ruling out the possibility that those who performed better were more confident (Spearman $r = .300$, $p = .155$, Shapiro-Wilk = $p > .656$; $t(22)=1.48$; SE = .194) .

Finding no evidence of memory for the flanker information is a bit ambiguous. It is possible that attention was not allocated to flanker processing even though an indirect effect on task performance was observed. However, it seems more likely that either this information failed to be encoded into memory for subsequent report or that other factors interfered with access to the established memory.

Experiments 1 and 2 were virtually identical except for the need to generate summaries of mean orientation. Experiment 1 showed that summary representations of the displayed orientations failed to produce a reliable flanker effect, whereas Experiment 2 showed that these same averages could be used to speed performance if they were provided directly. The explanation that variably mapped flankers caused the negative finding in Experiment 1 can be ruled out since the same variability occurred in Experiment 2. Taken together, the results from these experiments are inconsistent with the view that summaries of mean orientation occur preattentively and consistent with the view that summaries require attention.

We would like to acknowledge the possibility that the difference in results between Experiments 1 and 2 may be accounted for by the Rescorla-Wagner model without need to appeal to statistical summaries (Rescorla & Wagner, 1972). According to

111

this model, changes in conditioning strength are large at the beginning of the experimental session and then taper off with each new trial until learning is complete. The associative strength of two conditioned stimuli – the target-response as one and the flanker-response as the other in the case of a correlated flankers task – depends in part on the relative salience and learning rate of each CS. It could be argued that the failure to find a flanker effect in Experiment 1 but not Experiment 2 is due to differences in variability of the flanker stimuli across these tasks. The flanker-response association using the heterogeneous flankers of Experiment 1 is less salient and more slowly learned than the target-response conditioning. This large difference in salience and learning rate between the CS stimuli may cause either blocking or overshadowing of the flanker CS by the target CS. One of the ways in which this might happen is that as the target CS quickly approaches the maximum level of associative strength, it causes the flanker CS to have either low or no associative value. A correlated flanker is not observed as a consequence. Supporting this view is evidence that variability weakens associative learning (e.g., Young & Wasserman, 1997). In contrast, there is a smaller discrepancy in terms of salience and learning rate for homogeneous flankers relative to the targets, allowing the flanker CS to be learned faster, and gain more associative strength, to ultimately produce a reliably flanker effect. While we cannot rule out this possibility, we can direct attention to the fact that heterogeneous flanker items have elicited correlated flanker effects in previous studies (e.g., find that citation).

A related concern is that the flanker-response relationship for variable items takes longer to learn than homogenous items. There may appear to be no flanker effect when the data from multiple blocks are collapsed, when in fact one exists later in the session.

112

To address this concern, we analyze performance in the beginning (*first* 6 blocks), middle (*middle* 6 blocks), and end portion of the experiment (*last* 6 blocks). The correlated flanker effect (incongruent minus congruent performance) in Experiment 1 was 39 ms (first) -62 ms (middle) -6 ms (last). This pattern of results suggests that a flanker effect did not emerge at the end of the experiment. In contrast, a flanker effect appeared early on when the means were provided directly: 6 ms (first) 39 ms (middle) 119 ms (last).

5.9   General discussion

The view that summary representations play a fundamental role in early visual perception rests in part on whether these representations can be established prior to the deployment of selective attention. We argue that prior investigations of this question measured summary statistics under conditions of divided attention but concluded that the observed effects occurred in the complete absence of attention. The current study used a correlated flankers task that offered a number of advantages over previous methods. Experiment 1 demonstrated that observers could not generate mean summaries of the to-be-ignored flanker sets if those sets were composed of various orientations. However, Experiment 2 showed that these same stimuli could exert conditioned responses if the means of the flankers from the first experiment were physically displayed (therefore making summary representations an unnecessary part of the task). An end-of-experiment questionnaire demonstrated that observers established neither an explicit or implicit association of the flanker-response correlation. The data from these experiments suggest that the source of the negative finding in Experiment 1 was due to a failure of statistical

113

extraction. We conclude that summary statistics require attention. As a consequence, the claim that the function of summary representations is to provide a sense of visual completeness in unattended regions of the visual field should be reconsidered.

Appendix A. Questionnaire for Experiment 1. An identical questionnaire was used in Experiment 2 with the exception that the pictured flankers reflected the mean (see Figure 5.1A, bottom row).

<div align="center">Questionnaire</div>

1) Did you notice any relationship between the mean orientation of the flankers and the response keys?

   Yes (what did you notice?_____)
   No

2) If I told you that the mean orientations of the flankers were correlated with the response keys, then could you tell me which two orientations from the four depicted below were positively correlated with the "F" response and which two orientations were positively correlated with the "J" response?

