February 2023

Temporal and Spatial Properties of Orientation Summary Statistic Representations

Jacob S. Zepp
University of South Florida

Follow this and additional works at: https://digitalcommons.usf.edu/etd

Part of the Cognitive Psychology Commons

Scholar Commons Citation

This Thesis is brought to you for free and open access by the USF Graduate Theses and Dissertations at Digital Commons @ University of South Florida. It has been accepted for inclusion in USF Tampa Graduate Theses and Dissertations by an authorized administrator of Digital Commons @ University of South Florida. For more information, please contact digitalcommons@usf.edu.
Temporal and Spatial Properties of Orientation Summary Statistic Representations

by

Jacob S. Zepp

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Arts
Department of Psychology
College of Arts and Sciences
University of South Florida

Co-Major Professor: Chad Dubé, Ph.D.
Co-Major Professor: Kenneth Malmberg, Ph.D.
Elizabeth Schotter, Ph.D.
Kristen Salomon, Ph.D.

Date of Approval:
December 8, 2022

Keywords: ensemble averaging, summary perception, memory, perception, visual processing, visual gist

Copyright © 2022, Jacob S. Zepp
# TABLE OF CONTENTS

List of Tables ........................................................................................................................................ iii

List of Figures ........................................................................................................................................ iv

List of Abbreviations ............................................................................................................................ vi

Abstract ................................................................................................................................................ vii

Chapter One: Introduction .......................................................................................................................... 1
  “Early” Summarizing .......................................................................................................................... 3
  Debate on the Necessity of a New Mechanism .............................................................................. 6
  The Efficiency of Information Integration ......................................................................................... 8
  The Zoom-Lens Model ...................................................................................................................... 10
  The Fidelity-Based Integration Model ............................................................................................... 12
  Challenging the Automaticity Assumption ....................................................................................... 14
  Iconic vs vSTM Item Contributions .................................................................................................. 18
  Controlling for the “When” and the “What” of SSRs ....................................................................... 20
  The Current Work ............................................................................................................................ 23

Chapter Two: Statistical Methods ............................................................................................................ 24
  A Brief Introduction to Circular Statistics ......................................................................................... 24
  Orientation Averaging Analysis .......................................................................................................... 27

Chapter Three: Experiment 1: The “What” of SSRs ........................................................................... 29
  Subjects ............................................................................................................................................... 30
  Design ............................................................................................................................................... 30
  Stimuli ............................................................................................................................................... 31
  Procedure ......................................................................................................................................... 34
  Orientation Task Analysis .................................................................................................................. 36
  Results ................................................................................................................................................. 39
  Discussion .......................................................................................................................................... 40
  Whole-Report Response Characteristics ............................................................................................ 41
  Results ................................................................................................................................................. 42
Chapter Four: Whole-Report as a Subsampling Model .................................................................48
  Averaged Whole-Report Model ........................................................................................................48
  Whole-Report as a Model of Non-Selective Transfer .......................................................................50
  Alternative Models...........................................................................................................................53
    Automatic Averaging Model ...........................................................................................................53
    Selective Sampling Model ............................................................................................................55
    Early + Late Averaging Model .......................................................................................................58
  Discussion .........................................................................................................................................60

Chapter Five: Experiment 2: The “When” of SSRs .......................................................................62
  Subjects ...........................................................................................................................................63
  Design ...............................................................................................................................................63
  Stimuli ...............................................................................................................................................63
  Procedure ..........................................................................................................................................63
  Analysis .............................................................................................................................................65
  Results ...............................................................................................................................................66
  Discussion .........................................................................................................................................68

Chapter Six: Modeling Encoding Time Effects .............................................................................70
  Simulating Subsamples ....................................................................................................................70
  Precision as a Function of Encoding Time .......................................................................................71
  Model of Encoded Information .........................................................................................................73
  Discussion .........................................................................................................................................75

Chapter Seven: Conclusions ..........................................................................................................80
  Information and the SSR....................................................................................................................80
  Experimental Design Contributions ....................................................................................................81
  Future Directions ...............................................................................................................................82
  Implications for Perception and Memory ...........................................................................................84
  Concluding Remarks ..........................................................................................................................85

References ........................................................................................................................................87

Appendix A: Experimental Instructions ..........................................................................................94
  Experiment 1 Instructions ..................................................................................................................94
  Experiment 2 Instructions ..................................................................................................................97

Appendix B: IRB Approval Letter ....................................................................................................99
LIST OF TABLES

Table 1: An example of the stimuli presented for the first and last 3 trials within each block for the first two subjects .................................................................33

Table 2: Results from the Bayesian mixed-effects regression for Experiment 1 ...............38

Table 3: Results from the Bayesian mixed-effects regression for Experiment 2 ...............67
**LIST OF FIGURES**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>The processing stages within the zoom-lens model of SSRs</td>
<td>10</td>
</tr>
<tr>
<td>Figure 1.2</td>
<td>A simplified diagram detailing the stages of the FIM model for a mean orientation estimation task over simultaneously presented stimuli</td>
<td>12</td>
</tr>
<tr>
<td>Figure 1.3</td>
<td>Experimental design of Zepp, Dubé, and Melcher (2021)</td>
<td>16</td>
</tr>
<tr>
<td>Figure 1.4</td>
<td>A simplified representation of the separate modes of memory for visual information</td>
<td>18</td>
</tr>
<tr>
<td>Figure 3.1</td>
<td>An example of the 4 orientation feature location structural configurations</td>
<td>32</td>
</tr>
<tr>
<td>Figure 3.2</td>
<td>An example of the Experiment 1 trial progressions</td>
<td>35</td>
</tr>
<tr>
<td>Figure 3.3</td>
<td>Circular histograms of the orientation averaging response error distributions for each condition</td>
<td>37</td>
</tr>
<tr>
<td>Figure 3.4</td>
<td>Highest density intervals for Experiment 1 model parameter estimate posterior distributions</td>
<td>39</td>
</tr>
<tr>
<td>Figure 3.5</td>
<td>Probability distribution of the number of items reported in the whole-report task</td>
<td>43</td>
</tr>
<tr>
<td>Figure 3.6</td>
<td>Probability distribution of the number of items reported in the whole-report task during each of the two range conditions</td>
<td>43</td>
</tr>
<tr>
<td>Figure 3.7</td>
<td>Probability distribution of the number of items reported in the whole-report task during each of the four stimulus configuration conditions</td>
<td>44</td>
</tr>
<tr>
<td>Figure 3.8</td>
<td>A heatmap displaying the proportion of responses given in each of the 16 possible response locations</td>
<td>44</td>
</tr>
<tr>
<td>Figure 3.9</td>
<td>A heatmap displaying the proportion of responses given in each possible response locations, segmented by the range condition</td>
<td>45</td>
</tr>
<tr>
<td>Figure 3.10</td>
<td>A heatmap displaying the proportion of responses given in each possible response locations, segmented by the stimulus configuration condition</td>
<td>46</td>
</tr>
</tbody>
</table>
Figure 4.1: Model of the relationship between ensemble averaging and whole-report response..............................................................48

Figure 4.2: Circular histograms of the orientation averaging task error distributions with averaged whole-report response error distribution model fits..........................50

Figure 4.3: Circular histograms of the simulated averaged subsample response error distributions for each condition .............................................................................52

Figure 4.4: Alternative model where averaging occurs early over all items in the iconic store ........................................................................................................................................54

Figure 4.5: Circular histograms of the simulated early averaging model response error distributions for each condition.................................................................55

Figure 4.6: An alternative model where the central four elements of the display are sampled for averaging........................................................................................................56

Figure 4.7: Circular histograms of the simulated central four item subsampling model response error distributions for each condition....................................................57

Figure 4.8: Alternative model which bridges the gap between early and late models........58

Figure 4.9: Circular histograms of the simulated unequal weighting model response error distributions for each condition.................................................................59

Figure 5.1: An example of the Experiment 2 trial progressions.................................................................64

Figure 5.2: Highest density intervals for Experiment 2 model parameter estimate posterior distributions ........................................................................................................66

Figure 5.3: The effect of stimulus frame presentation time and presence of a mask on the precision of the SSR error distribution ...............................................................68

Figure 6.1: An example of simulating the sampling function Experiment 1 whole-report data .........................................................................................................................................71

Figure 6.2: Precision of orientation averaging as a function of available encoding time.................................................................................................................................72

Figure 6.3: Fit of the encoded subsampled model to Experiment 2 subject data .......................75

Figure 6.4: An example implementation of a sampling procedure that is further limited by encoding time ........................................................................................................78
# LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviations</th>
<th>Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td>CI</td>
<td>Credible Interval</td>
</tr>
<tr>
<td>FIM</td>
<td>Fidelity-Integration Model</td>
</tr>
<tr>
<td>HDI</td>
<td>Highest-Density Interval</td>
</tr>
<tr>
<td>ICC</td>
<td>Intracluster Correlation Coefficient</td>
</tr>
<tr>
<td>LTM</td>
<td>Long-Term Memory</td>
</tr>
<tr>
<td>ms</td>
<td>Milliseconds</td>
</tr>
<tr>
<td>s</td>
<td>Seconds</td>
</tr>
<tr>
<td>SSR</td>
<td>Statistical Summary Representation</td>
</tr>
<tr>
<td>VM</td>
<td>Von-Mises Distribution</td>
</tr>
<tr>
<td>VPC</td>
<td>Variance Partition Coefficient</td>
</tr>
<tr>
<td>vSTM</td>
<td>Visual Short-Term Memory</td>
</tr>
</tbody>
</table>
ABSTRACT

The aim of the current work was to determine the amount of information that contributes to the formation of summary statistical representations (SSRs), as well as the time course over which these representations are formed. While the prevailing interpretation of SSRs within literature is that the summaries are formed through a compulsory rapid integration across all information in a scene, debate exists on the necessity of this unique processing mode. To investigate the formation of SSRs, two experiments were conducted. In the first, results from an orientation averaging task were compared to results from a whole-report task, over equivalent stimulus displays. The purpose of this experiment was to determine if items reported in a visual short-term memory (vSTM) task could predict responses on an orientation averaging task, to inform a likely set of items that contribute to SSRs. The second experiment was conducted to determine whether SSRs are generated by a time dependent process, and the rate at which information is accumulated into the averaging process. This experiment consisted of conducting an orientation averaging task on masked or unmasked displays, with variably brief exposure durations. Results indicated that SSRs are predicted by whole-report responses, and follow a pattern of temporal results consistent with encoding information into vSTM. Together, the results provide evidence against a unique early summarization process, and instead indicate an averaging control process acting on items within vSTM. The results were incorporated into a simple computational model of the SSR generation process.
CHAPTER ONE:

INTRODUCTION

The visual system is tasked with making efficient use of a constant input stream of information to allow navigation of our environment. Within this flood of information, very little can be processed into usable and distinguishable units at any given time (Cowan, 2001). Luckily, there exists a great amount of redundancy and consistency within the visual features of the environment. Recent research has suggested a new perspective on how the visual system wades through this overload of visual information. The theory of summary statistic representations (SSR; also known as summary perception or ensemble perception) posits that the visual system creates distributional representations of visual features from which estimates of statistical moments can be extracted for quick and efficient judgements of the scene. The classic analogy for this idea is that one sees the forest before the trees, highlighting a global or ‘gist’ representation that can be used to compensate for limitations in the processing of individual objects within a scene. In a review of the theory, Whitney and Yamanashi Lieb (2018) operationally define SSRs based on the following characteristics:

- Ensemble perception is the ability to discriminate or reproduce a statistical moment.
- Ensemble perception requires the integration of multiple items.
• Ensemble information at each level of representation can be precise relative to the processing of single objects at that level.

• Single-item recognition is not a prerequisite for ensemble coding.

• Ensemble representations can be extracted with a temporal resolution at or beyond the temporal resolution of individual object recognition.

While this definition provides a good overview of the current state of the theory, considerable debate exists on the specifics of how the SSR is constructed, particularly in the number of items integrated by the mechanism (Ariely, 2001; Chong et al., 2003; 2005a,b; 2008; Myczek & Simons, 2008; Simons & Myczek, 2008; Corbett & Oriet, 2011; Marchant, et al. 2013; Utochkin & Tiurina, 2014; Zepp, et al., 2021), as well as what point in the visual processing stream the averaging mechanism may be acting (Chong & Treisman, 2003, 2005a,b; Joo, et al., 2009; Boduroglu & Yildirim, 2020; Yörük & Boduroglu, 2020; Whiting & Oriet, 2011; Epstein, et al., 2020; Zepp, et al., 2021). Given that this ability is posited as an elemental process within visual processing, clarity on how and when exactly this averaging mechanism operates is key for an understanding of visual scene perception.

Within the literature on SSR production, the vast majority of work has centered on the idea that perceptual averaging is a rapid process that integrates information from all stimuli within a scene (Alvarez & Oliva, 2008, 2009; Chong & Treisman, 2003, 2005b, 2005a; Chong, Joo, Emmanouil, & Treisman, 2008; Emmanouil & Treisman, 2008; Oriet & Brand, 2013). This interpretation generally states that averaging either requires significantly fewer attentional resources to be extracted and/or is an automatic biproduct of an initial attempt to process a scene with limited attentional resources (Alvarez, 2011; Baek & Chong, 2020a,b). As the deployment of focused attention resources to encode individual stimuli takes a non-trivial amount of time,
such as in a visual search task (Wolfe, 2015), an automatic averaging process would allow for a much more rapid extraction of stimulus gist features by bypassing this processing bottleneck. The time course of such a rapid summary extraction process over several items would exceed the capacity and encoding limitations that are normally reported within the visual working memory and visual perception literature for the set of individual items (Cowan, 2001; Sperling, 1967; Vogel et al., 2006). This class of theories therefore amount to an *early* generation process of SSRs, in that they are characterized by a rapid extraction of the summary which occurs before focused attention to individual items deploys.

In contrast, conclusions drawn by other authors detail a process that occurs after focused attention deployment and therefore results in a *late* generation of SSRs over a subset of items within a display (Myczek & Simons, 2008; Simons & Myczek, 2008; de Fockert & Marchant, 2008; Marchant & de Fockert, 2009; Marchant et al., 2013; Zepp et al., 2021). Proponents of the *late* theory of SSRs generally point towards various findings demonstrating that phenomena associated with SSRs do not require a process that exceeds capacity or temporal limits present within the existing understanding of perceptual processing. While the *early* generation of SSRs is the current predominant theory, the debate surrounding this distinction is detailed in the following section.

**“Early” Summarizing**

In what may be considered the beginning of the current explosion of research into SSRs, Ariely (2001) developed two experimental paradigms to introduce the idea of “set representations”, a precursor to SSRs. These two methods, the mean discrimination and member identification tasks, allowed for an experimental comparison of the “gist” information of a stimulus set vs the individual item representation retention abilities of subjects. In the mean
discrimination task, subjects were asked to determine whether a test disc was larger or smaller than the average size of a set of previously presented discs. In the member-identification task, subjects were tasked with determining whether a test disc was present within the previous stimulus set of discs. Using these methods, Ariely was able to show that subjects were highly accurate at reporting information about the average feature values of a set, even without the ability to report on individual item information. Equally surprising, the research found no decrements in performance when the number of items in the display was increased from 4-16 items. The author suggested that these results show a process within which the visual system deals with quickly extracting information from sets of similar items that are greater in number than the limited bottleneck of visual perception would be expected to handle. Ariely concluded that subjects represented the features of all items as a global statistical representation, which serves as a precursor to individual item representations. Further, the author suggested that unless necessitated by the task the individual item representations are discarded in favor of the SSRs for efficient visual perception (Ariely, 2001). This work proposed a unique hierarchy of visual information representations, in that the default way the visual system may represent information is just as a summary of the external environment with extra resources devoted to capturing certain items or features when needed. Given a system with a limited set of resources, this may be considered a logical way that humans may capture the most amount of information at a time.

