Effective perceptual and cognitive functioning in a noisy world: Computing summaries and identifying randomness

Bjorn Hubert-Wallander

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2015

Reading Committee:

Geoffrey M. Boynton, Chair

Scott O. Murray

Geoffrey R. Loftus

Program Authorized to Offer Degree:

Psychology

University of Washington

Abstract

Effective perceptual and cognitive functioning in a noisy world: Computing summaries and identifying randomness

Bjorn Hubert-Wallander

Chair of the Supervisory Committee:
Professor Geoffrey M. Boynton
Department of Psychology

The world around us contains a vast amount of information, but this information is contaminated by a similarly vast amount of noise. It is the job of human perception and cognition to promote effective functioning by separating the signal from the noise. While many perceptual and cognitive processes are involved in this mission, this dissertation reports on behavioral investigations into two specific aspects of how we deal with our noisy world: summary computation and identifying randomness. Chapters 2 and 3 present and discuss a line of experiments designed to understand how we determine the average value of a series of visual objects, focusing on how this summary computation differs depending on what visual feature is being averaged. Chapter 4 reports on a series of experiments investigating how we identify randomness in a set of stimuli or outcomes, and whether these perceptions can be predicted based on the stimuli themselves.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

Any acknowledgements here must start with the principal investigators in the lab: Professors Geoff Boynton, Scott Murray, and Ione Fine. Each of them has contributed enormously to my personal and professional development, and to the work presented here. For their teaching, advice, and fun conversation across five years, I am thoroughly grateful.

I am also grateful to the lab members at large. They have each influenced this work in some way, and have also made my time in the lab more than a little fun. So thanks to Jess, Kit, Sung Jun, Erik, Libby, Zach, Jeff, Jason, Paola, Fang, Maria, Andrew, Alex, Michael-Paul, and Ani for making the lab the happy, collaborative environment that it is.

I am eternally thankful for the members of the Department's support and administrative staff. They have tolerated my poorly filled out forms, already-answered questions, and general bothering for over five years. Jeanny Mai in particular has been an invaluable resource as the Program's administrator and general knower-of-all-things. The extent to which she makes the lives of the graduate students easier cannot be understated. I am also thankful to all those who have worked to make the Department and Graduate Program better than it was yesterday.

Additionally, none of this work would exist without the generous contribution of time and effort by those participating in my experiments. Their willingness to endure truly boring experiences is appreciated. I am also thankful to Fabian, Zeke, Blake, and Wendy, who helped collect data from these valiant participants over the years.

Last but certainly not least comes the vital funding provided to both me and my lab by the National Science Foundation and the National Institutes of Health. Without this necessary resource provided by the public, none of this research would have been possible.

DEDICATION

This work is dedicated equally to two groups of people, and only slightly less equally to a third.

My family has always been a place I know I can turn if I need advice, if things are difficult, if I need to complain, or if I want to think about anything other than school for a while. Though I'm sure they'd all be happy if I called more often, their support has been unwavering for as long as I can remember, and is something I have often taken for granted. I hope they're proud of this work, and of me.

I also dedicate this work to Jess and Ashwin. I can scarcely put into words the meaning that these two have to me, and the value that they've brought to my life since coming to Seattle. The extent to which they both amaze and impress me, and the extent to which I'm humbled by their affection for me, truly knows no bounds. I love you both.

Finally, I dedicate this work to all other friends and partners who have helped me enjoy and grow in my time in Seattle. These years didn't have to be as fun and fulfilling as this, but they were.