   _____   _____   _____   _____

   Tilted Left   Tilted Right   Vertical   Horizontal
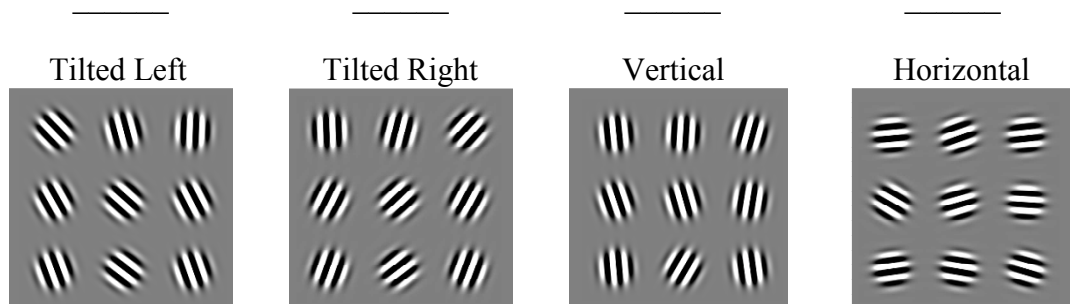
3) Rate your level of confidence in response to Question 2:

   1 = not at all sure
   2 = somewhat sure
   3 = very sure

CHAPTER 6: GENERAL DISCUSSION

6.1   Summary

The visual system has been likened to an intuitive statistician that effortlessly summarizes groups of similar items into efficient representations that guide behavior (e.g., Im & Chong, 2009; Balas, Nakano, & Rosenholtz, 2010). It is of widespread belief that the function of these summaries is first to act early in the stream of visual processing to reduce uncertainty in a cluttered world and second to serve as a foundational representation for which more complex processing can be based. Summaries are believed to play a central role in organizing the millions of bits of information received by the eyes in a way that leads to rapid visual scene perception and the subjective impression that we see more than we do. Therefore, it is thought that the inclusion of statistical summaries is necessary for theories of early visual processing as well for understanding how we develop unified, coherent visual percepts.

But from what evidence are these claims drawn? The claim that summary representations unfold across the visual field prior to any processing bottleneck are drawn from tasks that produce flat search slopes when set size is manipulated over items with a very high degree of item regularity. Interestingly, when item regularity is reduced, significant set size effects abound, which makes it difficult to determine whether processing occurs with or without interference. Furthermore, set size manipulations vary factors other than the number of items in the display, such as the amount of noise feeding into decision processes, the number of responses to be made, and the duration of any

given stimulus. The simultaneous-sequential method, in contrast, holds these factors constant. Another advantage that the simultaneous-sequential method has over set-size manipulations is that different models of processing interference have been formalized for ease of interpretation (Scharff et al., 2011a). Chapters 2-4 tested the most popular summary representations on which discussions of capacity limitations are based. The results indicate that while multiple within-feature summaries cannot be established without engaging limited-capacity processes, multiple between-feature summaries can. Specifically, establishing multiple sets is capacity limited while establishing multiple individual items and multiple feature dimensions is capacity unlimited. We conclude that the visual system cannot effortlessly generate multiple coarse representations of information across the entire peripheral visual field; a tradeoff exists between establishing a set of summary statistics in one region and establishing them in another.

Similarly, claims of preattentive statistical processing have been drawn from studies wherein the "unattended" stimuli could not possibly be ignored. Surely these studies are good examples of how summary statistics are processed under conditions of divided attention (and thus it is little surprise to find effects of summary statistics). Chapter 5 used a correlated flankers task in which statistical extraction was measured via conditioned responses. The advantage of this method is that it allows the unattended stimuli to remain ignored for the entire duration of the experimental session. No flanker effect was found when averaging was required. In contrast, a significant flanker effect was found when the means were provided directly. The conclusion is that summary representations require attention.

117

These findings challenge the view that the functional role of summaries is to reduce complex information across the visual field to support later processes and the sense of perceptual continuity in unattended areas. Summary representations do not bypass the limited aspects of our attentional and cognitive systems because they require attention and are limited themselves.

6.2 Discussion

6.2.1   Automaticity

Recent studies are contributing to the emerging picture that summary representations may not be such an early aspect of perceptual processing after all. Brown, Gore, and Carr (2002) outline several generally accepted criteria that a given process should meet in order to be considered automatic. First, the process in question should be insensitive to capacity demands. However, summaries do appear to be constrained by such demands. Summary performance is sensitive to input at stages beyond the initial registration of features such as object-substitution masking and visual working memory (Attarha, Moore, & Vecera, unpublished data; Jacoby, Kamke, & Mattingley, 2013; Myzczek & Simons, 2008; Poltoratski & Xu, 2013; see also Im & Chong, 2014). Additionally, Marchant, Simons, and de Fockert (2013) minimized the item homogeneity typically present in early reports of statistical averaging and found that performance decreased as the number of items within the set increased. These results indicate that there is a cost for sampling over larger sets. Finally, in the current study, we report

significant interference when generating multiple summaries within the same feature dimension, which means that summary representations must not be immune to interference from concurrent processing (Attarha & Moore, in press; Attarha, Moore, & Vecera, 2014; see also Brand, Oriet, & Tottenham, 2012).