Following Ariely, the influential work of Chong and Treisman on the properties of SSRs sits at the heart of the rapid and distributed process view of SSR generation. The authors’ initial investigation into SSR generation began by using a mean comparison task. In this task, subjects are shown a display of stimuli with varying sizes, spatially divided into two groups on each side of the display. Subjects are asked to determine which half of the display contained the set of
stimuli with the larger mean size. The authors modulated both the distribution shape of features within the stimulus set and the exposure duration of the stimulus display (Chong & Treisman, 2003). The authors were interested in comparing single item representation abilities to that of multiple items of varying distributional spreads (i.e. adjusting the similarity of items within the scene), and how dissimilarities between these representations evolve over time.

The authors’ results showed that subjects were as accurate at distinguishing between two single stimuli as they were for sets of heterogeneously distributed stimuli, and even more accurate at discriminating between sets of homogeneous redundant copies of the same stimuli. Strikingly this effect persisted at as little as 50ms exposure time. With longer exposure times, single item performances matched that of the homogeneous displays, and performance on heterogeneously distributed displays increased to a lesser degree. While it is important to note that the stimulus displays were not masked, and therefore processing time was not adequately controlled to make inferences about temporal processing speed (see Whiting & Oriet, 2011), the results demonstrated that high levels of performance were obtainable in the mean size averaging task at short display exposure durations. The patterns of superior performance in the homogeneous condition across exposure durations, and the diminishing returns of increasing exposure duration on the heterogeneous condition, led the authors to propose an advantage to redundant feature presentations with a ceiling on improvement due to internal noise cancellation in averaging heterogeneous items (Chong & Treisman, 2003).

Following up on these results, the authors investigated the attentional distribution over space of SSRs by using a dual-task design that required either focused or distributed attention. Concurrently, a mean discrimination task was paired with each attention condition (Chong & Treisman, 2005a). These experiments found that performance on the mean discrimination task
was supported when attention was cued to be distributed rather than localized. When attention was cued to be distributed, there did not appear to be a detriment in performance between completing the mean estimation task as a concurrent task versus completing it as a singular task, implying that attentional selection is not required to generate a summary representation. In another follow-up, Chong and Treisman (2005b) compared the ability to compute SSR averages over stimuli that are segmented by spatial groupings to stimuli that are intermixed spatially but segregated into sets by color. This experiment found that averages over stimuli segregated by color were as accurately reported as those discriminated by location, with little effect of set size, density, or colored distractor items. Taken together the results of this body of work posit that SSRs are generated very rapidly, though after items are segregated by color and location, in a diffuse attentional state which allows for summary judgements to be made as accurately as judgements between single items.

**Debate on the Necessity of a New Mechanism**

The proposal of an early mechanism of SSRs, unbound by the number of items to be processed or the time available for processing, is surprising within the idea of a working memory storage mode that is limited by information capacity (Cowan, 2001; Luck & Vogel, 1997, 2013; Ma et al., 2014) and information transfer rate (Sperling, 1967; Gegenfurtner & Sperling, 1993; Vogel et al., 2006). Before accepting that a new mechanism of perceptual processing is responsible for the observed phenomenon though, it must first be shown that a new mechanism is required to produce the reported experimental results. Testing this necessity, Myczek and Simons (2008) performed a conceptual replication of the previously discussed research by Ariely (2001) and Chong and Treisman (2003, 2005a, 2005b). They simulated results of subjects using a conventional focused attention strategy to generate responses, instead of averaging over all
items, by sampling a small subset of items (between 1 and 4 items) from the stimulus display. The subset was then averaged for the SSR response, modeling a strategic approach to completing averaging tasks consistent with classic working memory limitations. The results of these simulations showed that not only does a strategic approach of subsampling the stimulus array produce comparable results to experiments which conclude a global averaging mechanism, but that comparable results are found by “averaging” only 1 or 2 sampled items. While the authors were careful to note that their results do not necessitate that subjects are employing said strategy of serially sampling items strategically, the results do show that a new mechanism is not required to produce comparable averaging results under the conditions of the previously discussed SSR experiments (Myczek & Simons, 2008).

In reply to Myczek and Simons (2008), Chong and colleagues (2008) performed a series of experiments targeting the subsampling strategies that were proposed to account for the findings in Chong and Treisman’s previous work (Chong & Treisman, 2003, 2005a,b). Each experiment consisted of a modified version of the previously described experimental trials of Chong and Treisman (2003, 2005b). The first experiment intermixed trials of differing disc size distribution configurations to investigate expected performance costs that would appear if subjects were adjusting a focused attention strategy for subsampling disc sizes. The results show that switching between focused attention subsampling strategies was implausible, as no decrement in performance between the blocked and intermixed trials was observed. The second experiment found that presenting a physical subsample of the display (i.e. only 1 or 2 discs from an 8 disc display) produced less accurate averages of the full display than presenting all items. The third experiment showed that the possible strategy of sampling just the largest disc was unlikely, due to no observed differences in performance when this strategy would provide the
correct answer versus when it would not. The authors concluded that the results were more consistent with those of a global statistical averaging process rather than a subsampling account.

In reply to this work, Simons and Myczek (2008) highlighted that there are two key assumptions in the preceding conclusions of early SSR generation that have yet to be adequately proven empirically: 1) that performance on averaging tasks exceeds any prediction that may be provided within the limits of focused attention, and 2) that performance necessitates that all items in the display have been incorporated into the representation. Simons and Myczek argue that Chong and colleagues’ assumption that the subjects employ the focused attention strategies described in their work, or that they must change strategies between the trial configurations, is not well founded or controlled empirically (Simons & Myczek, 2008). A different choice in the focused attention strategy, such as sampling the largest and smallest items from each display and averaging over this subsample, could provide satisfactory results within all present experiments. Importantly, the experiments by Chong and colleagues (2008) tackle the first assumption above, though proving this assumption does not necessitate a proof of the second assumption. The acceptance of a new processing mechanism requires that the new mechanism be necessary to describe observed phenomena.

The Efficiency of Information Integration

The previous discussion produces an obvious question: if all items are not integrated into an SSR, then how many are? Measures of this efficiency have varied widely based on stimulus type, task, and assumptions of the efficiency estimation method (see Whitney & Yamanashi Lieb, 2018 for a review). Nonetheless, the number of items to be integrated within an SSR has been reported to approximately follow a square root function of the number of items presented (Dakin, 2001; Solomon, 2010; Solomon, et al., 2011; Whitney & Yamanashi Lieb, 2018) and to
be as high as 90 items (Dakin, 2001), which is well beyond the number of items generally thought to be available to focused attention. While this large capacity supports an early pre-focused-attention interpretation of SSRs, it should be noted that if the attentional demands of a task are increased then the effective number of stimuli that are integrated decreases (Dakin, et al., 2009). This implies a necessary relationship between the amount of attentional resources available and the fidelity of the summarization process.

One popular explanation for how an early mechanism of SSR generation creates an accurate response from a large amount of noisy information is the power of averaging (Alvarez, 2011; Sun & Chong, 2019). This theory states that averaging over several noisy item representations may cancel out individual noise, resulting in a less noisy summary representation. Consistent with this idea, while increasing the number of stimuli in a display generally hinders performance in tasks that require focused attention, various studies have reported increased accuracy in averaging tasks when set size is increased and the range of feature values is low (Robitaille & Harris, 2011; Utochkin & Tiurina, 2014; Lee, et al., 2016; Baek & Chong, 2020b). Further studies have reported that when individual item representations would be unavailable to a subject, either through instruction not to attend (Alvarez & Oliva, 2008, 2009; Oriet & Brand, 2013) or through visual crowding (Parkes, et al., 2001; Dakin, et al., 2009; Whitney & Levi, 2011), SSR estimates over the scene are still reported accurately. In these situations, one would expect that individual item representations would be highly noisy though the summary representation persists accurately. The cancelation of noise while averaging over large numbers of items sits at the heart of most models of SSR generation.
Baek and Chong (2020a,b) have recently proposed a computational model that represents SSRs as an early process (represented in Figure 1.1). The “zoom-lens model” builds on the framework provided in Lu and Dosher’s perceptual template model (Lu & Dosher, 1998; Dosher & Lu, 2000), which describes the effects of attention within visual processing in terms of the modulation of separate internal noise sources: a nonlinear perceptual template that acts as a feature detector to some degree of precision, a multiplicative noise source that comes from processing the input stimulus, and an additive noise source that applies to the output of the processed stimulus information before a decision is made. Within this framework, attention acts by either tuning the precision of the nonlinear template function (for distractor exclusion), or by reducing the noise present within the multiplicative stimulus processing function (for signal enhancement).
The zoom-lens model uses Lu and Dosher’s framework to explain SSR generation by assuming a limited capacity for attentional resources, which acts equally across all items presented within a display. Attention acts to reduce the internal multiplicative noise of the item representations, leading to signal enhancement of all items in the display. When attention is distributed across a wide attentional window (such as when the number of items in the display is large), each individual item receives relatively few attentional resources and therefore the noise reduction of each item is low. In contrast, a narrow attentional window allots greater resources to the reduction of early item noise (Baek & Chong, 2020a). When an average judgement of display features is required, an integrator averages over all present noisy representations while adding some amount of late noise to the SSR judgement (see Figure 1.1). The authors propose this early distributed attention model as separate from focused attention over all items, though the output of this early mode can still be modulated by focused attention effects, such as expectations (Baek & Chong, 2020b). This model formalizes the claims described earlier by Chong and Treisman (2003, 2005a,b, 2008) of an early and rapid mechanism of SSR production.

The authors show that the zoom-lens model fits empirical data patterns very well through an assumption of distributed attention acting on all items. The model was tested by comparing several configurations of global pooling models across varying set sizes; including a noisy-selection subsampling model (Allik et al., 2013), a spotlight model that only reduces early noise on some items, and models with no early attentional reduction (Baek & Chong, 2020b). The comparisons showed that the subsampling, focused, and distributed attention models fit the data well, with the distributed attention model slightly outperforming the others in fit statistics. The authors favor the distributed attention model due to theoretical considerations of the difference in purpose between SSRs vs focused attention in visual perception, findings that unattended items
contribute to visual processing, and that the distributed attention model did not require a parameter to fit the attention selection window size (Baek & Chong, 2020b). The framework put forward within the zoom-lens model significantly advances the understanding of SSRs by incorporating an attention component that modulates item representation fidelity and computationally formalizing a theory of early SSR generation.

The Fidelity-Based Integration Model

![FIDM Diagram]

*Figure 1.2:* A simplified diagram detailing the stages of the FIM model for a mean orientation estimation task over simultaneously presented stimuli. In this model, items contribute to the averaged representation with variable weights determined by the representational fidelity of each item. The higher fidelity items (10° and 15°) contribute more to the integrated representation compared to the lower fidelity (25° and 30°) items, leading to an output that is skewed towards the higher fidelity item values (Tong, 2020).

Anderson’s information integration theory provided a conceptual model to describe how humans can combine various sources of perceptual input information to output decisions (Anderson, 1981). This conceptual model framework consists of three stages to describe the process: input valuation, integration, and a response mechanism. The input valuation function maps the stimulus features onto an internal representation, the integration function combines the separate internal representations into a singular representation, and the response function produces the output of the process. Building on this conceptual framework, Tong (2020) developed the fidelity-based integration model (FIM) to provide a computational expression of
Anderson’s model (see Figure 1.2). In FIM, the valuation function maps external stimuli to an internal scale by representing the stimuli as probability distributions over feature space. Stimuli are encoded with variable fidelity, which is represented by the spread of the aforementioned probability distributions. The fidelity of the representation is a consequence of a limited amount of cognitive resources, and the demands and features of the task at hand. The distributions are then integrated by sampling these separate distributions into a single representative distribution. The relative representation of each item distribution within the integrated representation is dependent on the fidelity of the item representations, with higher fidelity items being sampled more into the integrated representation. The sampling function therefore creates a fidelity-based weighting for the integrated representation. Finally, the response function is dependent on the structure of the task and specifies the method in which a value is sampled from the integrated representation for output.

In addition to this computational framework, the FIM model provides a categorization of task dependent SSR influences into implicit and explicit effects. The sampling process of items into the integrated representation is modulated by the demands of the task at hand. When the task does not require a response based on the average stimulus feature value, such as in responding with a single item, implicit averaging occurs. Implicit averaging refers to the integration of samples from task irrelevant information, both non-target items on the current task and information from the previous tasks, into the target response. For example, when responding with a single line length from a set of line stimuli, the response will be primarily driven by the target line stimulus with a bias occurring towards samples from the irrelevant lines. For tasks in which the goal is to produce or to directly make use of the average feature value of a stimulus set, explicit averaging occurs. In explicit averaging, the task is to make use of information from
the full set of stimuli and so the samples that are integrated are more distributed across item representations. Importantly, integration of information within exemplar tasks is sampled both from stimuli within the set, as well as from previous trials. Once again, the relative contributions to the integrated representation of samples from the current trial item information and the previous trial information in both implicit and explicit averaging is determined by the fidelity of the item representations.

This framework provides a significant advancement to formalizing a model of SSR phenomenon, particularly in the categorization of various types of SSR tasks, though where the components of the framework occur within the information processing stream is left ambiguous in the original work. Similar to Baek and Chong’s zoom-lens model (Baek & Chong, 2020a,b), this framework assumes that attention affects the fidelity (internal multiplicative noise) of the stimulus representations. In contrast where the zoom-lens model assumes that all items contribute equally to the final SSR representation, FIM assumes that the degree to which each item will contribute is governed by linking the integrator sampling function to the fidelity of the individual representations. Localizing each of the stages of the model will provide further information on the dynamics of the averaging mechanism and provide a path towards integrating this framework into more generalized models of human memory and perception.

**Challenging the Automaticity Assumption**

While dominant theories of SSR generation suggest a rapid, early extraction of statistical codes, very little work has been done thus far to determine a locus of SSRs within the visual processing stream. The previously described zoom-lens model, for example, distinguishes contributions from progressive stages of noise and attention, though provides no description of *where* in the visual processing stream the items are integrated or stored within memory. The
model concludes only that focused attention is not necessary for generating SSRs (Baek & Chong, 2020b). Most commonly, conclusions of automaticity stem from brief stimulus presentations (Chong & Treisman, 2003; 2005a,b; 2008) with an assumption that presentation brevity, high response accuracy, and lack of apparent set-size effects logically insinuate an early mass integration of feature values that precludes the transfer bottleneck between iconic and short-term memory (Gegenfurtner & Sperling, 1993).

Our lab compared performance on classic iconic memory tasks with those of an averaging task using identical experimental procedures to determine whether this assumption is valid (Zepp, Dubé, & Melcher, 2021). Iconic memory is an early high-capacity storage locus of memory, that decays rapidly but acts as a buffer zone for information to be transferred from the visual system to durable short-term memory through attentional deployment (Coltheart, 1980). Iconic memory also acts as an integration area for information presented in rapid succession. As an “early” SSR generation mechanism is thought to create the summary rapidly, over a large amount of information, before focused attention, and through an integration process, it is logical to assume that this process may occur within iconic memory. Following this logic, if SSRs are generated within iconic memory, the information that the SSR is generated from should be subject to fundamental features of iconic memory. More specifically, the iconic store is the locus of information integration (information presented close in time is represented as a single perceptual event) and segregation (information spaced in time is perceived as a series of independent events), with a distinct temporal dependence on the relative appearance of each effect within data. Our goal was to determine whether this same pattern of temporal dependence was present within the SSR estimate, to determine the likelihood of the iconic store as a locus.
Figure 1.3: Experimental design of Zepp, Dubé, and Melcher (2021). Each task followed the same time course as shown in (A), with only the goal and response of the subject varying across tasks. The composite image in the “Frame 1 + Frame 2” overlay sections of B, C, and D represent the total stimulus set from frame 1 and frame 2. The red circles in (B) and (C) represent the target of the task, to which the subject should move the mouse to click. The red arrow in (C) represents the rotation of the central probe for the averaging response.