Chapter 1. INTRODUCTION

From the moment that we wake every morning, our senses are confronted with a constant and vast stream of information. Nearly all of this information is contaminated by some amount of noise, or inherent variability that complicates the problem of handling the information stream effectively. This problem of noise appears in nearly all perceptual domains. For example, light reflecting off the surface of our brown desk actually contains hundreds of slightly different color shades. Similarly, the leaves hanging outside our window all vary slightly in their shape and size and the sounds of traffic from the nearby freeway are constantly varying in small ways. Noise contaminates signal not only in sensory domains, but also in more conceptual ones. The financial markets, for example, fluctuate daily in unpredictable ways even as they follow a larger trend across months or years. Similarly, day to day weather can vary seemingly randomly even within a season, and the moods and feelings of those around us often seem to vary without apparent cause. And finally, coin flip, lottery, and gambling outcomes are almost entirely unpredictable.

The primary job of perception and also a large part of cognition is to separate the signal from the noise. As an example, consider an early human who must hunt or gather food. If her habitat is periodically visited by a group of dangerous animals, two general outcomes are possible. In the first, she is able to discern the pattern in the noisy data and know that animals usually (but not always) visit the area once a week in the warmer seasons and usually (but not always) visit once a month in the cooler seasons. In this case, the human can adjust her behavior to account for this observation For example, she can avoid the area only during times when they are likely to appear, or even lay a trap for them before they are predicted to arrive. In this case the human has become more suited to her environment and has increased her chances of survival. However, in the case that the human is not able to discern the pattern and therefore the rule that governs the animals' visitations, the only suitable response is to avoid the area entirely since the visitations cannot be understood or exploited in any productive way. This likely would incur some significant cost to the human, who must now expend energy and safety to venture to some unknown other habitat. She also forfeits the opportunity cost of whatever benefits she could have gained by being able to predict the animals' arrival. Thus, discerning the patterns, regularities, and structures around us is immensely important for our survival, since only by using them can

we infer and understand the rules and systems that govern the world. And if we cannot understand the laws of our environment, we are poorly poised to operate effectively in it.

Our struggle to productively function despite variability in our environment encompasses a vast number of perceptual and cognitive processes, and the study of them is as old as psychology itself. In this dissertation, I will present and discuss scientific investigations into two specific aspects of how we deal with our noisy world.

1.1 SUMMARY COMPUTATION

Thankfully, we are often up to the task of dealing with variability in our environment. For example, if I look closely at the section of beige wall that sits behind my computer monitor as I write this, I can see many tiny variations in its color, created by bumps in the surface that all reflect the ambient light slightly differently. However, unless I'm carefully attending to the variability in the color coming off the wall, I do not perceptually experience a hundred slightly different hues. Instead I simply see beige. This is because my visual system has, at some point, deemed this minor color variability irrelevant and has discarded it in favor of a less complicated but still functional percept. In other words, it has extracted the signal (beige) and discarded the noise (beige #301, beige #756, beige #42, etc.). In perception research, this process is studied under the label of *summary statistics*.

Many reports over the past fifteen years have shown that our visual system is adept at extracting summary statistics (usually the mean) from visual input with great speed and facility. These reports have also shown that this process can operate in a large number of visual feature domains. However, very little is known about how summary statistics are actually built from the input, and whether this process differs depending on what is being summarized. For example, are summaries of mean object location meaningfully different from summaries of mean object size? Additionally, it is not well-known how summarization or averaging processes operate when the objects to be summarized vary over time, as in a face whose expression changes as you look at it. Chapter 2 of this dissertation describes a series of behavioral studies designed to address these heretofore unanswered questions. Chapter 3 follows up on several unexplained findings documented in Chapter 2 and attempts to discover their causes and implications.

1.2 IDENTIFYING RANDOMNESS

Unfortunately, our ability to discern signal from noise can be overmatched if the noise is too severe. For example, many people and companies go to great lengths to predict the stock market, but the amount of day-to-day variability in any given stock has consistently proved too great for people or even sophisticated software to extract useful patterns. The same thing can happen in human cognition. If no useful information can be obtained from some stimulus or series of events we should spend our time and energy elsewhere. This is not only because more might be gained from interacting with or studying other phenomena, but because environments that are too noisy to predict can be actively hazardous, as in the above example of the early human attempting to discern the movements of dangerous predators. How do we decide when the noise is too much, and that we should simply label some underlying process as random?