Interpreting the results of the current study within the context of other studies using the simultaneous–sequential method also points to the possibility that SSR formation commences at later, rather than earlier, stages of visual processing. Specifically, the current application of the extended simultaneous-sequential method contributes to a developing picture regarding capacity limitations in perceptual processing more generally. Processes found to engage only unlimited-capacity processes using this method include, but are not limited to, contrast discrimination (Scharff et al., 2011a), image shape (Scharff et al., 2013), size discrimination of individual items (Huang & Pashler, 2005), modal and amodal surface completion (Attarha & Moore, 2010; Attarha et al., 2013), symmetry detection (Huang, Pashler, & Junge, 2004), and letter identification (Shiffrin & Gardner, 1972). Processes that yield results consistent with fixed-capacity include object categorization (Scharff et al., 2011b), object shape identification (Scharff et al., 2013), word categorization (Scharff et al., 2011a), and now multiple within-feature summary statistics. These processes constitute extreme conditions, with unlimited-capacity processing on the one hand and maximally limited-capacity (i.e., fixed-capacity) processing on the other. Together the results indicate that at some point (or points) within the stream of visual processing between contrast discrimination and object identification, severe limitations ensue. When drawing similarities between the processes at each extreme, it appears as though sensory and

119

segmentation processes have unlimited capacity while object and semantic processes have fixed capacity. The present study suggests that the formation of multiple within-feature summary statistic representations is more like object and semantic processing than it is like sensory or organizational processing.

A second proposed criterion of automaticity is that processing should be established quickly enough to avoid serial shifts of attention. Initial studies reported that summary extraction occurs in as little as 50 ms (Chong & Treisman, 2003). But these displays were never masked. When observers can no longer rely on sensory memory to inform their estimates of the mean, the amount of time required to achieve adequate performance increases four-fold to 200 ms (Whiting & Oriet, 2011; see also Attarha, Moore, & Vecera, 2014). This duration exceeds that of other purportedly automatic processes (e.g., Rayner et al., 1981).

Finally, a third criterion proposed to define a basic perceptual process is evidence of involuntarily processing. I argue that the results of chapter 5 offer the strongest evidence to date that summary representations require attention to establish.

Taken together, the more recent findings mentioned above suggest that summary representations may not meet the most basic of criteria that are used define automatic processing. The view that summaries engage parallel processes that unfold quickly and automatically without interference should be updated to say that summaries engage limited processes that unfold slowly with attention.

6.2.2 An alternative account of summary statistics

Debate persists over whether all items, most items, or only a very small subset of all items, are incorporated into the average (Ariely, 2008; Chong, Joo, Emmanouil, & Treisman, 2008; Dakin, 2001; Myczek & Simons, 2008). The fact that observers' performance improves as the number of items increase suggests that many items are averaged (Robitaille & Harris, 2011). However, there exists compelling evidence that the sampling algorithm used to compute SSRs may be much smaller in scope than current reports claim. Using static displays with spatially-distributed items, simulations by Myczek and Simons (2008) demonstrate that averaging over a subsample of only 2-3 items from a large set was sufficient to yield performance levels similar to those observed in the foundational papers of this literature (Ariely, 2001; Chong & Treisman; 2003, 2005a, 2005b). The number of items sampled by this algorithm is suspiciously similar to the storage capacity of visual working memory. Attarha, Moore, & Vecera (unpublished data) implemented an analog simulation to temporal summary statistics (i.e., items that change size over time). Like Simons and colleagues, we found that averaging performance does not exceed what would be expected using a working memory strategy; simulations based on averaging the sizes of only four items from the entire set modeled the data well. These findings suggest that temporal summary statistic representations are restricted by the storage capacity of visual working memory and that the processes involved in averaging over time cannot be engaged online continuously (e.g., Awh, Barton, & Vogel, 2007; Luck & Vogel, 1997; Vogel, Woodman, & Luck, 2001).

Obtaining decisive evidence to show that adequate averaging performance could not be achieved with a limited subsampling strategy is difficult. Such evidence would require tasks that demonstrate that averaging performance is significantly better than what is possible using visual working memory. That is, averaging the whole array of items must produce adequate performance while averaging just a portion of the display must produce poor performance. The challenge of implementing such tasks, however, is that subsampling of any combination of 3-5 items (e.g., Luck & Vogel, 1997) within a many-item set will approach the population mean of the set. Future work should implement such simulations to determine whether the results obtained can be produced without appealing to specialized averaging abilities. In the words of Myzczek and Simons (2008, p. 774), "It seems prudent to eliminate known mechanisms in order to support the existence of new ones". The range of effects cited in the SSR literature may be accounted for by known psychophysical principles (Allik et al., 2013) or by existing cognitive mechanisms (Myzczek & Simons, 2008).

In light of the evidence above, the application of summary representations to daily perceptual life is greatly diminished. The weakest conclusion that can be reached with this collection of results is that summaries, while possibly used in perception, are not a critical, fundamental component of early visual processing and later visual awareness (at least with respect to the generally accepted criteria that are typically used to define "early processes"; Brown, Gore, & Carr, 2002; Schmidt & Dark, 1998). The strongest conclusion that can be reached is that specialized averaging mechanisms may not exist at all, especially in the event that known processes are found to account for the complete range of effects (e.g., Myczek & Simons, 2008).

6.2.3   The role of between-feature summaries in visual perception

The finding of unlimited-capacity processing for between-feature summaries contributes to the discussion of whether there exists a general-purpose mechanism for summary extraction or more specialized mechanisms. The hierarchical view of summary statistics, which states that different statistical summaries are established in separate visual pathways, predicts that it should be possible to form summaries between dimensions insofar as each summary type engages different subsets of processes. Following this logic, I conclude that summaries of orientation and summaries of size are generated in separate areas in the stream of perceptual processing.