We presented subjects with 16 oriented Landolt-C style stimuli, divided over two frame presentations by a variable interframe interval, with an exposure duration of 16ms for each frame. In a blocked design, we then asked the subjects to complete either an integration task that required the subject to maintain a superposition of the two frames, a segregation task that required the subjects to maintain separate percepts of each frame, or an averaging task that required subjects to reproduce the average orientation of all stimuli (see Figure 1.3).

Tasks requiring the integration and segregation of rapidly presented displays have a long history for examining iconic memory phenomena, and robust characteristic patterns of
performance related to the interframe interval duration typically hold. When the interframe interval decreases frame integration becomes easier and frame segregation becomes more difficult (Sperling, 1967; Eriksen & Collins, 1967; Wutz, et al., 2016). We first sought to determine whether the integration of items into an SSR was related to integration across frames. The results showed that averaging performance was consistent across all interframe durations, while integration task performance varied as expected. These findings argued against a relationship between classic early iconic item integration and SSR generation, as the temporal dependence that defines iconic integration was not present in SSR responses.

We were then able to determine that, across all interframe intervals, subjects appeared to be making their averaging judgements based on only two of the sixteen items. This finding echoed the previously discussed work of Myczek and Simons (2008). Further, the subjects appeared to be sampling from only the most recently presented frame for all levels of interframe interval duration, except the shortest two. Within the shortest exposure durations, subjects appeared to be sampling two items randomly from either frame. Contrary to the assumptions made by Myczek and Simons, we did not have to imply that this was an intentional strategy being used by subjects. This is exactly the pattern of results that would be expected if subjects were making use of items available within visual short-term memory (vSTM), based on the duration of the stimuli and prior estimates of information transfer rates from iconic memory (Vogel et al., 2006; Sperling, 1960). This surprising result has motivated the current investigation of the role vSTM plays in SSR formation and added to the continued discussion of the automaticity of the underlying mechanism at play.
Iconic vs vSTM Item Contributions

Figure 1.4: A simplified representation of the separate modes of memory for visual information. Information is initial fed into a high-capacity iconic store, which rapidly decays. Using attention, items can then be transferred to the durable short-term memory to halt decay. Items within short-term memory can then be consciously acted upon through control processes. Information within short-term memory can then be moved to a more permanent long-term memory store. Previously stored items within long-term memory can also be recalled into short-term memory (Atkinson & Shiffrin, 1968).

In his classic work, Sperling (1960) detailed two distinct storage modalities present within the short-term store. The first, an early high capacity, but fast decaying, iconic store within which a large proportion of stimulus information is contained and available briefly (<2 seconds). The second, a later durable store that allows the portion of items successfully encoded from the iconic store to be consciously acted upon (see Coltheart, 1980 for a review). Within his initial procedure, Sperling demonstrated that a greater number of stimuli are available for reproduction when the subject is cued to certain items, versus when the subject is asked to report from the full set of stimuli. The advantage of this partial report response to that of the whole-report characteristically decayed as the time between the stimulus offset and the cue onset increased. The defined ability to sample more information accurately into vSTM for a brief time frame, when the task made a subset of the display relevant, demonstrated that there was an early persistence of information from the visual environment available for the short-term store to pull from. Subsequent reports have shown that subjects have both a selective (strategic) and non-selective (passive) ability to sample from this early store, based on the demands of the task,
which can be modeled well computationally (Gegenfurtner & Sperling, 1993; Sperling, 2018). Importantly, the full information available in the iconic store is registered prior to the onset of attentional selection (pre-attentive) and must be sampled into vSTM through either a random, uncontrolled process (non-selective) or by selectively attending to a subset of the information available to the observer. Consequently, a large fraction of the information in a scene will fail to result in conscious awareness within a single scan of the display, despite having been held in iconic storage.

To parse out the contributions of items within iconic versus vSTM stores, a common practice is to limit the encoding time available for a stimulus display with backwards masking (Coltheart, 1980; Sperling, 1960). Work within visual masking has shown that masks are most effective in halting encoding when the features of the chosen mask are of the same category as the target feature to-be-masked (Turvey, 1973; Bhardwaj, Mollon, & Smithson, 2012). An appropriately chosen mask will cause erasure of the information within the iconic store, while leaving the information within the durable vSTM store unaffected. This erasure is hypothesized to occur by overloading the mechanisms required for the transfer and maintenance of the relevant features within the icon (Turvey, 1973). Therefore, one can effectively halt the transfer of information out of the iconic store, and view only the information available to the observer within vSTM at a given timepoint, using backwards masking.

The importance of the distinction between these two storage states to the idea of SSRs comes in determining what, and how much, of the visual environment is available to the averaging mechanism. If the SSR is generated by a process that resides within the iconic store, then it potentially has a substantially large amount of fast decaying information to act on, and subsequently must generate representations rapidly to counteract the decay of iconic information.
If instead the mechanism lies within the later vSTM, then the attentional bottleneck substantially limits the amount of information available to the mechanism. This limit would imply that averages are generated from subsets of the complete display, as was shown in the previously discussed work of Zepp, Dubé, and Melcher (2021).

**Controlling for the “When” and the “What” of SSRs**

While conclusions of automaticity and unlimited processing capacity for SSRs are common, as described above, the actual control of the processing time in experiments is much less common. Nonetheless in the relatively rare instances in which the processing time has been experimentally controlled through masking, a much different picture of the mechanism has appeared.

Whiting and Oriet (2011) employed a backward mask and variable stimulus duration scheme in a 2-AFC size averaging task. Stimulus sets of 12 circles at varying sizes were presented for 0 (no circles presented), 50, 100, or 1000ms followed by a 400ms pattern mask. Afterwards, subjects were shown sets of two circle probes and asked to choose which corresponded to the average of the previously observed set. Critically, results indicated that not only were subjects unable to reliably derive the average size of the stimulus set until exposure times above 200-ms, but that they appeared to be relying heavily on previous trial information to make their judgements. This result is also consistent with predictions and experimental results of the previously described FIM model (Tong, 2020). Whiting and Oriet noted that due to regularities often present in the structure of SSR designs (i.e., using equivalent frequencies of each trial mean’s presentation throughout a study) this strategy of making guesses based on previous trial information would appear as fairly accurate generations of the SSR, leading to possibly erroneous conclusions of automatic processes in averaging. By controlling the encoding
time available to information within the iconic store through backwards masking, Whiting and Oriet demonstrated that averaging over size stimuli necessitates a much longer timeframe than previously assumed. Further, this study demonstrated that disruption in the contents of the iconic store, and maybe more importantly in the ability to transfer items into the visual short-term store from the iconic store, resulted in drastic deviations from expected early averaging performance.

In another study of backward masking effects within SSR research, Epstein, et al. (2020) employed the paradigm in conjunction with an SSR reproduction response probe to investigate the integration of outliers within orientation feature averaging. The results once again showed a clear trend relating the latency of masking to averaging performance. Additionally, beyond just showing that SSR generation is augmented by encoding time allotments, the results demonstrate that the process appears to be iterative in the way items were represented within the SSR. Concurrently, the iterative integration of stimuli shown in Epstein’s work is what one might expect if a subject was serially transferring information from the iconic store into the averaging mechanism (Sperling, 1960; Averbach and Coriell, 1961; Gegenfurtner & Sperling, 1993), though the same result may also be achieved with a parallel process that acts unevenly across information within the store.

In contrast to the results of the previously discussed studies, two recent studies by Boduroglu and colleagues on SSR generation showed no significant degradation in averaging performance when a backward mask is introduced after a short stimulus presentation time. Both studies aimed to investigate the SSR mechanism’s ability to integrate pairs of stimulus features concurrently, with the first focusing on comparing center-of-mass and centroid judgements (Boduroglu & Yildirim, 2020) and the second focusing on the ability to generate separate orientation and size averages over the same set of stimuli (Yörük & Boduroglu, 2020). The
papers each used a reproduction of the estimated average as a response, and each contained an experiment to test the time course of average generation through the implementation of a backward mask. Even at a 50ms stimulus presentation time, the accuracy of feature averages reproduced by subjects within each of the four domains did not significantly differ from the respective averages produced when masking was not implemented. These results imply that the generated feature average can be processed within the 50ms of stimulus presentation, before the mask was shown, and was passed to a durable store. This conclusion appears to conflict with the previous two reported demonstrations of masking reducing feature averaging performance.

One explanation for the discrepancy between these results may come in the details of how the masks are chosen. In the first two experiments on center-of-mass and centroid judgements, the backward mask was characterized by a clustered pattern of many overlapping lines at varying orientations. The next two experiments, with orientation and size as the relevant features, employed a random noise mask. Previous research into the efficiency of backwards masking on encoding disruption has found that when the mask and stimulus are dissimilar in feature space, the stimulus resists the mask and appears as a degraded copy of itself within the mask (Turvey, 1973; Bhardwaj, Mollon, & Smithson, 2012). This leads to the possibility that the conflicting results may come down to the characteristics of the masks used. Pattern masks that are defined by identical features to that of the target stimulus feature likely provide a superior effect to that of masks defined by non-target features or noise. This indicates that the masking present in Yörük and Boduroglu’s work may not have been strong enough to erase relevant information present within the iconic store, leading to misleading conclusions. Further investigation is necessary to determine whether this explanation explains the discrepancy.
The Current Work

The current work accomplished two overarching goals. The first was to determine the temporal and spatial characteristics of SSR generation. As was described in the introduction, there exists considerable debate regarding how much information contributes to SSRs, and how rapidly SSR generation occurs. The first experiment gave insight on the amount of information that contributes to an averaged representation and the structure of the mechanism used to sample the items. The second experiment investigated the speed at which SSRs are generated and the effect of limiting encoding time on summarization. Using known properties of the visual processing stream and structure of various modes of memory, this experiment provided insight on the debate of whether SSRs are generated early, from information within the iconic store, or late, on items that have been sampled into vSTM.

The second goal was the expansion of current models of SSRs to include the collected experimental results into an account of “what” is averaged and “when” the averaging mechanism occurs. The model I formulated gives a description of the relationship between information within the iconic store and that of the short-term store, as well as relative fidelity of the features within the displays at each timescale. Combined, these results advance averaging models beyond some of the current hurdles in theory and provide a clearer understanding of summary representations.
A Brief Introduction to Circular Statistics

As the feature of interest is orientation, care must be taken to account for the biasing effect of periodicity in directional data when summarizing the average feature value. To account for periodicity, circular statistical estimates will be used throughout the rest of this work. Therefore, a brief introduction to the calculation of these adjusted measures is warranted. Unless otherwise stated, the mean and precision parameters within this paper will be calculated as follows.

As the data are cyclical on \([-180^\circ, 180^\circ]\), the common normal distribution with infinite tailed supports will not be appropriate for fitting. A common approximation to the normal distribution for circular data is the Von-Mises distribution. The probability density function of this distribution is:

\[
    f(\theta | \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} e^{(\kappa \cos(\theta - \mu))},
\]  

(2.1)

where \(\theta\) is a set of angles (in radians), \(\mu\) is the circular mean of the set of angles, \(\kappa\) is the concentration parameter, and \(I_0\) is a modified Bessel function of order 0. This distribution has support on \(\theta \in [-\pi, \pi]\) (or any support of length \(2\pi\)). The mean and concentration parameters will be discussed below.
The first statistic is given by a circular mean, which is used in place of the arithmetic mean to account for the biasing effect of periodicity in directional data summarization. The circular mean is given by the equation:

$$
\mu_c(\theta) = \text{atan2}\left(\frac{1}{N} \sum_{j=1}^{N} \sin \theta_j, \frac{1}{N} \sum_{j=1}^{N} \cos \theta_j\right),
$$

(2.2)

which converts each orientation angle ($\theta_j$, in radians) of a stimulus set ($\theta$) from polar coordinates on the unit circle to cartesian coordinates, averages over each set of cartesian coordinates, and then converts the averaged cartesian coordinates back to polar space using the 2-argument arctangent of the coordinates. To demonstrate the reasoning of this choice, consider the average of the angle set [0°, 360°]. The arithmetic average would be 180° but, as circles are periodic and 0° is equivalent to 360°, a more appropriate average value would be the circular average, 0°. Each set of 2 angles will always have 2 mean values; one using the "long" arc between two points (i.e. the central point of the arc wrapping around the whole circle to get from 0° to 360°) and one "short" arc (i.e. the non-existent distance between 0° to 360°). Throughout this work, the shorter arc measurements will always be the arcs of interest and the circular mean calculation will be used to average over these arc values. Though the calculation is different, for the purposes of this work the interpretation of the circular mean is equivalent to that of the arithmetic mean.

The concentration parameter of the Von-Mises distribution ($\kappa$) can be found by solving:

$$
\frac{I_1(\kappa)}{I_0(\kappa)} = \frac{\|\Sigma_j^N \theta_j\|}{N},
$$

(2.3)

Where $I_1$ and $I_0$ are modified Bessel functions of the first kind at orders 1 and 0, respectively. $\theta_j$ represents a set of orientation angles. The solution can be approximated by (Sra, 2012):
\[ \hat{\kappa} = \frac{\bar{R}(2-\bar{R}^2)}{1-\bar{R}^2}, \text{where } \bar{R} = \frac{\sum_j \theta_j}{N}. \] (2.4)

When \( \kappa = 0 \), the distribution is uniform around the circle and as \( \kappa \) tends towards infinity, the distribution becomes a point estimate around the mean orientation value. For large \( \kappa \), the Von-Mises distribution approximates the normal distribution, with \( \kappa \approx 1/\sigma^2 \).

As a measure of performance within this work, focus will be placed on the spread of the subject’s error distribution (i.e., the spread of the difference between the subject’s judgements and the true value of the display). Deviations from 0° in the mean value of the error distribution may be interpreted as a systematic propensity to over/underestimate the true value. Given a sufficient number of estimation opportunities, this mean error bias is generally very close to zero for continuous orientation value reports across experimental design manipulations. The spread of a subject’s estimation errors (the precision of the estimate error distribution, \( \kappa \)) varies with experimental manipulation (Zepp, et al., 2020) and can be interpreted as an estimate of representational fidelity given task demands.

To discuss and display subject performance via the error distribution spread more naturally, precision (\( \kappa \)) will be used instead of the variance. Precision ranges between 0 (errors are uniform around the circle and there is no true mean value), and infinity (all errors are exactly equal). For the current work, as the value of precision increases, subject performance is also increasing.

Moving forward, parameter estimates from the Von-Mises distribution will be used to describe SSR responses. As the stimulus space is periodic, Von-Mises distributions are a natural choice for representing individual item representation distributions, though the SSR is an average over these item distributions. While the SSR distributions appear to be shaped as Von-Mises
distributions, mathematically the average of Von-Mises distributions is not itself a Von-Mises distribution. The resulting distribution is complex and has not been defined well enough to determine a functional relationship between the item distribution parameters and the averaged distribution. Given the growing use of circular distributions within cognitive research, a future formalization of the averaged Von-Mises distribution will be required for accurate descriptions of models over periodic representations.