Most research into randomness perception has shown that humans are surprisingly poor at understanding how randomness behaves, and what outcomes from random processes look like. The primary finding has been that we are surprisingly reluctant to label sets of outcomes from random generating processes as actually coming from truly random sources. For example, a streak of four heads in a row from a flipped coin would cause many to question the fairness of the coin, perhaps speculating on weight differences on its two sides, or on the moral integrity of the coin flipper. This is despite the fact that a truly fair, truly random coin does produce streaks of four heads or tails with regularity. While much research has focused on whether our judgments around randomness are good or bad, little is known about what actually causes people to label specific sequences of outcomes as random or non-random. Chapter 4 of this dissertation describes a series of behavioral experiments that attempt to understand how we decide to identify some set of outcomes as random. It also describes the development and evaluation of a mathematical model that attempts to predict how random any given binary sequence will appear to people.

Chapter 2. NOT ALL SUMMARY STATISTICS ARE MADE EQUAL: EVIDENCE FROM EXTRACTING SUMMARIES ACROSS TIME

2.1 ABSTRACT

Over the past fifteen years, a number of behavioral studies have shown that the human visual system can extract the average value of a set of items along a variety of feature dimensions, often with great facility and accuracy. These efficient representations of sets of items are commonly referred to as summary representations, but very little is known about whether their computation constitutes a single unitary process or if it involves different mechanisms in different domains. Here, we asked participants to report the average value of a set of items presented serially over time in four different feature dimensions. We then measured the contribution of different parts of the information stream to the reported summaries. We found that this temporal weighting profile differs greatly across domains. Specifically, summaries of mean object location (Experiment 1) were influenced approximately 2.5x more by early items than by later items. Summaries of mean object size (Experiment 1), mean facial expression (Experiment 2), and mean motion direction (Experiment 3), however, were more influenced by later items. These primacy and recency effects show that summary representations computed across time do not incorporate all items equally. Furthermore, our results support the hypothesis that summary representations operate differently in different feature domains, and may be subserved by distinct mechanisms.

2.2 INTRODUCTION

The human visual system is constantly confronted with a large, complex, and dynamic stream of information that far outstrips its processing capacity. One way the visual system is thought to deal with this problem is through the use of *summary representations* (also referred to as *ensemble representations, summary statistics,* or *set representations*), wherein a central tendency is extracted from a set of stimuli that vary along one or more feature dimensions. Recent behavioral studies show that participants can accurately report summary representations for mean size (Ariely, 2001; Chong & Treisman, 2003, 2005a, 2005b), mean orientation (Dakin,

2001; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001; Robitaille & Harris, 2011), mean position (Alvarez & Oliva, 2008; Greenwood, Bex, & Dakin, 2009; Spencer, 1961, 1963), mean color of a group of objects (de Gardelle & Summerfield, 2011), and even the mean expression or identity contained in a set of faces (de Fockert & Wolfenstein, 2009; Haberman, Harp, & Whitney, 2009; Haberman & Whitney, 2007, 2009). Given the accuracy, efficiency, and automaticity with which they appear to be computed, some have speculated that summary representations are an important contributor to our subjective impression of a complete visual world, since they could potentially provide a rough sketch of areas or objects we're not currently focusing on (Whitney, Haberman, & Sweeny, 2013).

2.2.1 *Extracting summaries across time*

Most studies of summary representations have used static arrays, where all the samples that participants are expected to summarize are presented concurrently. However, real world visual input is inherently dynamic, and the properties of objects that we wish to know about often change across time. For example, the expression on a friend's face evolves as he or she listens to what we say. Finally, even for a static visual stimulus, information is effectively sampled serially in time through shifts in visual attention and eye position.