According to the hierarchical view, any between-feature averaging processes that do not overlap in resources, should not interfere, and therefore should unfold with unlimited capacity. While summary representations may be more spatially constrained than previously thought, they may be perceptually richer in localized, behaviorally relevant regions of space. After all, a summary representation of the leaves on a single tree, for example, may include the average color, size, and shape of the leaves and would be potentially more useful than representations of size alone for all visible trees. The results of Experiment 3C in Chapter 5 should therefore shift focus to understanding how multiple single-feature summaries help the system derive a global sense of completeness in at least one area of the peripheral visual field. It may be the case that between-feature summaries of a single collection of items, rather than within-feature summaries of multiple collections, are a relevant factor in theories of visual perception.

This conclusion is supported by the last experiment of chapters 2-4, which demonstrate equality of the simultaneous and sequential conditions for single sets. A point of concern, however, may be that the way in which our stimuli were divided in the sequential condition unintentionally produced the perception of two separate sets. If observers were unable to integrate these items into a single representation then the apparent equality of the simultaneous-sequential conditions may instead reflect a failure to group the stimuli the way they were intended. Across most experiments, we find that sequential performance across the multi-set experiments (e.g., Chapter 4, Experiment 1) and single-set experiments (e.g., Chapter 4, Experiment 3C) remains constant (58% vs 57%, respectively), suggesting that these stimuli were grouped in a perceptually similar manner. Simultaneous performance, however, improved for single-set summaries (55%) and decreased for multi-set summaries (42%), suggesting that these two displays were treated differently. The conclusion that single sets are truly unlimited may therefore require additional experimental support to ensure that an advantage of the sequential condition would not have be observed even if perceptual grouping cues were better equated across conditions.

6.3 The 'Grand Illusion' revisited

The conscious experience of our daily life and the development of our most adaptive behaviors are derived in part by processes that allow us to establish an internal representation of the external world. The usefulness of this internal representation in daily life, it was once thought, depended entirely upon how accurately the system could

process all of the available sensory information. The idea was that somewhere in the brain was a rich and detailed internal representation of the rich and detailed environment. It is now known that our perceptual and cognitive systems are severely limited in the quantity of information that they can process, suggesting instead that the poverty of the internal representation is analogous to the poverty of the visual input (e.g., Luck & Vogel, 1997; Nakayama, 1990; Raymond, Shapiro, & Arnell, 1992; Rensink, O'Regan, & Clark, 1997; Simons & Chabris, 1999). Herein lies a problem. How do we construct a meaningful internal representation that can guide behavior from relatively little information? When performance on daily tasks relies on an impoverished internal copy, then shouldn't behavior be similarly compromised? How is it, then, that many tasks are performed successfully?

From the discovery that observers can represent the average of a group of items, it was proposed that highly specialized statistical mechanisms complement the inherent limitations of our perceptual and cognitive systems to produce a relatively complete internal representation at minimal cost. The use of summary representations in daily life became a hot topic in the visual perception literature since they were thought to underlie the subjective experience of seeing (e.g., Alvarez, 2011; Corbett & Melcher, 2014). The current set of studies reevaluates this role of summary representations in daily life. But if summary statistics cannot explain our conscious experience of seeing then what can? The 'Grand Illusion' may instead result from the operation of established processes (e.g., Henderson & Hollingworth, 2002) or from a failure to encode perceptual experience into a reportable format (e.g., Moore & Egeth, 1997). For the time being, though, suffice it to say that there is not enough empirical support for the 'Grand Illusion' (Irwin, 1991;

O'Regan, 1992; O'Regan & Noë, 2001); this illusion may be a matter of faulty logic wherein the evidence of processing failures does not necessarily support the conclusion of a sparse internal representation (Cohen, 2002; Simons & Rensink, 2005), and furthermore that ecologically relevant behaviors need not require a representation of all of the available sensory information, only the information relevant to the perceptual observer's current goals.

Finally, it is critical to note that the sort of averaging defined by statistical summary representations is qualitatively distinct from the sort of averaging achieved by existing mechanisms. Up to this point, the term statistical summary representations has referred to processes that exhaustively analyze all or most items available within a given set for all or most sets appearing throughout the visual field. Although I use evidence from the current sets of projects to challenge this standard conceptualization of SSRs, "summary statistics" certainly exist to the extent that redundant information is pooled and averaged from a minimum number of two items and a maximum number of items defined by the perceptual observer's attentional and memory limitations. This latter form of statistical representation may therefore warrant either a new name for differentiation or, at the very least, a modifier to emphasize its limited scope. A question that emerges then is: what is the function of limited summary statistics if not to solve the problem of the grand illusion? Limited statistical averaging, broadly speaking, may contribute to efficient behavioral interaction in isolated, relevant areas of the visual field to which attention has already been directed through reducing at least some of the computational burden when large collections of similar items require encoding. Summary statistics may also subserve statistical learning. Zhao et al. (2011) found that learning statistical

regularities and performing a summary perception task greatly impeded each other such that effects of one task could not be observed while engaged in the other. These results suggest that summary statistics and statistical learning are fundamentally related. Under this view, summary statistics may be necessary to our ability to predict and learn visual regularities.