**Orientation Averaging Analysis**

The data for the orientation averaging task for each experiment were analyzed using a Bayesian linear mixed-effects distributional regression of the form:

\[
SSRerror_{i,j,r} \sim VM(\mu_{i,j}, \kappa_{i,j}), \\
\mu_{i,j} \sim \sum_{j=0}^{J}(s_{i,j}^\mu + \beta_{i,j}^\mu X_j), \\
\log(\kappa_{i,j}) \sim \sum_{j=0}^{J}(s_{i,j}^\kappa + \beta_{i,j}^\kappa X_j), \\
(2.5)
\]

\[
\beta_{i,j}^\mu \sim N(0, \sigma), \\
\beta_{i,j}^\kappa \sim N(0, \sigma).
\]

Here, \(SSRerror_{i,j,r}\) is the error, in radians, of the orientation averaging response for subject \(i\) given experimental condition set \(j\) on trial \(r\). Errors are drawn from a Von-Mises distribution, with mean \(\mu_{i,j}\) and precision \(\kappa_{i,j}\). The parameter estimates of the Von-Mises distribution, from which the subject’s error responses are drawn, are modeled as a linear summation of the estimated effect coefficient for each subject \(\beta_{i,j}\) with random effects on the slopes and intercepts \(s_{i,j}\) for the fixed-effect condition \(X_j\). As the precision parameter for the Von-Mises distribution can only be positive, a log-link function is used for the estimation of the precision parameter.
Normal priors for both the intercept and effect estimates are used, centered around 0 with equivalent $\sigma$ values. This allows for the vectorization of the priors, which greatly speeds up computation time.
CHAPTER THREE:

EXPERIMENT 1: THE “WHAT” OF SSRS

The goal of this experiment was to determine the contents of what is passed to the averaging mechanism. Specifically, whether all items from a display contribute to the SSR estimate, or only a subsample of items within the display. To accomplish this, subjects were asked to either reconstruct as many items from the stimulus frame as they can (a whole-report task) or to report the average orientation feature value of the frame (an averaging task). The tasks were completed in separate blocks over equivalent stimulus displays, and the average of the items reported in the whole-report procedure was compared to the average feature value reported in the averaging task. The distribution properties of the stimuli, and the locations of each item in the display, were controlled to allow for comparisons of the responses between the two conditions.

The whole-report task is one which requires the subject to respond with the subset of item information that they have been able to transfer from the fleeting iconic store to the durable short-term store (Sperling, 1960). If subjects were similarly using a subsampling strategy to feed items into vSTM, where they are averaged into an SSR, then the subsample reported in the whole-report task should be predictive of the response in the averaging task over an equivalent stimulus frame. Importantly, the exact same stimulus sets were presented to a subject across both tasks (though the stimulus configurations were varied across subjects). This one-to-one matching
of what was shown between the two tasks, with only the response varying, allowed for a direct comparison to be made between the whole-report subsample and the SSR response.

It may be the case that even with equivalent stimulus frames the subsampling behavior present within the whole-report and averaging tasks may not be equivalent, due to differing task demands. The structure of the subsampling strategy could be dependent on consistently sampling certain locations within the display (the central items over the peripheral items for example, which may be predicted by a higher fidelity for items overtly attended to (Tong, 2020)). If instead the sampling mechanism was distinct between the two tasks, and therefore the specific items sampled within the whole-report task did not map well onto the ones sampled for the averaging task, then a separate distinct pattern of results will be present within the various stimulus configurations. Several models of different sampling strategies were considered in the analysis. Combined, this analysis provides insight on the mechanism of SSR construction.

**Subjects**

20 participants participated in this study. 60% of subjects were female, 30% were male, and 10% were non-binary. Subjects ranged in age from 18-26 (µ = 19.15, σ = 1.95). All subjects were recruited from the USF SONA subject pool. All participants had normal to corrected normal vision.

**Design**

The experiment consisted of two tasks: a whole-report task to report as many items as the subjects can remember, and an averaging task in which subjects were asked to recreate the average orientation value of the full display. Each task was presented in 2 separate blocks each, consisting of 224 trials within each block, totaling 448 trials per task per subject. Blocks were
counterbalanced to reduce order effects, and subjects were asked to take a short break between each block to reduce fatigue. Each subject completed all conditions. Each trial lasted about 2 seconds, and the complete experiment took about 1.5 hours to complete.

**Stimuli**

Stimuli varied monotonically along the orientation dimension, bounded between -90° and +90° with 0° being vertical “up”. Subjects were seated approximately 100 cm from the center of the display. Sixteen black arrows (each subtending 0.5° of visual angle) were organized into a 4x4 imaginary grid of equally spaced presentation locations, which was positioned centrally on the screen. Each element was spaced 0.5° of visual angle away from the next element, measured center-to-center. The full grid of arrows had a length and height of 3.5° of visual angle. The stimuli were presented against a uniform gray background at 50% of the monitor’s RGB range ([127 127 127]).

Individual positions of stimuli within each grid square were slightly jittered on each trial. The range of orientations varied between two levels: a narrow range of µ±15° and a broad range of µ±30°. Four unique orientation mean values (µ= -60°, -30°, 30°, 60°) were chosen for each display around which the 16 stimulus orientation values were uniformly spaced around, creating 8 unique stimulus distributions (4µ x 2 ranges). The feature values were distributed equally around the true mean value, though the mean feature value was never part of the display feature set. The orientation feature value of a particular arrow that appears within the central 4 locations of the 4x4 grid was structured to fit into one of 4 different categories (see Figure 3.1):
Figure 3.1: An example of the 4 orientation feature location structural configurations. The central 4 elements, which are the controlled elements for each structure type, are highlighted in the green box. A sample set of 16 orientation feature choices, centered here around a mean value of 0°, is included.

- (1) The central 4 elements were selected to be the 4 feature values closest to the frame mean; central 4 items are low variance and centered around the frame mean.
- (2) The central 4 elements were selected as the 4 values with the furthest positive distance from the frame mean; central 4 values are low in variance and centered around a value that is systematically larger than the frame mean.
- (3) The central 4 elements were selected as the 4 furthest endpoints of the distribution; central 4 values are high in variance and centered around the frame mean.
- (4) The central 4 elements were randomly selected from the pool of 16 orientation feature values, as is common in most SSR experiments within the literature.
Table 1. An example of the stimuli presented for the first and last 3 trials within each block for two subjects.

<table>
<thead>
<tr>
<th>Block</th>
<th>Task</th>
<th>Trial 1</th>
<th>Trial 2</th>
<th>Trial 3</th>
<th>Trial 222</th>
<th>Trial 223</th>
<th>Trial 224</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td><strong>Subject 1</strong></td>
<td></td>
<td></td>
<td><strong>Subject 2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Averaging</td>
<td><strong>Stimulus #1</strong></td>
<td><strong>Stimulus #2</strong></td>
<td><strong>Stimulus #3</strong></td>
<td><strong>Stimulus #28</strong></td>
<td><strong>Stimulus #28</strong></td>
<td><strong>Stimulus #24</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mu= 60, Range= 60,</td>
<td>Mu= -30, Range= 60,</td>
<td>Mu= 60, Range= 30,</td>
<td>Mu= -30, Range= 60</td>
<td>Mu= -30, Range= 60</td>
<td>Mu= 60</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Random Set, Positive</td>
<td>Random Set, Positive</td>
<td>Random Set</td>
<td>Positive Set</td>
<td>Positive Set</td>
<td>Positive Set</td>
</tr>
<tr>
<td>2</td>
<td>Whole-Report</td>
<td><strong>Stimulus #1</strong></td>
<td><strong>Stimulus #2</strong></td>
<td><strong>Stimulus #3</strong></td>
<td><strong>Stimulus #28</strong></td>
<td><strong>Stimulus #28</strong></td>
<td><strong>Stimulus #24</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mu= 60, Range= 60,</td>
<td>Mu= -30, Range= 60,</td>
<td>Mu= 60, Range= 30,</td>
<td>Mu= -30, Range= 60</td>
<td>Mu= -30, Range= 60</td>
<td>Mu= 60</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Random Set, Positive</td>
<td>Random Set</td>
<td>Random Set</td>
<td>Positive Set</td>
<td>Positive Set</td>
<td>Positive Set</td>
</tr>
<tr>
<td>3</td>
<td>Whole-Report</td>
<td><strong>Stimulus #23</strong></td>
<td><strong>Stimulus #6</strong></td>
<td><strong>Stimulus #26</strong></td>
<td><strong>Stimulus #15</strong></td>
<td><strong>Stimulus #24</strong></td>
<td><strong>Stimulus #20</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mu= 30, Range= 60,</td>
<td>Mu= -60, Range= 60,</td>
<td>Mu= 30, Range= 60,</td>
<td>Mu= -60, Range= 60</td>
<td>Mu= -60, Range= 60</td>
<td>Mu= 30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Central Set, Positive</td>
<td>Positive Set</td>
<td>Positive Set</td>
<td>Random Set</td>
<td>Random Set</td>
<td>Random Set</td>
</tr>
<tr>
<td>4</td>
<td>Averaging</td>
<td><strong>Stimulus #23</strong></td>
<td><strong>Stimulus #6</strong></td>
<td><strong>Stimulus #26</strong></td>
<td><strong>Stimulus #15</strong></td>
<td><strong>Stimulus #24</strong></td>
<td><strong>Stimulus #20</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mu= 30, Range= 60,</td>
<td>Mu= -60, Range= 60,</td>
<td>Mu= 30, Range= 60,</td>
<td>Mu= -60, Range= 60</td>
<td>Mu= -60, Range= 60</td>
<td>Mu= 30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Central Set, Positive</td>
<td>Positive Set</td>
<td>Positive Set</td>
<td>Random Set</td>
<td>Random Set</td>
<td>Random Set</td>
</tr>
</tbody>
</table>

Note that the same order of stimulus frames was presented across each pair of task blocks within a subject. The order of stimulus presentations was randomly reshuffled within the second pair of blocks (blocks 3 and 4) relative to the first pair of blocks, though the same set of 32 stimuli were used. Stimuli were recreated independently for each subject, such that “Stimulus #1” for Subject 1 is not the same as “Stimulus #1” for Subject 2.
Each orientation value that is chosen to occupy a location is selected from the feature distribution, without replacement. This means that structuring the set of orientation values that occupy the central 4 positions will also constrain the surrounding 12 values (e.g. if the central 4 items were chosen to be the 4 items closest to the frame mean, the surrounding 12 grid positions would contain items that are all farther from the frame mean than any of the central 4 items; see Figure 3.1 above).

The set of 4 mean value conditions, 2 range value conditions, and 4 structural configurations provided 32 unique stimulus frames per subject. The unique frames were repeated 7 times, in a randomized order within a block. The same set of stimulus frames was used for both the averaging and whole-report task blocks, with the presentation order of each frame matched between tasks but randomized between the first and second pair of task presentations (see Table 1 above).

Procedure

Subjects first completed a series of practice trials to become familiar with each task before the first block that starts each task. To begin, a still image of full timeline of the trial progression (similar to Figure 3.2) was displayed to the subject. Subjects were instructed that each stimulus varied between -90° (left from center) and +90° (right from center) from 0° (vertical), and that all stimulus arrows pointed some degree of “upwards”. The subjects were then presented with an example stimulus frame paired with an example response frame. For the averaging task, the example response frame displayed the correct average of the stimulus frame. The whole-report task example response frame displayed a 4x4 grid of oriented arrows, a subsample of which have been adjusted to match the orientation values in the relevant locations from the example stimulus frame. The practice trials then began with a slow presentation of each
display and progressed in speed until reaching the same display duration present in the experimental trials. After completing a set of practice trials, the experimental trials immediately followed (see the Appendix for the full instructions presented to subjects.)

![Figure 3.2](image)

*Figure 3.2:* An example of the Experiment 1 trial progressions. The response on these tasks varied by block and were either a recreation of the mean orientation value or a whole-report of as many individual items as possible.

Each trial began with a fixation cross presented in the center of the screen for 300ms. After an interval jittered between 200-400ms, the stimulus frame was presented for 48ms. As the temporal properties of the procedure are not the object of this experiment, the displays remained unmasked. After a 300ms post-stimulus offset delay, subjects were prompted for a response.

In the averaging condition, subjects received an arrow in the center of the screen that they rotated to match what they believed was the average orientation of the previous display. Responses were entered by rotating the central arrow from a randomly chosen starting value (sampled from a uniform distribution between -180° and 180°) via moving the mouse up and
down and were locked in by pressing the space bar. Subjects’ responses were unrestricted, and they could respond with any orientation value between -180° and 180°.

In the whole-report procedure, subjects received a 4x4 grid of dots in the same positions as the arrow stimuli on the previous display. If a subject remembered an arrow’s orientation at a given location, subjects clicked on the dot in the location of the remembered item. Clicking on the dot would cause an arrow to appear in the clicked location, with an initial orientation of 0°. Responses were recorded by clicking each arrow and dragging the mouse up and down until the arrow was at the desired orientation. Subjects then pressed the space bar to lock in their response, once they reported as many items as they could. The order of report for each stimulus location was recorded, in addition to the reported orientation value.

Subjects in both tasks were instructed to respond as quickly and accurately as they can, though an unlimited response duration was allotted. The trial proceeded once the subject pressed the space bar.

**Orientation Task Analysis**

Results for the orientation averaging task were analyzed using a Bayesian linear mixed-effects distributional regression of the form shown in *equation 2.5*. Error on each trial was expressed in radians within the regression analysis. Fixed effects for each condition, as well as all interactions, were modeled with random slopes and intercepts. The intercept, to which other conditions are compared, was the low range and random configuration condition. Normally distributed extremely weakly informative priors, with \( \mu=0 \) and \( \sigma=30 \), were set on all effect estimate and intercept parameters.
Figure 3.3. Circular histograms of the orientation averaging response error distributions for each condition. Each column is a different stimulus configuration, and each row is a range condition. Longer blue bars indicate more concentration around that degree value. The average error value for each condition is noted with a dashed red line. Estimate average error degree value and precision estimate are listed below the histograms.

The model was estimated using the BRMS package (Bürkner, 2017,2018) in R (R Core Team, 2022), with 4 chains of 10,000 samples each and a burn in period of 1,000 samples. All chains were visually checked for convergence, and the Gelman-Rubin statistic (Gelman and Rubin, 1992) $\hat{R} = 1.00$ for all estimated parameters. Parameter estimates with 90% posterior sampling credible intervals that exclude 0 are discussed. These intervals indicate that there is a 90% chance that the true effect estimate is non-zero, given the evidence from the data. The variance partition coefficient (VPC) is also calculated using the performance package in R (Lüdecke, et al., 2021) and reported. This value is an alternative to the intracluster correlation coefficient (ICC) for models that use random slopes and intercepts, which indicates the proportion of variance accounted for by the random effects clustering structure.
Table 2. Results from the Bayesian mixed-effects regression for Experiment 1.