While several studies have shown that it is *possible* to extract a summary representation from stimuli presented across time, (Albrecht & Scholl, 2010; Albrecht, Scholl, & Chun, 2012; Corbett & Oriet, 2011; Haberman, et al., 2009; Piazza, Sweeny, Wessel, Silver, & Whitney, 2013), very little is known about *how* the summary is built in this case. How are the individual stimuli incorporated into a summary? Work by Juni and colleagues (Juni, Gureckis, & Maloney, 2012) has demonstrated that summary-like judgments are sensitive to manipulations of how informative early and late items are, with more reliable items contributing more to participants' judgments than less reliable ones. But it still remains to be seen how a summary is constructed from a set presented over time in the default case where all items carry equal information. Do all items or parts of the information stream contribute equally to the perceived mean, or do early items or late items contribute more heavily?

Unequal weighting of information over time, here referred to as primacy and recency, might be consequences of underlying neuronal processes or mechanisms, or might reflect optimal

behavior for a particular task. For example, if making a decision rapidly is important and the first few items provide sufficient information for the task, one might expect primacy. The costs of time and cognitive resources to incorporate later items might outweigh any additional accuracy they might contribute. On the other hand, one might expect recency if an accurate representation of the most recent state of the world is desired, or if early information is lost due to memory or attentional limitations.

2.2.2 *Extracting summaries across feature domains*

Summary representation has been invoked to describe behavior across a wide variety of visual features, but surprisingly little is known about how summarization mechanisms might or might not differ across those domains. Is summary extraction a general cognitive mechanism that operates similarly across stimulus domains, or does it depend on the feature domain of interest?

Some studies have attempted to address this question by comparing various properties of summaries across different feature domains, but reports on these properties are generally few, unclear, or conflicted. For example, some researchers have compared the accuracy of summary representations in two or more domains (Albrecht, et al., 2012; Emmanouil & Treisman, 2008), but using accuracy measures may not be optimal since, as Albrecht et al. note, it is likely highly dependent on the statistical properties of the particular stimuli used. One can also compare domains by examining how summary accuracy varies with the number of items present, but even within a domain accuracy sometimes increases with set size (Ariely, 2001; Chong & Treisman, 2003, 2005b; Parkes, et al., 2001) and sometimes does not (Robitaille & Harris, 2011; Solomon, Morgan, & Chubb, 2011). Finally, one could compare domains by examining what portion of items in a set are incorporated into a summary. However, not enough is known about this property to understand whether it differs across domains. Most researchers are only able to conclude that more than two but fewer than all items present are used to summarize (Dakin, 2001; Morgan & Glennerster, 1991; Piazza, et al., 2013; Solomon, 2010; Solomon, et al., 2011; Watamaniuk & Duchon, 1992). Comparing how summaries are computed across time in different feature domains may help to address this unresolved issue.

2.2.3 *The present study*

Here, we explored both how summary statistics are constructed when stimuli are presented serially across time and whether such summary computation differs across feature domains. We found that not all items contribute equally to summary representations built across time, with different temporal weighting profiles appearing in different feature domains. In particular, judgments of mean object position appeared special, apparently operating differently from those of mean object size, mean facial expression, and mean motion direction. These differences in how information is used over time to compute a summary representation raise the possibility that different mechanisms are associated with different feature domains.

2.3 EXPERIMENT 1: AVERAGING POSITION AND SIZE ACROSS TIME

To understand how summary representations are constructed when the items are presented serially across time, we presented participants with a sequence of small white dots and asked them to report either the mean size or the mean position of the group. We then quantified the influence of the temporal position of each dot on the participants' estimates across many trials.

2.3.1 *Method*

2.3.1.1 Participants

Twenty five students at the University of Washington were recruited from the Department of Psychology's undergraduate participant pool, where students may volunteer as participants for studies in exchange for course credit. This number of participants to recruit was decided upon after initial simulations conducted based on pilot testing showed that approximately 20 participants would result in relatively stable estimations of group-level weights (see Results section) in both tasks. All participants had normal or corrected-to-normal vision. Recruitment and study procedures in all experiments presented here were conducted in accordance with the ethical policies set forth by the University of Washington's Human Subjects Division, and those in the Declaration of Helsinki.