REFERENCES

Ackermann, J. F., & Landy, M. S. (2014). Statistical templates for visual search. *Journal of Vision*, *14*(3), 1–17. doi:10.1167/14.3.18

Albrecht, A. R., & Scholl, B. J. (2010). Perceptually Averaging in a Continuous Visual World: Extracting Statistical Summary Representations Over Time. *Psychological Science*, *21*(4), 560–567.

Albrecht, A. R., Scholl, B. J., & Chun, M. M. (2012). Perceptual averaging by eye and ear: Computing summary statistics from multimodal stimuli. *Attention, perception & psychophysics*, *74*(5), 810–815.

Allik, J., Toom, M., Raidvee, A., Averin, K., & Kreegipuu, K. (2013). An almost general theory of mean size perception. *Vision Research*, *83*, 25–39.

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, *15*(3), 122–131.

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*, 106-111.

Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science*, *19*(4), 392–398.

Alvarez, G. A., & Oliva, A. (2009). Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *PNAS*, *106*(18), 7345–7350.

Ariely, D. (2001). Seeing sets: representation by statistical properties. *Psychological Science*, *12*(2), 157–162.

Ariely, D. (2008). Better than average? When can we say that subsampling of items is better than statistical summary representations? *Perception & psychophysics*, *70*(7), 1325–1326.

Attarha, M. & Moore, C.M. (2014). Orientation summary statistics are limited in processing capacity. *Visual Cognition, 22*(8), 1018-1022. DOI: 10.1080/13506285.2014.960667

Attarha, M. & Moore, C.M. (in press). The capacity limitations of orientation summary statistics. *Attention, Perception, & Psychophysics.*

Attarha, M., & Moore, C. M. (2010). Testing capacity limitations of surface completion using the simultaneous-sequential method. *Visual Cognition*, *18*(10), 1487–1491.

Attarha, M., Moore, C. M., Scharff, A., & Palmer, J. (2014a). Evidence of unlimited-capacity surface completion. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(2), 556–565.

Attarha, M., Moore, C.M., & Vecera, S.P. (2014b). Summary statistics of size: Fixed processing capacity for multiple ensembles but unlimited processing capacity for single ensembles. *Journal of Experimental Psychology: Human Perception and Performance, 40*(4), 1440-1449.

Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, *18*(7), 622–628.

Balas, B., Nakano, L., & Rosenholtz, R. (2010). A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision*, *9*(12), 1–30. doi:10.1167/9.12.13

Banno, H., & Saiki, J. (2012). Calculation of the mean circle size does not circumvent the bottleneck of crowding. *Journal of Vision*, *12*(11), 1–15.

Bauer, B. (2009). Does Stevens's power law for brightness extend to perceptual brightness averaging? *The Psychological Record*, *59*, 171–186.

Bilsky, A. B., & Wolfe, J. M. (1995). Part-whole information is useful in visual search for size x size but not orientation x orientation conjunctions. *Perception & Psychophysics*, *57*(6), 749–760.

Bouma, H. (1970). Interaction effects in parafoveal letter recognition. Nature, 226, 177-178.

Brady, T. F., & Alvarez, G. A. (2011). Hierarchical Encoding in Visual Working Memory: Ensemble Statistics Bias Memory for Individual Items. *Psychological Science*, *22*(3), 384–392.

Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of Vision*, *11*(5), 1–34.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10,* 433– 436.

Brand, J., Oriet, C., & Tottenham, L. S. (2012). Size and emotion averaging: Costs of dividing attention after all. *Canadian Journal of Experimental Psychology*, *66*(1), 63–69.

Broadbent, D. E. (1958). *Perception and communication*. New York: Pergamon Press.

Brown, T. L., Gore, C. L., & Carr, T. H. (2002). Visual attention and word recognition in Stroop color naming: Is word recognition "automatic?." *Journal of Experimental Psychology: General*, *131*(2), 220–240.

Cavanagh, P. (2001). Seeing the forest but not the trees. *Nat. Neurosci*, *4*, 673–674

Cavanagh, P., Arguin, M., & Treisman, A. (1990). Effect of surface medium on visual search for orientation and size features. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(3), 479–491.

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, *43*(4), 393–404.

Chong, S. C., & Treisman, A. (2005a). Statistical processing: computing the average size in perceptual groups. *Vision Research*, *45*(7), 891–900.

Chong, S. C., & Treisman, A. (2005b). Attentional spread in the statistical processing of visual displays. *Perception & Psychophysics*, *67*(1), 1–13.

Chong, S. C., Joo, S. J., Emmanouil, T. A., & Treisman, A. (2008). Statistical processing: not so implausible after all. *Perception & Psychophysics*, *70*(7), 1327–1334. doi:10.3758/PP.70.7.1327

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. LEA: Hillsdale, NJ

Cohen, J. (2002). The grand grand illusion. *Journal of Consciousness Studies*, *9*, 141-157.

Corbett, J. E. & Song, J-H. (2014) Statistical extraction affects visually guided action. *Visual Cognition, 22*(7), 881-895. DOI: 10.1080/13506285.2014.927044

Corbett, J. E., & Melcher, D. (2014). Characterizing ensemble statistics: mean size is represented across multiple frames of reference. *Attention, Perception & Psychophysics*, *76*(3), 746–758.

Corbett, J. E., & Oriet, C. (2011). The whole is indeed more than the sum of its parts: perceptual averaging in the absence of individual item representation. *Acta psychologica*, *138*(2), 289–301.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology, 1*, 42□45.