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Estimates</th>
<th>Est. Error</th>
<th>CI (90%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>μ Predictors</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-0.01</td>
<td>0.01</td>
<td>-0.02 – 0.00</td>
</tr>
<tr>
<td>Range</td>
<td>0.02</td>
<td>0.01</td>
<td>0.00 – 0.04</td>
</tr>
<tr>
<td>Central Configuration</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00 – 0.02</td>
</tr>
<tr>
<td>Positive Configuration</td>
<td>0.07</td>
<td>0.01</td>
<td>0.05 – 0.08 *</td>
</tr>
<tr>
<td>Endpoint Configuration</td>
<td>0.00</td>
<td>0.01</td>
<td>-0.01 – 0.02</td>
</tr>
<tr>
<td>Range: Central Configuration</td>
<td>-0.03</td>
<td>0.01</td>
<td>-0.05 – 0.00</td>
</tr>
<tr>
<td>Range: Positive Configuration</td>
<td>0.05</td>
<td>0.01</td>
<td>0.03 – 0.08 *</td>
</tr>
<tr>
<td>Range: Endpoint Configuration</td>
<td>0.01</td>
<td>0.02</td>
<td>-0.03 – 0.05</td>
</tr>
<tr>
<td><strong>Log(κ) Predictors</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>2.44</td>
<td>0.10</td>
<td>2.27 – 2.61 *</td>
</tr>
<tr>
<td>Range</td>
<td>-0.31</td>
<td>0.07</td>
<td>-0.42 – -0.19 *</td>
</tr>
<tr>
<td>Central Configuration</td>
<td>-0.05</td>
<td>0.08</td>
<td>-0.18 – 0.09</td>
</tr>
<tr>
<td>Positive Configuration</td>
<td>0.02</td>
<td>0.07</td>
<td>-0.09 – 0.12</td>
</tr>
<tr>
<td>Endpoint Configuration</td>
<td>-0.14</td>
<td>0.07</td>
<td>-0.25 – -0.03 *</td>
</tr>
<tr>
<td>Range: Central Configuration</td>
<td>0.11</td>
<td>0.10</td>
<td>-0.05 – 0.27</td>
</tr>
<tr>
<td>Range: Positive Configuration</td>
<td>0.11</td>
<td>0.11</td>
<td>-0.06 – 0.29</td>
</tr>
<tr>
<td>Range: Endpoint Configuration</td>
<td>-0.10</td>
<td>0.12</td>
<td>-0.29 – 0.10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random Effects</th>
<th>Estimates</th>
<th>CI (90%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>σ² Without Random Effects</td>
<td>0.12</td>
<td>0.10 – 0.14</td>
</tr>
<tr>
<td>σ² With Random Effects</td>
<td>0.15</td>
<td>0.15 – 0.16</td>
</tr>
<tr>
<td>VPC</td>
<td>0.19</td>
<td>0.05 – 0.32</td>
</tr>
<tr>
<td>N Subjects</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>8960</td>
<td></td>
</tr>
</tbody>
</table>
Results

The results show that the average error in orientation SSR report was significantly biased positively away from 0 when the central set of features was also positively biased away from the true mean of the frame (estimate = 0.07, 90% CI: [0.05 – 0.08]), as compared to when the feature distribution of the display was randomly distributed. The average response error for the positive configuration was additionally biased from 0 when the range of the stimulus was increased (estimate = 0.05, 90% CI: [0.03 – 0.08]). Changing the range alone did not lead to a significant change in the average SSR error, and similarly no other stimulus configuration had a significant biasing effect on the average SSR error.

The precision of the error distribution was significantly affected by the range of the feature distribution on a stimulus frame, with the precision of responses decreasing as the range of the display rises (estimate = -0.31, 90% CI: [-0.42 – -0.19]). The only feature configuration condition that significantly affected the precision of the errors was when the central elements of
the display were sampled at the endpoints of the frame feature distribution and centered around
the true mean value (estimate = -0.14, 90% CI: [-0.25 – -0.03]). The intercept of the precision
parameter was also non-zero, though this is no surprise as the kappa intercept is just the precision
of the low range and random configuration conditions (estimate = 2.44, 90% CI: 2.27 – 2.61]). It
is important to make clear that a log-link function was used for the estimation of the kappa effect
coefficients, and therefore would need to be exponentiated to convert back to the scale that kappa
is generally on.

The data showed significant variation between subjects, with 19% (90% CI: [0.05 –
0.32]) of the variance being accounted for by the random effects clustering of the subject
individual differences.

Discussion

The results of the mixed effects modeling are about what were expected from the original
hypotheses. It is key to reiterate that for each subject, the same stimulus feature sets were used
between each of the stimulus configurations. The only change between configurations was the
locations of features from the common feature distribution. If all items were equally weighted in
their contribution to the average orientation representation, then there should be no effect of
feature locations on the SSR error distributions as all items would be represented. Nonetheless,
biasing the center of a scene to only contain values from the tail of the scene feature distribution
concurrently biased the response of the subjects towards the tail. Similarly, making the center of
the scene contain the endpoints of the feature distribution increased the spread of errors in SSR
estimation. Increasing the range of the feature distribution also decreased the precision of
responses, regardless of the feature-location configuration (see Figure 3.3). These results imply
that features in the display are at the very least not equivalently weighted in the SSR estimation
process, and that a location-based factor plays some role in the feature contribution weights (Tong & Dubé, 2022).

An interesting result, contrary to the hypothesized results, was that fixing the center of the display to be tightly clustered around the center of the feature distribution did not lead to a reduction in the spread of errors relative to a random location structure for the features in the display. One explanation for this lack of effect may be that there is an upper bound on the precision of ensemble representations. Support for this idea comes from studies which have compared displays in which all features in a display are homogenous versus displays where items were heterogenous (Ariely, 2001; Chong & Treisman, 2003; Haberman & Whitney, 2007, 2009; Haberman, Harp & Whitney, 2009). Results from these experiments show that averaging performance in the homogeneous condition is greater than that of the heterogeneous condition, though the difference approaches an asymptote with decreasing early noise. This indicates a limit to the degree in which increasing the similarity between features can increase the fidelity of the average representation. For the corollary, the results of the current study indicated that increasing the dissimilarity between features in the display decreased averaging precision. Increasing dissimilarity influenced precision both when increases occurred across the full feature distribution, by increasing the range, or simply within the center of the display, by placing the feature display endpoints in the center.

Whole-Report Response Characteristics

One hypothesis for the results obtained within the orientation averaging task, in which equivalent displays produced different patterns of results based only on swapping where features were placed in the display, is that subjects were averaging over a subset of items in the display. Contrary to more common evaluations of subsampling models in ensemble averaging, where
subsamples of a specific size are either randomly drawn from a display (Myczek & Simmons, 2008; Chong et al., 2008; Im & Halberda, 2013; Maule & Franklin, 2016) or are drawn from a set of salient items (Chong et al., 2008; Kanaya et al., 2018; Iakovlev & Utochkin, 2021), this work makes the subjects their own subsample generator. This is done through asking the subjects to complete a whole-report task, over the same stimuli that would be presented within the orientation averaging task. Doing so allowed for the subsampling mechanism to both capture variation in the number and locations subsampled by each subject, and therefore will not need to make any assumptions about either factor.

We will begin by looking at the characteristics of responses for the whole-report task, in which subjects were asked to recreate as many items from the previous display of oriented arrows as they could remember. The subject could respond with anywhere between 0 items (no response) and 16 items (responding to all item locations). Each response would have its own error from the item in the response position that the subject was attempting to recreate.

Results

Figure 3.5 below displays the probability distribution of whole-report sizes that were reported. Subjects were most likely to report between 2 to 4 items on average. The whole-report task requires items to be passed into vSTM for report, and therefore performance is limited by an individual’s information transfer bottlenecks and memory capacity (Sperling, 1960). The retrieved whole-report size distribution is well within what would be expected given the brief stimulus presentation. The distribution of whole-report sizes was not affected by either a change in the range of the stimulus display feature distribution, or a change in the configuration of the distribution of the central items (see Figure 3.6 and Figure 3.7 below, respectively).
Figure 3.5. Probability distribution of the number of items reported in the whole-report task.

Figure 3.6. Probability distribution of the number of items reported in the whole-report task during each of the two range conditions.
Figure 3.7. Probability distribution of the number of items reported in the whole-report task during each of the four stimulus configuration conditions.

### Proportion of Item Responses by Location

<table>
<thead>
<tr>
<th># of Items Reported</th>
<th>Central</th>
<th>Positive</th>
<th>Endpoint</th>
<th>Random</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.03</td>
<td>0.05</td>
<td>0.03</td>
<td>0.01</td>
</tr>
<tr>
<td>2</td>
<td>0.03</td>
<td>0.23</td>
<td>0.14</td>
<td>0.01</td>
</tr>
<tr>
<td>4</td>
<td>0.03</td>
<td>0.22</td>
<td>0.13</td>
<td>0.02</td>
</tr>
<tr>
<td>6</td>
<td>0.02</td>
<td>0.03</td>
<td>0.01</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Figure 3.8. A heatmap displaying the proportion of responses given in each of the 16 possible response locations.
Further characterizing the subjects’ response characteristics on the whole-report task, 
*Figure 3.8* displays an overwhelming bias for subjects to respond in the central four positions of 
the response frame. Responses were particularly concentrated within the upper half of the central 
four elements within the display. When a response is given at a certain location on the response 
screen, the subject indicated that they remembered the item from the stimulus frame that was in 
the same location. All items were presented within the range of foveal vision, and therefore 
would be expected to have approximately equivalent spatial saliency. While subjects were 
instructed to respond with as many items as they could remember, the overwhelming dominance 
of the center on response locations is almost certainly the compounded effect of the prompt to 
fixate on the center of the display and limited number of remembered items (as displayed in 
*Figure 3.5*). Once again, the location report distributions were relatively identical between the 
two range conditions (*Figure 3.9*), and the four stimulus configuration conditions (*Figure 3.10*).

### Proportion of Item Responses by Location and Range

<table>
<thead>
<tr>
<th>Low Range</th>
<th></th>
<th>High Range</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.03</td>
<td>0.03</td>
<td>0.02</td>
<td>0.03</td>
</tr>
<tr>
<td>0.05</td>
<td>0.22</td>
<td>0.24</td>
<td>0.03</td>
</tr>
<tr>
<td>0.03</td>
<td>0.13</td>
<td>0.14</td>
<td>0.23</td>
</tr>
<tr>
<td>0.01</td>
<td>0.02</td>
<td>0.02</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>0.01</td>
<td>0.01</td>
<td>0.14</td>
</tr>
</tbody>
</table>

*Figure 3.9.* A heatmap displaying the proportion of responses given in each possible 
response locations, segmented by the range condition.
**Figure 3.10.** A heatmap displaying the proportion of responses given in each possible response locations, segmented by the stimulus configuration condition.

It should be noted that, for the single subject who occasionally reported all 16 items, the response patterns on trials where all 16 items were reported changed from a central focused pattern to one of a serial left-to-right-top-to-bottom structure. This separate pattern also appeared in pilot testing of this experiment on trials where subjects reported all 16 items, though in both situations the centralized response pattern overwhelming dominated among subjects. The left-to-right bias can still be seen to a lesser degree within the subjects who responded almost exclusively centrally, such that first responses were most common in the central-top-left location, followed by the central-top-right location. It is unclear whether this behavior is just a response bias or an attentional bias (for a discussion see Dickinson & Intraub, 2009).
As stated previously, one explanation for the central bias may be the instruction to attend to the center of the screen, where the fixation cross was shown. This instruction was given to make sure that subjects were able to see the full display, given the short exposure duration. Similarly, asking subjects to attend to the center of the display was done to try and ensure that each stimulus position appeared in a similar spot retinotopically. Nonetheless, it is possible that the instruction to attend to the center of the display may have narrowed the attentional window and contribute to the central bias in the whole-report data. An instruction to attend centrally is common in studies of visual perception and memory, though when investigating effects of attention, each instruction from the experimenter should be considered when evaluating results.
CHAPTER FOUR:
WHOLE-REPORT AS A SUBSAMPLING MODEL

Averaged Whole-Report Model

![Diagram](image)

*Figure 4.1.* Model of the relationship between ensemble averaging and whole-report response. For both tasks, a subset of items are drawn from the iconic store via a probabilistic non-selective transfer into vSTM. For a whole-report task, items are drawn from vSTM to be recreated at the response screen. For the averaging task, the items that occupy vSTM are fed into a control process that averages over the contents within memory. The averaged representation is then fed into the response system.

A hypothesized model of SSR generation is that subjects are using an average over a subset of feature values that have been sampled into vSTM to summarize the display. Sampling occurs through a non-selective mechanism that is shared between the orientation averaging and whole-report tasks. The logic of this model is presented in *Figure 4.1*. Noisy information stored in iconic memory is sampled into vSTM through a passive, but centrally biased, non-selective transfer mode of attention (i.e. no cueing to any individual item is presented, and therefore
attentational sampling for durable storage occurs over as many items as possible). For the SSR task, information that makes it into vSTM is then acted on by a control process within vSTM that averages over stored items and outputs the ensemble estimate. For the whole-report task, the subject passes as many items as possible to the response system for reproduction.

Importantly, within this model the two tasks draw directly from the same limited storage locus and therefore should share the same encoding-based biases. Specifically, if certain items are more likely to be sampled into vSTM than others, characteristics of these sampled items will appear in both the whole-report and SSR task responses. If this model is valid, we can use the responses from the whole-report task as a representation of vSTM contents. As the SSR report is an average over the items within vSTM, we can instead average over the whole-report responses as a representation of vSTM items and compare the distribution of these averaged response values with the SSR task response distribution. If both reports share a common subset of items, similar biases should appear between the SSR distribution and the averaged whole-report distribution.

Viewing the error distribution of the averaged whole-report subsamples in Figure 4.2 gives a superficially similar pattern of results to those presented in the orientation averaging task. Average error biasing is present in the positive configuration, with the amount of bias being exacerbated by an increase in stimulus feature distribution range. The endpoint distribution also has the lowest precision when the range is high. The simple model is not perfect though. Most notably, the precision of the averaged whole-report distributions is greater for every condition. One major contributing factor to this is certainly that late noise is being cancelled out in averaging, and therefore a source of noise present in the orientation averaging results is not present in the averaged whole-report distributions. Regardless, the major effects present in the
orientation averaging data persist in the whole-report average model, giving credence to the idea that the two responses are indeed related and warrant deeper investigation.

![Figure 4.2](image)

**Figure 4.2.** Circular histograms of the orientation averaging task error distributions with averaged whole-report response error distribution model fits. Model fits are shown as red lines within each plot. The blue histograms are the orientation averaging task error distributions from Figure 3.3. The average orientation task error value for each condition is noted with a dashed black line. Estimate average error and precision for the SSR task are listed with the $o$ subscript, and parameter estimates for the averaged whole-report model are listed with the $w$ subscript.

**Whole-Report as a Model of Non-Selective Transfer**

While averaging over the whole-report responses provides promising results, a cleaner method that allows for more control over sources of noise in the model would be to use the responses from the whole-report task to inform a representational model of the spatial non-selective sampling function. To elaborate, instead of using the subsample of actual responses in
each location from the whole-report task, we instead use only the locations that the subject sampled. The true feature values from these locations are then used, instead of the subject’s recreation of the feature value from each sampled location. The precision of the individual item representations and response biases can then be estimated and evaluated. Once again, the assumption is that the link between the whole-report task responses and the orientation averaging response is that they both are generated from a common subset of items sampled into vSTM. Therefore, common abnormal distribution behavior between the two tasks should be due to the vSTM contents, and more specifically the items that are sampled for encoding.

Following this logic, a location-based subsample model was created. In this model, the subset of response locations from the whole-report task was used to create vectors of item feature values present within vSTM. Each sampled feature value was then used as the average orientation value of an independent Von-Mises distribution. For simplicity, each item sampled into vSTM was assumed to be represented with equivalent fidelity, given by the Von-Mises precision parameter $\kappa$. The common item precision parameter estimate is the only free parameter in this model, with a best fitting value of $\kappa = 4.3$. As shown in Figure 4.1, representations within vSTM were then averaged and a single value from the averaged distribution was sampled as a response.