2.3.1.2 Apparatus

All study procedures took place in a dimly lit room, with the participants seated 50 cm from a CRT monitor subtending 40.4° x 30.8° of visual angle. A chinrest was used to ensure constant viewing distance to the monitor, on which all stimuli were shown against a black background. A faint grey grid pattern was always present as part of the background in order to provide spatial reference. All stimuli were generated by custom software written using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) for MATLAB (The Mathworks, Natick, MA).

2.3.1.3 Stimuli and procedure

As shown in Figure 2-1, participants were precued at the beginning of each trial with the word "Location" or "Size" presented slightly above a white central cross approximately 0.8° in width and height. Despite the presence of the central cross, no instructions about fixation were given and participants were free to look wherever they wished over the course of the experiment. This word precue was present for a random period of time between one and two seconds and indicated what the participant would be asked to report at the end of the trial.

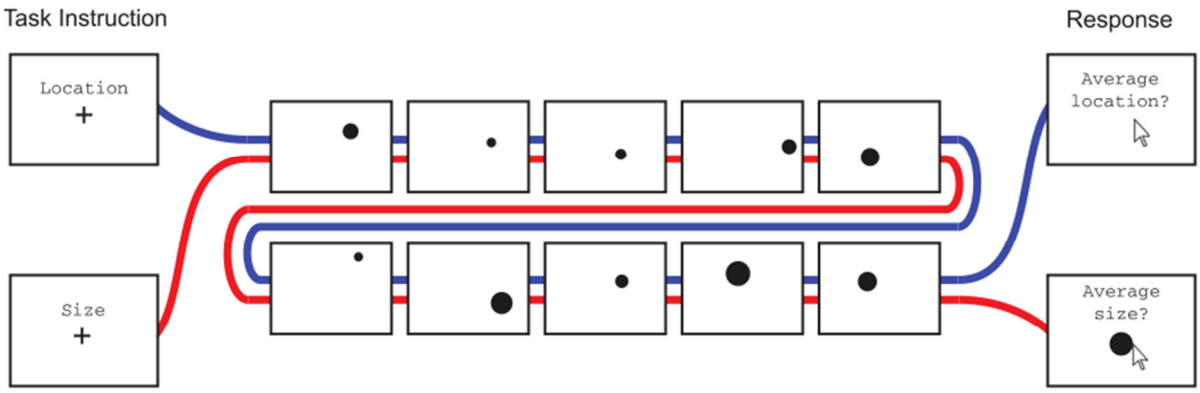

Figure 2-1. Trial schematic from Experiment 1.

Each trial consisted of a series of ten white dots that varied in their position and size. Each dot was shown for 150 ms and was followed by a 50 ms blank inter-dot interval, resulting in a dot presentation rate of 5 Hz (Figure 2-1). On a given trial, dot locations were chosen by sampling ten times from a bivariate Gaussian probability distribution with a standard deviation of 2.3°. The center of the bivariate Gaussian distribution was drawn randomly on each trial from a

square-shaped uniform probability distribution centered on the middle of the screen and subtending 21.1° x 21.1°, corresponding to 70% of the vertical height of the screen. If a dot's sampled location was outside the borders of the screen, it was moved to the point on the screen nearest to its originally sampled location. Dot radii were similarly sampled from a Gaussian distribution, the center of which was sampled on each trial from a uniform distribution ranging from 0.3° to 1.5°. The standard deviation of the Gaussian distribution was always exactly 0.3 times the center of the same distribution in order to minimize the possibility of sampling dot radii below zero. If a dot's radius was sampled to be below zero, it was resampled until it was above zero. The series of ten dots was followed by a blank period of 300 ms, followed by a response period.