Dakin, S. C. (2001). Information limit on the spatial integration of local orientation signals. *Journal of the Optical Society of America*, *18*(5), 1016–1026.

Dakin, S. C., & Watt, R. J. (1997). The computation of orientation statistics from visual texture. *Vision Research*, *37*(22), 3181–3192.

de Fockert, J., & Wolfenstein, C. (2009). Rapid extraction of mean identity from sets of faces. *The Quarterly Journal of Experimental Psychology*, *62*(9), 1716–1722.

Demeyere, N., Rzeskiewicz, A., Humphreys, K. A., & Humphreys, G. W. (2008). Automatic statistical processing of visual properties in simultanagnosia. *Neuropsychologia*, *46*(11), 2861–2864.

Eckstein, M. P., Thomas, J. P., Palmer, J., & Shimozaki, S. S. (2000). A signal detection model predicts the effects of set size on visual search accuracy for feature, conjunction, triple conjunction, and disjunction displays. *Perception & Psychophysics*, *62*(3), 425–451.

Emmanouil, T. A., & Treisman, A. (2008). Dividing attention across feature dimensions in statistical processing of perceptual groups. *Perception & psychophysics*, *70*(6), 946–954.

Eriksen, C. W., & Spencer, T. (1969). Rate of information processing in visual perception: some results and methodological considerations. *Journal of Experimental Psychology*, *79*(2), 1–16.

Estes, W K., & Taylor, H. A. (1964). A detection method and probabilistic models for assessing information processing from brief visual displays. *Proceedings of the National Academy of Sciences*, USA, 52, 446-454.

Feldman, J. A. (1985). Four frames suffice: A provisional model of vision and space. *Behavioral and Brain Sciences*, *8*, 265–289. doi:10.1017/S0140525X00020707

Fouriezos, G., Rubenfeld, S., & Capstick, G. (2008). Visual statistical decisions. *Perception & Psychophysics*, 70(3), 456–464.

Gabor, D. (1946). Theory of communication. *J. Inst. Electr. Eng, 24*, 891–910.

Gillen, C. & Heath, M. (2014). Perceptual averaging governs antisaccade endpoint bias. *Exp Brain Res, 232*, 3201-3210. DOI: 10.1007/s00221-014-4010-1

Haberman J. & Whitney D. (2011). Ensemble perception: Summarizing the scene and broadening the limits of visual processing. In Wolfe J., Robertson L. (Eds.), *A Festschrift in honor of Anne Treisman*. Oxford University Press.

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology*, *17*(17), R751–R753.

Haberman, J., & Whitney, D. (2009). Seeing the mean: Ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 718–734.

Haberman, J., Harp, T., & Whitney, D. (2009). Averaging facial expression over time. *Journal of Vision*, *9*(11), 1–13.

Halberda, J., Sires, S. F., & Feigenson, L. (2006). Multiple Spatially Overlapping Sets Can Be Enumerated in Parallel. *Psychological Science*, *17*(7), 572–576. doi:10.1111/j.1467-9280.2006.01746.x

Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(1), 113–136. doi:10.1037//0096-1523.28.1.113

Huang, L., & Pashler, H. (2005). Attention capacity and task difficulty in visual search. *Cognition*, *94*(3), B101–11.

Huang, L., Pashler, H., & Junge, J. A. (2004). Are there capacity limitations in symmetry perception? *Psychonomic Bulletin & Review*, *11*(5), 862–869.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol*, *160*(1), 106–154.

Im, H. Y., & Chong, S. C. (2009). Computation of mean size is based on perceived size. *Attention, Perception & Psychophysics*, *71*(2), 375–384.

Im, H. Y., & Chong, S. C. (2014). Mean size as a unit of visual working memory. *Perception*, *43*(7), 663–676. doi:10.1068/p7719

Im, H. Y., & Halberda, J. (2013). The effects of sampling and internal noise on the representation of ensemble average size. *Attention, perception & psychophysics*, *75*(2), 278–286.

Irwin, D. E. (1991). Information integration across saccadic eye movements. *Cognitive Psychology*, *23*(3), 420–456.

Jacoby, O., Kamke, M. R., & Mattingley, J. B. (2013). Is the whole really more than the sum of its parts? Estimates of average size and orientation are susceptible to object substitution masking. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(1), 233–244.

Jennings, J. R., & Wood, C. C. (1976). The e-adjustment procedure for repeated-measures analyses of variance. *Psychophysiology, 13*, 277–278.

132

Joo, S. J., Shin, K., Chong, S. C., & Blake, R. (2009). On the nature of the stimulus information necessary for estimating mean size of visual arrays. *Journal of Vision*, *9*(9), 1–12.

Kanwisher, N. G. (1987). Repetition blindness: type recognition without token individuation. *Cognition*, *27*(2), 117–143.

Kooi, F. L., Toet, A., Tripathy, S.P., Levi, D.M. (1994). The effect of similarity and duration on spatial interaction in peripheral vision. *Spatial vision, 8*(2), 255-279.

Luck, S. J., & Vogel, E.K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*, 279-281.

Lüschow, A., & Nothdurft, H. C. (1993). Pop-out of orientation but no pop-out of motion at isoluminance. *Vision Research*, *33*(1), 91–104.