Figure 4.3 displays the fit results of the location-based subsample model. The fits for this model are much better, and all major effects are preserved from the orientation averaging data. While there are some discrepancies in the estimates of the precision parameters, particularly for the random configuration, the shape of the distributions predicted by the model are very similar to those of the subject data.
One interesting prediction of the model was that the precision of errors in the central configuration would remain mostly unaffected by increases in range. Items within the central configuration that occupy the center of the display are chosen to be the closest grouping of feature values around the true display mean feature value. With increasing range, these stimuli also increase in distance from one another though continue to be the closest grouping around the display average. As the subject data do show an effect of stimulus range on precision for the central configuration, it may be the case that orientation stimuli are encoded on a non-linear scale such that small increases in the difference between two angles are magnified when translated from the physical scale to the perceptual feature scale (see Schurgin, Wixted, & Brady, 2020;
though also see Tomić, & Bays, 2022). Overall, the single free parameter location-based subsampling model, derived from subject whole-report responses, provides an excellent fit to the subject orientation averaging error distributions.

The success of the whole-report task-based models to replicate the distributions of the SSR responses gives support to the idea that SSRs are formed over items within vSTM and not over the full display feature distribution that is present within iconic memory. It should be restated again that each configuration condition in Experiment 1, within each of the two range conditions, had the same feature distribution. Only the locations of the features within the distributions were intentionally arranged. If an average was generated over all items, then each configuration would be expected to have the same response distribution. This is not the case, and the deviations between the configuration conditions are well explained by just the items sampled into vSTM during the whole-report task and a certain level of noise. This is not the only possible model though, and we will now explore some other popular theories of SSR generation.

**Alternative Models**

**Automatic Averaging Model**

The most common idea of SSR formation is that averaging occurs automatically and early over all items within the display. A representation of this theory, placed within a normal memory structure is shown in *Figure 4.4*. In this model, very noisy representations of items within the iconic store are averaged through an early compulsory process which can then feed the early representation to vSTM for response. The key aspect of this model is that the averaged representation is formed using information from all items within the display.
Figure 4.4. Alternative model where averaging occurs early over all items in the iconic store. This model simulates more popular early models of averaging, where all items are sampled early with equal contribution to the SSR report. Averaging is automatic in the iconic store and is fed into vSTM to be reported.

Like the location-based subsampling model, we can simulate this model with a single free shared precision parameter between item feature distributions centered on the item’s true feature value. The distributions are then averaged, and a value is sampled for the response. For this model, the best fitting item representation distribution precision parameter was $\kappa = 1.2$. Results are displayed in Figure 4.5.

As we can see in Figure 4.5, the assumption of averaging over all items provides a worse fit to the data than the previous vSTM based models. As noted previously, averaging over all items in the display would lead to similar distributions over all conditions that would not capture the pattern of biases present within the orientation averaging data given the configuration or range conditions. Each distribution is roughly centered around 0° with similar precision, and variations in parameter estimates are due primarily to randomness. The model’s lack of ability to capture biases arising from item configurations in the display, and range of the feature distribution, gives strong evidence that at the very least all items are not represented equally.
within the SSR representation (Tong & Dubé, 2022; Iakovlev & Utochkin, 2021; Tong, 2020; Epstein et al., 2020; Kanaya et al., 2018)

**Figure 4.5.** Circular histograms of the simulated early averaging model response error distributions for each condition. Item representational precision was fit to $\kappa = 1.2$ for each item distribution. Model fits are shown as red lines within each plot. The blue histograms are the orientation averaging task error distributions from Figure 3.3. The average orientation task error value for each condition is noted with a dashed black line. Estimate average error and precision for the SSR task are listed with the $o$ subscript, and parameter estimates for the early model fits are listed with the $w$ subscript.

**Selective Sampling Model**

The most common way that subsampling models are represented for comparison with other models in literature is that of a strategic selective subsample of the display. Usually, the subsample is assumed to be taken from the most salient region of the display, and usually this assumed to be the central items in the display. In this model formulation, noisy item
representations are drawn from iconic memory through strategic selective attention transfer process. Representations are then averaged through a control process before being fed to a response system. This model is very similar to the whole-report based models presented in the previous section, except the sampling procedure is selective and more consistent.

![Diagram](image)

*Figure 4.6. An alternative model where the central four elements of the display are sampled for averaging. This model assumes a consistent selective transfer model that samples only central elements. Items are then stored in vSTM and averaged through a control process that feeds to a response system.*

The model simulation procedure is like that of the previous models, where noisy representations with a common precision parameter are averaged, except this model only averages over the items that were presented in the central four positions in the display. The best fitting precision parameter value for the item feature distributions within this model was $\kappa = 2.7$. A similar model where only the upper two central locations from the set of four central locations were sampled was also simulated (with $\kappa = 5$) though the two models had similar results, therefore only the central-four sampling model will be discussed.
The results for the central-four items subsampling model are shown in Figure 4.7. This model does a better job of capturing the mean bias present in the positive configuration, though it appears to overestimate the bias in comparison to the orientation averaging responses. The effect of the endpoint configuration was also not captured particularly well, though it is represented more than the early sampling model over all items. The effect of range on reducing the precision of SSR responses is mostly lost in this model, and interestingly increasing the range also increased the model’s estimated precision for the central and positive configuration conditions. As adjusting the features that appear in the central four locations of the display were the defining...
characteristics of the configuration conditions, the current results are not much of a surprise. If a subject was only strategically sampling the central items and neglecting all other items, we would expect to see only a real effect of the configuration that biased the mean of the display: the positive configuration. The other two structured configurations, the central and endpoint configurations, were centered around the true mean just with tighter and wider spreads respectively. The poor quality of this model’s fit to the data highlights that the subjects were not completely consistent in their sampling, even though they were more likely to sample the center of the display than outside the center.

**Early + Late Averaging Model**

![Diagram](image)

*Figure 4.8.* Alternative model which bridges the gap between early and late models. In this model, averaging occurs both automatically over all items in vSTM as well as later within vSTM. Items are selectively transferred into vSTM and averaged together with the automatically generated average representation, weighting sampled vSTM items more heavily in the response.

A compromise between the central item selective transfer model and the early averaging model over all items is a model that performs both averaging processes to form an estimate. As shown in *Figure 4.8*, it is possible that the averaging response occurs early in memory over noisy items within the iconic store in parallel with attention selectively transferring a select few items
into vSTM. The early average estimate would then enter vSTM alongside the sampled set of items. The late averaging control process will then average over both the items in vSTM and the early averaged estimate. This would create an SSR estimate that both takes the full feature distribution from the display into account, while giving higher weighting to items that are selectively sampled. Often within early averaging models, there will be an aside that mentions that unequal weights can be incorporated into the model for specific items though it is rare that the unequal weighted models are evaluated or simulated. This current model aims to simulate this possibility.

**Figure 4.9.** Circular histograms of the simulated unequal weighting model response error distributions for each condition. Item representational precision for central items was fit to $\kappa = 2$, and precision for non-central items was fit to $\kappa = 1$ for each item distribution. Model fits are shown as red lines within each plot. The blue histograms are the orientation averaging task error distributions from **Figure 3.3**. The average orientation task error value for each condition is noted with a dashed black line. Estimate average error and precision for the SSR task are listed with the $o$ subscript, and parameter estimates for the unequal weighting model fits are listed with the $w$ subscript.
The results for the unequal weighting model fits are shown in Figure 4.9. The model slightly captures the mean bias of the positive configuration, as well as the effect of range, though to a much lesser degree than the orientation averaging data. The effect on precision within the endpoint configuration condition is non-existent. This is expected, as the central items within the endpoint configuration are centered around the true frame mean value, but with a high range around the feature distribution. Upweighting the central items would only lead to more precise clustering around the true feature value in comparison to equal sampling of all items. The small mean bias in the positive configuration, relative to the orientation averaging data, arise strictly from the upweighted central items. Adding the non-central items into the averaging mechanism greatly reduces the ability of these upweighted items to influence the SSR as more information from the full feature distribution is present. The more information from the full feature distribution that is added to the SSR estimate, the less item based biasing effects will be present in the SSR estimate. This model sits between the selective subsampling model (see Figure 4.6) and the automatic averaging model (see Figure 4.4) and inherits properties from both models based on the degree of difference between central and non-central item weights.

Discussion

Overall, the whole-report based sampling models, particularly the location-based subsampling model, outperformed the alternative models. As the whole-report task responses are drawn from items within vSTM, the success of this task at predicting SSR estimates gives evidence that SSRs are also generated from items within vSTM. A worse fit in the central sampling model demonstrates that while the SSR may be generated from vSTM contents, it is not generated from a strategic subsample of the display. The stochasticity of non-selective transfer present within the whole-report task is a probable way that items are sampled for SSR
estimation. As both tasks ask for responses based on the full display, it is not surprising that the sampling strategy is similar for both tasks. Further, assuming that both an early and late averaging process may be occurring also leads to a worse fit than a pure late averaging process (see Figure 4.9). While this unequally weighted model is more flexible than the early (Figure 4.4) or late (Figure 4.6) models, the representation of all items either adds more noise to the average to try and capture the bias or reduces the noise to try to capture the averaged noise. The level of noise in the subject data, as well as the bias present in the positive configuration condition, imply that less than all items are represented in the SSR. Together, the set of biases present within the orientation averaging error distributions and the collection of model simulation results give strong evidence for an SSR generation process that occurs over items non-selectively transferred into vSTM. Further investigation into the validity of this conclusion is given within Experiment 2.
CHAPTER FIVE:

EXPERIMENT 2: THE “WHEN” OF SSRS

This experiment was aimed at determining the temporal locus of the averaging mechanism, thought to generate the SSRs in explicit averaging tasks, within the visual information processing stream. To do this, a backwards mask was used with varying stimulus presentation durations to give a time-course of feature averaging precision relative to interruption of iconic contents. With longer delays before the mask is deployed, subjects had more opportunities to encode items into VSTM from the iconic store. If feature averages were automatically available to subjects, there should be no expected difference in precision among presentation times. If instead VSTM was required for averaging, then a characteristic pattern of performance increases should be seen with increasing presentation time, up to a timepoint where vSTM reaches a capacity limit. Following the previously described work of Whiting and Oriet (2011), the distribution from which the stimulus trial mean was chosen was also manipulated between blocks to map out reliance on previous trial information in the response of the current trial. If subjects were relying on previous trial information to generate their estimates, then responses would be more accurate when the current trial mean was close to the cumulative mean across trials within the block (Tong & Dubé, 2022; Whiting & Oriet, 2011). Therefore, within the present design, subjects were expected to respond with higher accuracy when the trial means are normally distributed, as opposed to when they are chosen from a bimodal distribution.
Subjects

25 participants participated in this study. 48% of subjects were female, 48% were male, and 4% were non-binary. Subjects ranged in age from 18-20 ($\mu = 18.64, \sigma = 0.7$). All subjects were recruited from the USF SONA subject pool. All participants had normal to corrected normal vision.

Design

This experiment was organized as a 2 (masking: masked vs unmasked) x 5 (stimulus duration: 16ms vs 32ms vs 48ms vs 96ms vs 256ms) x 2 (distribution: bimodal vs normal) within-subjects design. 25 Subjects participated in the experiment. Subjects completed a total of 960 trials each, with each trial lasting approximately 2 seconds on average. The total experiment lasted about 1.5 hours.

Stimuli

Stimulus design was identical to Experiment 1. The range of orientations was restricted to $\pm 24^\circ$ of the current trial mean value. The backwards mask was a randomized pattern mask of overlapping oriented lines and extended over all stimulus locations which were uniquely generated for each masked trial.

Procedure

Subjects first completed a series of practice trials to become familiar with the task and rapid stimulus presentation. To begin, subjects were shown a still image that displayed the full timeline of the trial progression (see Figure 5.1). Subjects were instructed that each stimulus varied between -90° (left from center) and +90° (right from center) from 0° (vertical), and that all
stimulus arrows pointed “up”. The subjects were then presented with an example stimulus frame paired with an example response frame that denotes the correct average of the stimulus frame. The practice trials then began with a slow presentation of each display and progressed in speed until reaching experiment speeds. Once adequate performance in the practice trials was achieved, the experimental portion began.

Figure 5.1: An example of the Experiment 2 trial progressions. The oriented arrows in the stimulus frame were either distributed normally or bimodally around a mean orientation value, for a variable amount of time. Afterwards, either a backward pattern mask or a blank frame was shown, followed by a reproduction response.

The conditions were split into 2 sections, each consisting of 2 blocks of 240 trial repetitions of each combination of masking and trial mean distribution. The conditions were crossed between masked vs unmasked trials and bimodally distributed trial means vs normally distributed trial means. Subjects were encouraged to take a short break between blocks. Each subject completed all conditions, for a total of 240 trials per combination of mean distribution and masking.
Each trial began with a fixation cross presented in the center of the screen for 300ms. After an interval jittered between 200-400ms, the stimulus frame was presented for either 16, 32, 48, 96, or 256ms. If SSRs are generated within vSTM (Coltheart, 1983; Gegenfurtner & Sperling, 1993; Sperling, 1967), these durations may correspond to an estimated transfer of 1, 2, 3, 6, or 16 items from the icon to the durable store, though this relationship between the stimulus frame duration and an exact number of discrete items is not particularly important for the predictions or conclusions of the current work. This instead acted as an informed choice of time values for an approximate expectation of information within vSTM over time, with a hope of capturing both fine details of information transfer at lower stimulus durations and limiting behavior at higher stimulus durations. After the stimulus offset, either a mask was presented for 300ms or the screen persisted blank until response. The response display consisted of a single arrow, oriented at a random initial orientation, presented in the center of the screen. By moving the mouse, subjects rotated the test probe until it matched their memory for the average of the previous display’s stimulus orientations. The subject then pressed the space bar to submit their response, and the next trial began after a 100ms-250ms jittered blank screen.

Analysis

Orientation averaging results were modeled similarly to the results from Experiment 1. All random slopes and intercepts were included in the model. The intercept comparison conditions for the categorical variables in this model were the unmasked and normally distributed distribution levels. 90% credible intervals that do not contain 0 are discussed below.
Results

The results show that the average SSR error is slightly biased away from 0 for the frame masking and overarching distribution configuration interaction (estimate = -0.03, 90% CI: [-0.06 - -0.01]) indicating a possible effect of a rolling average over trials on subject responses when the frames are masked. It should be noted that while the credible interval for the distribution parameter alone contained 0 (estimate = 0.02, 90% CI: [0.00 – 0.03]), the 90% HDI for the parameter did not contain 0 (90% HDI: [0.01 – 0.03]). While both the main effect and interaction for the distribution parameter are weak effects, the combination of the two effects could be interpreted as a slight positive bias in the average SSR error when the overarching distribution of trial means is bimodal instead of normally distributed. This difference between distributions mostly disappears in masked trials. No other parameter estimate’s credible interval excluded 0.

Figure 5.2. Highest density intervals for Experiment 2 model parameter estimate posterior distributions. The 90% HDI shows the smallest interval within which 90% of the posterior distribution is contained for each parameter value estimate.
Table 3. Results from the Bayesian mixed-effects regression for Experiment 2.