Mack, A. & Rock, I. (1998). *Inattentional blindness*. Cambridge, MA: MIT Press.

Marchant, A. P., Simons, D. J., & de Fockert, J. W. (2013). Ensemble representations: Effects of set size and item heterogeneity on average size perception. *Acta Psychol*, *142*(2), 245–250.

Mauchly, J.W. (1940). Significance test for sphericity of n-variate normal population. *Annals of Mathematical Statistics, 11,* 204-209.

Miller, A. L., & Sheldon, R. (1969). Magnitude estimation of average length and average inclination. *Journal of Experimental Psychology. 81*(1), 16–21.

Miller, J. (1987). Priming is not necessary for selective-attention failures: semantic effects of unattended, unprimed letters. *Perception & Psychophysics*, *41*(5), 419–434.

Mordkoff, J. T., & Halterman, R. (2008). Feature integration without visual attention: Evidence from the correlated flankers task. *Psychonomic Bulletin & Review*, *15*(2), 385–389. doi:10.3758/PBR.15.2.385

Moore, C. M., & Egeth, H. (1997). Perception without attention: evidence of grouping under conditions of inattention. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(2), 339–352.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Reason, 4*(2), 61-64.

Morgan, M., Chubb, C., & Solomon, J. A. (2008). A "dipper" function for texture discrimination based on orientation variance. *Journal of Vision*, *8*(11), 1–8.

Myzczek, K., & Simons, D. J. (2008). Better than average: Alternatives to statistical summary representations for rapid judgments of average size. *Perception & psychophysics*, *70*(5), 772–788.

Nakayama, K. (1990). The iconic bottleneck and the tenuous link between early visual processing and perception. *Vision: Coding and Efficiency*. (C. Blakemore Ed.) (pp. 411–422). Cambridge University Press.

Noë, A. (2002). Is the visual world a grand illusion? *Journal of Consciousness Studies*, *9*(5-6), 1–12.

Noë, A., Pessoa, L., & Thompson, E. (2000). Beyond the grand illusion: What change blindness really teaches us about vision. *Visual Cognition*, *7*(1-3), 93–106.

O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology*, *46*(3), 461–488. doi:10.1037/h0084327

O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *The Behavioral and Brain Sciences*, *24*(5), 939–1031.

Oriet, C., & Brand, J. (2013). Size averaging of irrelevant stimuli cannot be prevented. *Vision Research*, *79*, 8–16.

Palmer, J. (1994). Set-size effects in visual search: The effect of attention is independent of the stimulus for simple tasks. *Vision Research*, *34*, 1703-1721.

Palmer, J. (1995). Attention in visual search: Distinguishing four causes of set-size effects. *Current Directions in Psychological Science, 4,* 118–123.

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

Pashler, H. E. (1998). *The psychology of attention*. Cambridge, MA: MIT Press.

Pelli, D. G. (1997). The Video Toolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442.

Peterson, C. R., & Beach, L. R. (1967). Man as an intuitive statistician. *Psychological bulletin*, *68*(1), 29–46.

Pollard, P. (1984). Intuitive judgments of proportions, means, and variances: A review. *Current Psychology*, *3*(1), 5–18.

Poltoratski, S., & Xu, Y. (2013). The association of color memory and the enumeration of multiple spatially overlapping sets. *Journal of Vision*, *13*(8), 1–11. doi:10.1167/13.8.6

Pylyshyn, Z.W. and R.W. (1988). Storm, Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spatial Vision,* 3(3), 1-19.

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: an attentional blink? *Journal of Experimental Psychology: Human Perception and Performance*, *18*(3), 849–860.

Rayner, K., Inhoff, A. W., Morrison, R., Slowiaczek, M. L., & Bertera, J. H. (1981). Masking of foveal and parafoveal vision during eye fixations in reading. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 167–179.

Rensink, R. A. (2000). Seeing, sensing, and scrutinizing. *Vision Research*, *40*, 1469–1487.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, *8*(5), 368–373.

Robitaille, N., & Harris, I. M. (2011). When more is less: Extraction of summary statistics benefits from larger sets. *Journal of Vision*, *11*(12), 1–8.

Rosenholtz, R. (2011). What your visual system sees where you are not looking. In B. E. R. and T. N. Pappas, (Eds.) SPIE: Human Vision and Electronic Imaging, XVI. 7865, 786510.

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*(4), 1–17.

Rumelhart, D. E. (1970). A multicomponent theory of the perception of briefly exposed visual displays. *Journal of Mathematical Psychology, 7*, 191-218.

Scharff, A., Palmer, J. P., & Moore, C. M. (2011a). Extending the simultaneous-sequential paradigm to measure perceptual capacity for features and words. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(3), 813-833.

Scharff, A., Palmer, J. P., & Moore, C. M. (2011b). Evidence of fixed capacity in visual object categorization. *Psychonomic Bulletin & Review, 18*(4), 713-721.

Scharff, A., Palmer, J. P., & Moore, C. M. (2013). Divided attention limits perception of 3-D object shapes. *Journal of Vision*, *13*(2), 1–24.

Schmidt, P. A., & Dark, V. J. (1998). Attentional processing of "unattended" flankers: evidence for a failure of selective attention. *Perception & Psychophysics*, *60*(2), 227–238.