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Estimates</th>
<th>Est. Error</th>
<th>CI (90%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>μ Predictors</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>-0.01</td>
<td>0.01</td>
<td>-0.02 – 0.00</td>
</tr>
<tr>
<td>Mask</td>
<td>0.02</td>
<td>0.01</td>
<td>0.00 – 0.04</td>
</tr>
<tr>
<td>Time</td>
<td>0.02</td>
<td>0.03</td>
<td>-0.02 – 0.07</td>
</tr>
<tr>
<td>Distribution</td>
<td>0.02</td>
<td>0.01</td>
<td>0.00 – 0.03</td>
</tr>
<tr>
<td>Mask:Time</td>
<td>-0.05</td>
<td>0.05</td>
<td>-0.13 – 0.04</td>
</tr>
<tr>
<td>Mask:Distribution</td>
<td>-0.03</td>
<td>0.02</td>
<td>-0.06 – -0.01*</td>
</tr>
<tr>
<td>Time:Distribution</td>
<td>-0.01</td>
<td>0.04</td>
<td>-0.08 – 0.06</td>
</tr>
<tr>
<td>Mask:Time:Distribution</td>
<td>0.02</td>
<td>0.06</td>
<td>-0.09 – 0.13</td>
</tr>
<tr>
<td><strong>Log(κ) Predictors</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>2.16</td>
<td>0.10</td>
<td>2.01 – 2.32*</td>
</tr>
<tr>
<td>Mask</td>
<td>-1.07</td>
<td>0.08</td>
<td>-1.21 – -0.94*</td>
</tr>
<tr>
<td>Time</td>
<td>0.61</td>
<td>0.23</td>
<td>0.24 – 0.97*</td>
</tr>
<tr>
<td>Distribution</td>
<td>-0.03</td>
<td>0.08</td>
<td>-0.15 – 0.10</td>
</tr>
<tr>
<td>Mask:Time</td>
<td>4.78</td>
<td>0.52</td>
<td>3.94 – 5.62*</td>
</tr>
<tr>
<td>Mask:Distribution</td>
<td>0.11</td>
<td>0.12</td>
<td>-0.07 – 0.30</td>
</tr>
<tr>
<td>Time:Distribution</td>
<td>0.13</td>
<td>0.33</td>
<td>-0.40 – 0.66</td>
</tr>
<tr>
<td>Mask:Time:Distribution</td>
<td>-0.26</td>
<td>0.64</td>
<td>-1.31 – 0.79</td>
</tr>
<tr>
<td><strong>Random Effects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>σ² Without Random Effects</td>
<td>0.19</td>
<td></td>
<td>0.15 – 0.24</td>
</tr>
<tr>
<td>σ² With Random Effects</td>
<td>0.26</td>
<td></td>
<td>0.25 – 0.26</td>
</tr>
<tr>
<td>VPC</td>
<td>0.25</td>
<td></td>
<td>0.04 – 0.40</td>
</tr>
<tr>
<td>N Subjects</td>
<td>25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>24000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 5.3: The effect of stimulus frame presentation time and presence of a mask on the precision of the SSR error distribution. The precision for masked trials is shown in red, while the precision of unmasked trials is presented in blue. Increases in the y-axis value correspond to increases in the precision of subject responses in relation to the true SSR estimates. 95% Confidence intervals are shown for each precision estimate for each frame time value.

Discussion

The demonstration of a strong interaction effect of backwards pattern masking and stimulus presentation time on the response precision supports the hypothesis that SSRs are generated over time, rather than immediately. The use of the mask was intended to interrupt the encoding time of items within the stimulus frame, and therefore the detrimental effect of the mask on SSR performance indicates that the encoding process of the average can also be interrupted. If the SSR estimate was generated rapidly or automatically the interruption of item encoding would not influence the response precision, and no difference in the masked and unmasked conditions would be expected. While the grain of the frame duration interval is relatively sparse, Figure 5.3 shows that the difference in the error distribution spread between the
masked and unmasked conditions becomes negligible somewhere within the interval of 96ms and 256ms.

As was previously shown by Chong and Treisman (2003), increasing the presentation time only minimally increased the accuracy of responses for the unmasked stimuli. When compared to the decreased precision at lower presentation time intervals in the masked condition, the small effect of time in the unmasked condition indicates that encoding continues for a time after the stimulus frame is removed unless interrupted. Previous experiments on the effect of presentation time on SSR accuracy without adequate control on the encoding time are used to justify the idea of automaticity in SSR generation, though when these controls are instituted the dependence of SSR precision on encoding time becomes clear.

The effect of the overarching distribution when the trial is masked on the average bias of the SSR response is unclear, both in origin and whether it is a meaningful effect due to the small magnitude. As both the bimodal and normal distributions are symmetric around a grand average value, any effect of the change in distribution was expected to appear in the error precision. The precision of the error distribution was not affected by the distribution choice, when masked or unmasked. One explanation for the bias may be an expression of late response biases in uncertainty. The mask increases uncertainty in the current trial estimation response by decreasing encoding time, and the bimodal distribution condition makes each trial more likely to be distinct from the previous trial when compared to the normally distributed condition. The distinctiveness of the trials may lead to less of a bias towards the previous trial response, and more expression of late biases in the response when the estimate is already uncertain. Similar slight biases are found in Experiment 1 (Figure 3.3), as well as in our previous work on similar oriented stimuli with continuous responses (Zepp et al., 2021).
CHAPTER SIX:
MODELING ENCODING TIME EFFECTS

Simulating Subsamples

Continuing the modeling discussion from Experiment 1, we can extend the location-based non-selective subsampling model to also account for the results of Experiment 2 by incorporating temporal effects into the model. To begin, as a whole-report task was not part of this experiment, we need to generalize the non-selective subsampling element of the whole-report task beyond just the actual responses from Experiment 1. While constructing an unequal probability sampling function for the transfer function is beyond the current work (see Hanif & Brewer, 1980 for a review), an algorithm can be generated to create an arbitrary number of subsample sets. To do so, the conditional probabilities of sequential report are generated from the whole-report data of Experiment 1 (see Figure 6.1.)

Once the trials are simulated for Experiment 2, the same procedure from Experiment 1’s location-based subsampling model is used to simulate SSR estimates (see Figure 4.1). True feature values from each of the items in the subsampled locations are used as the average orientation values for a Von-Mises distributed item representation. For simplicity, the precision parameter estimate is equivalent for each item representation distribution. The sampled items are then averaged and reported as the SSR estimate.
Figure 6.1. An example of simulating the sampling function Experiment 1 whole-report data. This figure shows an example of one iteration of the sampling simulation algorithm for a single trial, but the same process is repeated within a trial until all available item locations are sampled or an item fails to be sampled. The same procedure is repeated from the start an arbitrary number of times to simulate the desired number of trials.

<table>
<thead>
<tr>
<th>First Report Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
</tr>
<tr>
<td>0.04</td>
</tr>
<tr>
<td>0.03</td>
</tr>
<tr>
<td>0.02</td>
</tr>
<tr>
<td>Stop</td>
</tr>
</tbody>
</table>

A location is sampled, based on the selection probabilities, and the location’s corresponding feature value is added to the stored item vector.

<table>
<thead>
<tr>
<th>First Report Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
</tr>
<tr>
<td>0.04</td>
</tr>
<tr>
<td>0.03</td>
</tr>
<tr>
<td>0.02</td>
</tr>
<tr>
<td>Stop</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Second Report Probability, Given First Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
</tr>
<tr>
<td>0.05</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>Stop</td>
</tr>
</tbody>
</table>

The probability of sampling an item from a certain location is generated. “Stop” indicates the probability that an item is not sampled, terminating the process.

The previously sampled item is removed from the available sample space, and a new set of conditional probabilities for sampling a new item, given the first sampled item is generated.

This process is repeated iteratively until the “Stop” item is sampled, or all 16 available locations are sampled. Note that the probability of sampling the “Stop” value grows as iterations increase. This creates the subsample of items for one trial. This process is repeated from the start for however many trials are needed.

**Precision as a Function of Encoding Time**

An assumption is made for the interpretation of the previous results, and the modeling process, that the mask limits encoding time of items in the display. An unmasked display
therefore continues encoding attempts to some degree until the response screen appears. Therefore, the encoding time available for unmasked displays is the time that the stimulus frame was displayed plus the 300ms delay until the response screen appears. For unmasked displays, encoding is assumed to continue for some time after the offset of the stimulus display. This is the general interpretation of the effect of backwards pattern masking in respect to iconic memory transfer (Sperling, 1960; Turvey, 1973; Coltheart, 1983), and has some support from neural recordings (Rolls et al., 1999). The results from Figure 5.3 can be expressed on an x-axis scale of encoding time instead of just frame duration to give a clearer picture of the pattern of results (Figure 6.2).

![Graph: Orientation Average Error Precision by Masking and Encoding Time](image)

*Figure 6.2. Precision of orientation averaging as a function of available encoding time. The lines are segmented by the masking condition, and 95% confidence intervals are shown for each precision estimate.*
Model of Encoded Information

The pattern of results in Figure 6.2 shows what appears to be an asymptote in precision improvement with increasing encoding time. This pattern of asymptotic increase in performance with increasing encoding time is common within literature on vSTM transfer. In an evaluation of iconic memory transfer Gegenfurtner and Sperling (1993) defined a function to model this asymptotic pattern as a function of the form:

$$N(t) = C_N[1 - e^{-(t/\tau)}], \quad (6.1)$$

where $N(t)$ is the number of items transferred given encoding time $t$. $C_N$ is the capacity of vSTM, and $\tau$ is a constant that determines the rate at which the function will approach the asymptote. The value of $\tau$ is the time at which $N(t)$ is approximately equal to 63% of $C_N$. Importantly, this model is based on discrete item responses (i.e. a report of which letter was presented) and discrete item representations. Given that the current work focusses on a continuous distributional model of memory representations, a translation between discrete and continuous representational structure is warranted.

Reinterpreting the discrete encoding function in terms of the continuous response data, the transfer function can be written as:

$$K(t) = K_A[1 - e^{-(t/\tau)}], \quad (6.2)$$

where $K(t)$ is the item representation precision given encoding time, $t$. $K_A$ is the asymptotic value of item representational precision, given extended encoding time, defined on $(0,\infty)$. $\tau$ is a time constant which determines the rate at which encoding time adjusts the item precision, defined on $(0,\infty)$. 
The full expression of the subsampling model up to this point is given by:

$$SSR_i = atan2\left(\frac{1}{N_i} \sum_{j=1}^{N_i} \sin \tilde{\theta}_{ij}, \frac{1}{N_i} \sum_{j=1}^{N_i} \cos \tilde{\theta}_{ij}\right),$$

$$\tilde{\theta}_{ij} \sim VM\left(\theta_{ij}, K(t)\right),$$

$$(6.3)$$

$$K(t) = K_A[1 - e^{-(t/\tau)}],$$

where $i$ is the trial number, within which $N_i$ items are encoded through a non-selective transfer function. $j$ is an item within vSTM, such that $\tilde{\theta}_{ij}$ is the memory representation of the encoded item $j$ on trial $i$. The value of random variable $\tilde{\theta}_{ij}$ is sampled from a Von-Mises distribution with a mean value equal to the true feature value of item $j$ within trial $i$, $\theta_{ij}$. The precision of representation $\tilde{\theta}_{ij}$ is determined by the transfer function $K(t)$, which is dependent on the available time for encoding, $t$, and an encoding rate parameter, $\tau$. The upper asymptote of item precision at extended encoding times is $K_A$. The non-selective transfer process is simulated by the procedure given in Figure 6.1.

Using this model, we can simulate the results for Experiment 2. Generally, $K_A$ would be a free parameter, though as the stimulus design of Experiment 1 was identical to that of the unmasked condition of Experiment 2 and the 48ms frame time (348ms of available encoding time) is within the asymptotic region of Figure 6.2, we can use the item precision fit value from the Experiment 1 ($\kappa = 4.3$). This leaves $\tau$ as the only free parameter, which was estimated to be $\tau = 77$. The model fit is shown in Figure 6.3 below.
Discussion

The addition of the simple model of encoding into the subsample model provides a fairly accurate fit to the shape of the data, using only one free parameter to adjust the information encoding rate. It is interesting to note that the original formulation of this model by Gegenfurtner and Sperling (1993) was for a completely different design and assumed discrete representations in storage, item representations, and transfer. The present reinterpretation shows that the same encoding function translates beyond only discrete representations, providing a link between earlier models of discrete memory and more modern distributional representations of items in memory. Further merging of older models of memory with modern theoretical developments is likely to provide similarly fruitful results.
While the ability to successfully model the shape of results with a single free parameter is encouraging, it is certainly an oversimplification of the information encoding process. A few deviations within the fit give insight on future directions for follow-up research needed to develop the current model. A decrease in precision within shorter frame durations of the unmasked stimuli, relative to the asymptote, indicates an unaccounted effect of exposure duration present within the data. One likely source of this discrepancy is a functional link between the decay rate of information available for encoding and the exposure duration, such that as stimulus duration increases the strength of the representation increases and more time will be needed for the representation to decay. An experiment that holds the stimulus exposure duration constant and adjust mask onset time would allow for a clearer picture of the decay of information over time.

A more interesting deviation present within the current results is that at the lowest encoding time value, the precision of subject response distributions is significantly greater than what would be expected from the model. The model predicts that precision should drop to 0 as time also goes to 0, but subject data is about 0.8 units higher than expected at 16ms. A model that moves the intercept of the model on the y-axis would statistically account for this deviation, though the idea that precision does not decrease to 0 in the limit is nonsensical in the context of theory.

One hypothesis is that the most difficult perceptual situations result in responses made from informed guessing, resulting in a lower bound of performance dictated by the quality of the information used in the guess. If this were the case, a main effect or interaction effect of the trial mean distribution configuration (which modulated the informativeness of a rolling average
across trials as a facilitating factor for SSR estimation) would have been expected in the Experiment 2 regression results (*Table 3*).

An alternative explanation may be that there is an initial burst in the encoding rate immediately coinciding with stimulus presentation that regularizes quickly to the form shown in the current model fit. It is plausible that the visual system attempts to quickly register a very small amount of stimulus information rapidly to indicate that a stimulus was presented, leading to a steep jump away from a precision of 0. Evaluation of this hypothesis would simply require carrying out a follow-up experiment at lower exposure durations to determine the gradient of the encoding rate.

A final possible explanation arises in simplifications made to the sampling function. For simplicity, the same sampling function is used across all duration and masking conditions, with only the common precision parameter varying with encoding time. Sampled items are assumed to be processed with equal resources and in parallel using this structure, though equivalent parallel processing of features is not necessarily the true structure of encoding nor is this debate a particular focus of this work. As the sampling function was generated from responses to unmasked displays with 48ms exposure durations (348ms available encoding time), any abnormalities that arise in sampling behavior with decreasing encoding time will not be captured. For example, it is possible that available encoding time modulates the deployment of attentional resources needed for items to be sampled or resolved to a strong enough precision level deemed informative by a transfer system.
Figure 6.4. An example implementation of a sampling procedure that is further limited by encoding time. A single prediction was generated for the shortest encoding time (16ms) from the experimental design, as an encoding time limitation is likely to be most relevant at the lower temporal bounds. The model assumed a 2-item limit at 16ms, and that the sparse distribution of attention leads to a doubling of each item’s precision value in vSTM. The estimated precision value from the unadjusted subsampling model is shown as a green square, and the adjusted encoding time limited model is shown as the 16ms black circle.