Shapiro, S. S. & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika, 52* (3/4), 591–611.

Shaw, M. L. (1980). Identifying attentional and decision-making components in information processing. *Attention and Performance, 8,* 277–296.

Shaw, M.L. (1984). Division of attention among spatial locations: A fundamental difference between detection of letters and detection of luminance increments. In: *Attention and Performance X* (Bouma, H. & Bouwhuis, E.G., eds.), pp.106-121. Hillsdale, NJ: Earlbaum.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures1. *Journal of Verbal Learning and Verbal Behavior*, *6*(1), 156–163.

Shiffrin, R. M., & Gardner, G. T. (1972). Visual processing capacity and attentional control. *Journal of Experimental Psychology, 93*(1), 72–82.

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception*, *28*(9), 1059–1074.

Simons, D. J., & Levin, D. T. (1998). Failure to detect changes to people during a real-world interaction. *Psychonomic Bulletin & Review*, *5*(4), 644–649. doi:10.3758/BF03208840

Simons, D. J., & Rensink, R. A. (2005). Change blindness: past, present, and future. *Trends in Cognitive Sciences*, *9*(1), 16–20. doi:10.1016/j.tics.2004.11.006

Sperling, G. (1960). *The information available in brief visual presentations* (Vol. 74, pp. 1–29). Psychological Monographs: General and Applied.

Suzuki, S. (2005). High-level pattern coding revealed by brief shape aftereffects. In C. Clifford and G. Rhodes (Eds.), *Fitting the mind to the world: Adaptation and aftereffects in high-level vision* (Advantages in Visual Cognition Series, Vol. 2), Oxford University Press.

Sweeny, T. D., Haroz, S., & Whitney, D. (2013). Perceiving group behavior: Sensitive ensemble coding mechanisms for biological motion of human crowds. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(2), 329–337.

Teghtsoonian, M. (1965). The judgment of size. *The American Journal of Psychology*, *78*(3), 392–402.

Townsend, J. T. (1990). Serial vs. parallel processing: Sometimes they look like Tweedledum and Tweedledee but they can (and should) be distinguished. *Psychological Science, 1*, 46-54.

Arnold Trehub (1991) *The Cognitive Brain.* Cambridge, Mass.: MIT Press.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136.

Treisman, A., & Souther, J. (1985). Search asymmetry: A diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology: General, 114*, 285-310.

Utochkin, I. S., & Tiurina, N. A. (2014). Parallel averaging of size is possible but range-limited: A reply to Marchant, Simons, and De Fockert. *Actpsy*, *146*(C), 7–18. doi:10.1016/j.actpsy.2013.11.012

Walker, D., & Vul, E. (2014). Hierarchical Encoding Makes Individuals in a Group Seem More Attractive. *Psychological Science*, *25*(1), 230–235.

Watamaniuk, S. N., & Duchon, A. (1992). The human visual system averages speed information. *Vision Research*, *32*(5), 931–941.

Watamaniuk, S. N., Sekuler, R., & Williams, D. W. (1989). Direction perception in complex dynamic displays: the integration of direction information. *Vision Research*, *29*(1), 47–59.

Webster, M. A., & De Valois, R. L. (1985). Relationship between spatial-frequency and orientation tuning of striate-cortex cells. *J. Opt. Soc. Am. A*, *2*(7), 1124–1132.

Weiss, D. J., & Anderson, N. H. (1969). Subjective averaging of length with serial presentation. *Journal of Experimental Psychology*, *82*(1), 52–63.

Whiting, B. F., & Oriet, C. (2011). Rapid averaging? Not so fast! *Psychonomic Bulletin & Review*, *18*(3), 484–489.

Whitney, D. (2009). Vision: Seeing through the Gaps in the Crowd. *Current Biology*, *19*(23), R1075–R1076.

Whitney, D., Haberman, J., & Sweeny, T. D. (2014). From textures to crowds: Multiple levels of summary statistical perception. In J. S. Werner & L. M. Chalupa (Eds.), The new visual neurosciences. Cambridge, MA: MIT Press. ISBN: 978-0-262-01916-3; 978-0-262-31789-4

Wolfe, J. M. (1994). Guided Search 2.0 A revised model of visual search. *Psychonomic Bulletin & Review*, *1*(2), 202–238.

Wolfe, J. M. (1998). Visual search. In H. Pashler (Ed.), *Attention*. Hove, England: Psychology Press.

Wolfe, J. M., Friedman-Hill, S. R., & Bilsky, A. B. (1994). Parallel processing of part-whole information in visual search tasks. *Perception & Psychophysics*, *55*(5), 537–550.

Wolfe, J. M., Võ, M. L. H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and nonselective pathways. *Trends in Cognitive Sciences*, *15*(2), 77–84. doi:10.1016/j.tics.2010.12.001

Wolfe, J. M., Yu, K. P., Stewart, M. I., Shorter, A. D., Friedman-Hill, S.R., & Cave, K.R. (1990). Limitations on the parallel guidance of visual search: Color × Color and Orientation × Orientation conjunctions. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 879–892

Zosh, J. M., Halberda, J., & Feigenson, L. (2011). Memory for multiple visual ensembles in infancy. *Journal of Experimental Psychology: General*, *140*(2), 141–158. doi:10.1037/a0022925