Following this reasoning, a small test to determine if making item sampling behavior time dependent at low exposure durations may be able to explain the limiting behavior. For the masked 16ms duration condition, the model was adjusted to have a maximum limit of 2 item locations able to be encoded by only retaining the first two simulated sampled item locations for each relevant trial. As the assumption is that these two items are receiving all available attention, the item precision for each of these items was doubled to model extra available resources available for resolving these items. The adjustment to the model greatly improved the fit of the predicted precision value for the relevant data point, giving a proof of concept for the theoretical
behavior of encoding in a limited environment. As an empirical investigation of sampling with encoding limitations was not conducted, further conclusions and fits related to this model will be left to future experiments.
CHAPTER SEVEN:

CONCLUSIONS

Information and the SSR

The presented work demonstrates a series of results that argue in favor of a “late” SSR generation framework which occurs over a subset of items that have been transferred to vSTM. Experiment 1 was centered around resolving debates within the literature regarding the amount of information that contributes to SSR estimates. While the dominant theory present within the area is that SSRs are generated across all information in a display, the results of Experiment 1 reject this theory. The orientation averaging task results demonstrate that subject SSR estimates can be predictably biased, and precision can predictably be reduced, by simply adjusting where certain features appear in the display. The key aspect of this result is that the feature distribution present within each display was equivalent, at least within each range condition. Only the spatial arrangement of features defined the stimulus configuration condition, with the range conditions acting to exasperate the effects of the stimulus configuration condition. Even with equivalent feature distributions, significant differences between subject error distributions were generated. If the prominent interpretation of SSR generation was indeed true, that summarization occurs rapidly and over all items, then rearranging where items are in the display would have no effect, as the distributions being summarized were equivalent. The finding that SSR estimates can be
predictably biased over equivalent feature distributions is incompatible with the idea that SSR generation occurs over all items in a display.

Experiment 2 provided additional support for the theory that SSRs are computed over the subset of the display that has been encoded into vSTM. Limitations placed on encoding time, or the transfer of items into vSTM, for a stimulus display led to a decrease in representational fidelity (see Figure 6.3). The exponential decay pattern present in the averaging task response precision is exactly what would be expected if item information encoding into vSTM is being disrupted by the presented backwards pattern mask.

The combination of results presented in this work provide clear evidence that SSRs are generated over information within vSTM, and that the precision of generated SSR responses is dependent on item encoding time.

**Experimental Design Contributions**

Each experiment within the current work employed methods that are uncommon within the literature on SSR generation, but that were very informative for investigating the underlying structure of orientation averaging.

Within Experiment 1, the success of using whole-report task responses to model responses on the orientation averaging task provides a clear framework for future experiments to investigate SSR characteristics as well as future computational model generation. The responses in the whole-report task provided a much more comprehensive view of a non-selective transfer function that provided a successful fit to SSR responses. The use of identically distributed feature distributions, with experimentally manipulated feature locations, was also very informative for drawing conclusions of which items were likely contributing to the averaging
mechanism. Most current models of SSRs assume that the error distribution of responses will always be centered around 0. Biases in the average are not accounted for, and focus is instead solely placed on the precision of the error distribution. As the results in the positive configuration condition of Experiment 1 demonstrate, average error biases are indeed present and predictable in SSR estimates. Future models of SSRs will need to incorporate an explanation for how the model could predict such biases if a subsampling model is not used.

Within Experiment 2, the use of visual backwards pattern masking demonstrated the importance of experimentally limiting processing time if one is to make conclusions based on the time course of a cognitive process. Previous SSR research with unmasked displays have concluded automaticity in the generation of SSR estimates, though this work displays clear evidence that this is not the case using backwards pattern masking. Future work that focusses on the speed of SSR estimation should employ similar constraints.

Future Directions

The current work sets up a body of follow-up experiments that can be conducted to provide clarity on current results and extend the theory further. The most logical continuation of the current work is a mixture of the design from Experiment 1 and Experiment 2, such that an orientation averaging task and a whole-report task are conducted over variable exposure durations, which are masked or unmasked. This experiment would provide further information on how the sampling function generated by the whole-report responses changes over time and how any changes relate to the distribution of results in the orientation averaging task. Using the sampling function from the unmasked stimuli presented within Experiment 1 to fit the results of Experiment 2 provided a good fit to the data, though deviations at the lowest encoding duration interval indicates that there may be a time dependence present in the sampling function.
Investigating the design of Experiment 1 at different encoding time intervals would provide clarity on this anomaly.

In the unmasked conditions of Experiment 2, a deviation from expected asymptote levels of performance was observed in shorter stimulus duration conditions. This likely indicates that there is an unaccounted-for relation between item exposure duration and the amount of information available to be encoded. This relationship could be determined by using a partial report task in conjunction with the design of Experiment 2, where either a single item is cued to be reported (partial report task) or subjects are asked to report the average of the display, over equivalent displays. The partial report task would provide clarity on the evolution of single item representational noise which could then be used to inform the item precision parameter in the subsampling model.

Throughout the modeling present in this work parallel processing and encoding of each item into vSTM was assumed, such that each item in vSTM shared a common precision parameter. This choice was made for convenience, and this structure need not be the case, though it is a common assumption made within SSR models. Clear biases are present within the whole-report data on which items from the display will be reported first at the response screen, and the implications of these report order biases on the SSR are unclear in the current work. Future work would be needed to determine a functional relationship between the sample serial order and relative representational fidelity, if an asymmetry in item precision exists. The sampling algorithm within the current work, and the procedure for generating the algorithm using the whole-report task, already generates an ordered sample for each trial that would naturally allow for an integration of such a function into the model.
The current work focused completely on orientation as a target feature to make inferences about broader processes in memory and perception. While no significant differences are theoretically expected to occur in different parameter spaces, beyond relative timescales, it is possible that different feature spaces may be represented on different scales within memory and therefore may have different considerations that need to be addressed. Validation of the presented models and theory within different feature spaces would aid in determining the generalizability of the current conclusions.

Implications for Perception and Memory

A non-selective transfer based summarization process has several key implications for the structure of percepts and memory. If summaries are based on a set of items in vSTM, then the system is likely interpolating the perception of the rest of a scene from the set of consolidated information. This may provide a perceptual explanation for failures in change detection (Simons & Ambinder, 2005) and inattentional blindness (Simons, 2000). In both cases, salient non-target features may gain priority for transfer to vSTM and therefore bias perception. This principle may be seen in vehicular accidents, such that even when the driver is attentive certain objects may gain priority for vSTM transfer. These objects generate the driver’s “gist” of the scene around themselves but may not include otherwise key features around themselves, such as a pedestrian. As only portions of scene information are represented in the summary percept, the driver may have truly not perceived the pedestrian even though the pedestrian was present in the scene. The competition for representation in the summary representation may bias perception to missing such key information.
The effects described at the perceptual level have branching implications for the structure of memories. If the vSTM summary representation is based on a few key items, then the LTM trace of the general scene would likely also be based on this sample of information from the scene. As an example, one might think of a movie they watched several years ago and remember it as an amazing film. Upon rewatch, they realize that the movie was mediocre though their memory of the event had been shaped by a few key good moments that were generalized to summarize the film in memory. The memory of the movie was an interpolation from a few salient bits of information. For the design of media that is meant to persist in someone’s memory, such as an advertisement or a call-to-action, salient moments are likely to frame the lasting perception of the full piece and should be carefully considered.

Concluding Remarks

Harkening back to a statement made in the introduction of this work, the acceptance of a new processing mechanism requires that the new mechanism be necessary to describe observed phenomena. Throughout this investigation, a series of results were presented that argue against a special new “early” automatic summarization mechanism. The generation of SSRs can be well explained as a simple control process operating over items that are encoded into vSTM. Considering the SSR as an estimate of a limited number of samples from a population feature distribution, many findings in the averaging literature become clear and easily explainable. Why does increasing the range of the feature distribution decrease the precision of ensemble estimates? Because the probability of obtaining an unbiased sample decreases with increasing population variability. Why does increasing the set size of items in the display increase the precision of SSR judgements, to an asymptotic limit? Because the probability of obtaining a sample of values close to, or approximately centered around the mean, will also increase to a
limit. While it is still possible that these findings may indeed be explainable by a new perceptual averaging mechanism, skepticism should be applied to claims of novelty when simpler explanations exist.
REFERENCES


92


APPENDIX A: EXPERIMENTAL INSTRUCTIONS

Experiment 1 Instructions

Practice trials:

In this experiment, you will be tasked with making responses on quickly displayed screens of oriented arrows. You will complete one of two tasks at a time, each in separate blocks. We will begin with a quick overview of what the display will look like in your first task and then move on to a few practice trials.

Each trial will begin with a small cross presented in the middle of the screen. When you see this cross, fixate your eyes on its center and hold them there while the screen with arrows is presented, until the response screen appears. The next screen you will be shown will contain 16 arrows, each pointing in a particular orientation. Each individual arrow will always either point to the left, right, or upwards to some degree. The screen with the arrows will be shown for a very short amount of time before it disappears. Afterwards you will be asked to make a response based on the display.

Before the Averaging Task:

The response screen will have a single arrow in the middle of the screen, as well as a line of instructions at the top to remind you of the current task. At this screen, you will try to move the center arrow to match what you think the average orientation was of all 16 arrows shown on
the previous display. To respond, you will click on the arrow in the center of the screen and drag it to the right or left until it matches what you think the average orientation was. Once the arrow is pointing in the correct orientation, hit the space bar to proceed to the next trial. Do you have any questions so far?

If you have no further questions, we will begin a set of practice trials to let you see how the trials will play out. We will begin with a slowed down version of the trial progression, so that you can see each piece, and then pick up speed to match how fast the experimental trials will go. If you are ready, we will now begin.

*During the slow trial presentation, the instructions on what to do for each frame will be repeated: “fixate your eyes on the center of the cross and hold them there”, “you will now see a set of oriented arrows, keep your eyes in the center of the screen until the next display”, “you will now respond with the average orientation of all the arrows on the previous screen by clicking on the arrow and moving it with the mouse. Lock in your answer with the space bar.”

**Before the Whole-Report Task:**

The response screen will contain 16 arrows in a grid on the screen, as well as a line of instructions at the top to remind you of the current task. Your goal will be to recreate as many of the previously shown arrows as you can. At this screen, you will move each arrow to match what you think the orientation was of the arrow shown in that position on the previous frame. To respond, you will click on the arrow and drag it to the right or left until it matches what you think the orientation was. If you do not remember the arrow orientation in a location, you may leave that location’s arrow unadjusted. Once every arrow that you remember is pointing in the correct orientation, hit the space bar to proceed to the next trial. Do you have any questions so far?
If you have no further questions, we will begin a set of practice trials to let you see how
the trials will play out. We will begin with a slowed down version of the trial progression, so that
you can see each piece, and then pick up speed to match how fast the experimental trials will go.
If you are ready, we will now begin.

*During the slow trial presentation, the instructions on what to do for each frame will be
repeated: “fixate your eyes on the center of the cross and hold them there”, “you will now see a
set of oriented arrows, keep your eyes in the center of the screen until the next display”, “you
will now respond with as many of the arrows on the previous screen as you can remember by
clicking on the arrow in each location and moving it with the mouse. Once you have reported as
many arrows as you can, lock in your answer with the space bar.”

At the end of the practice trials:

We will now begin the real experimental trials. The trials will progress just like they did
in the practice trials, except there will be quite a bit more. The experiment will be broken up into
4 blocks, 2 for each task that you will complete. After you have completed each block of trials, a
screen will appear telling you to take a break. When this appears, let me know and you will
receive a 3-minute break to rest your eyes before starting the next block. After you have
completed all 4 blocks, you will receive a message on the screen telling you that the experiment
has ended. Let me know when this screen appears, and we will document your points in SONA.
Experiment 2 Instructions

Practice trials:

In this experiment, you will be tasked with making responses on quickly displayed screens of oriented arrows. We will begin with a quick overview of what the display will look like in your first task and then move on to a few practice trials.

Each trial will begin with a small cross presented in the middle of the screen. When you see this cross, fixate your eyes on its center and hold them there while the screen with arrows is presented, until the response screen appears. The next screen you will be shown will contain 16 arrows, each pointing in a particular orientation. Each individual arrow will always either point to the left, right, or upwards to some degree. The screen with the arrows will be shown for a very short amount of time before it disappears. On half of the blocks, this screen will be followed by a pattern of lines. Afterwards you will be asked to make a response based on the display of 16 arrows. Your response will always be based on the screen with the 4x4 grid of arrows on it, not any of the other screens.

The response screen will have a single arrow in the middle of the screen, as well as a line of instructions at the top to remind you of the current task. At this screen, you will try to move the center arrow to match what you think the average orientation was of all 16 arrows shown on the previous display. To respond, you will move the mouse up and down to rotate the arrow on the screen until it matches what you think the average orientation was. Once the arrow is pointing in the correct orientation, hit the space bar to proceed to the next trial. Do you have any questions so far?
If you have no further questions, we will begin a set of practice trials to let you see how the trials will progress. We will begin with a slowed down version of the trial progression, so that you can see each frame, and then pick up speed to match how fast the experimental trials will go. If you are ready, we will now begin.

*During the slow trial presentation, the instructions on what to do for each frame will be repeated: “fixate your eyes on the center of the cross and hold them there”, “you will now see a set of oriented arrows, keep your eyes in the center of the screen until the next display”, “you will now respond with the average orientation of all the arrows on the previous screen by moving the mouse up and down to rotate the arrow on the screen until it matches what you think the average orientation was. Lock in your answer by pressing the space bar.”

At the end of the practice trials:

We will now begin the real experimental trials. The trials will progress just like they did in the practice trials, except there will be quite a bit more. The experiment will be broken up into 4 blocks, 2 will have a screen with a pattern of lines on it before you respond and 2 will be blank before you respond. After you have completed each block of trials, a screen will appear telling you to take a break. When this appears, let me know and you will receive a 3-minute break to rest your eyes before starting the next block. After you have completed all 6 blocks, you will receive a message on the screen telling you that the experiment has ended. Let me know when this screen appears, and we will document your points in SONA.
APPENDIX B: IRB APPROVAL LETTER

EXEMPT DETERMINATION

May 27, 2021

Jacob Zepp
Tampa, FL 33647

Dear Mr. Zepp:

On 5/27/2021, the IRB reviewed and approved the following protocol:

<table>
<thead>
<tr>
<th>Application Type</th>
<th>Initial Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRB ID</td>
<td>STUDY002740</td>
</tr>
<tr>
<td>Review Type</td>
<td>Exempt 2 and 3</td>
</tr>
<tr>
<td>Title</td>
<td>Temporal and Spatial Properties of Orientation Summary Statistic Representations</td>
</tr>
<tr>
<td>Funding</td>
<td>None</td>
</tr>
<tr>
<td>Protocol</td>
<td>2740 Protocol, Version #1, 05272021 .docx</td>
</tr>
</tbody>
</table>

The IRB determined that this protocol meets the criteria for exemption from IRB review.

In conducting this protocol, you are required to follow the requirements listed in the INVESTIGATOR MANUAL (HRP-103).

Please note, as per USF policy, once the exempt determination is made, the application is closed in BullsIRB. This does not limit your ability to conduct the research. Any proposed or anticipated change to the study design that was previously declared exempt from IRB oversight must be submitted to the IRB as a new study prior to initiation of the change. However, administrative changes, including changes in research personnel, do not warrant a modification or new application.

Ongoing IRB review and approval by this organization is not required. This determination applies only to the activities described in the IRB submission and does not apply should any changes be made. If changes are made and there are questions about whether these activities impact the exempt determination, please submit a new request to the IRB for a determination.

Institutional Review Boards / Research Integrity & Compliance
FWA No. 0001669
University of South Florida / 3702 Spectrum Blvd., Suite 165 / Tampa, FL 33612 / 813-974-5638

Page 1 of 2
Sincerely,

Various Menzel
IRB Research Compliance Administrator