During the response period, participants were reminded what to report by text appearing near the middle of the screen. On location trials participants reported the "average location" or "center" of the dots seen on that trial by moving the mouse cursor and clicking on their perceived center. The mouse cursor was visible to the participants only during the response period of location trials. On size trials participants reported the "average size" of the dots seen on that trial by adjusting the size of a centrally-presented test dot, the radius of which varied with the horizontal location of the mouse. Participants clicked to submit their response when the perceived mean size was obtained. The test dot's initial radius was chosen randomly on each trial from 0.1° to 3.2°, which also served as the limits of possible responses. The response period ended when the participant submitted his or her response, and was followed by a 1500 ms inter-trial interval.

Each participant received full instructions from an experimenter and then completed approximately ten practice trials in view of the experimenter before beginning the experimental trials. Each participant completed 320 experimental trials in blocks of 40 trials. Blocks alternated between all location trials and all size trials, with the first block type seen counterbalanced across participants. The participants were free to take breaks after each block, but could also do so at any point during the experiment by simply waiting to submit their response for a given trial. A full experimental session lasted about an hour.

2.3.2 *Results*

Weights that quantified the relative influence of each dot number (one through ten) on participants' responses were obtained by fitting a weighted average model (as used in Juni, et al., 2012) to each participant's data separately for size and location trials. The model took the form

$$R_j = \sum_{i=1}^{10} w_i x_{ij}$$

where R_j is the participant's response for trial j, x_{ij} is the position or radius of the dot at temporal position i and trial j, and w_i is the weight for temporal position i. Linear regression was used to obtain the least-squares best fitting set of weights for each participant for each task. The model fit the data well in all participants in both location (model R^2: mean = 0.98, SD = 0.02; all $p <$ 0.001) and size (model R^2: mean = 0.74, SD = 0.10; all $p < 0.001$) trials. Mean weights across participants for both trial types in Experiment 1 are shown in Figure 2-2, with error bars depicting 95% confidence intervals.

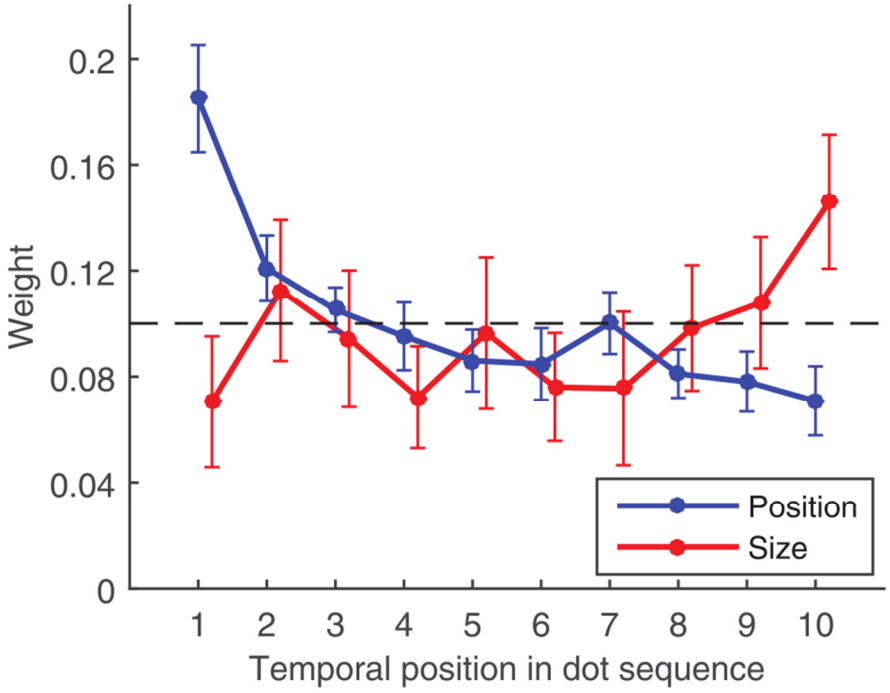

Figure 2-2. Results from Experiment 1.

The dashed line in Figure 2-2 shows the weight that would be obtained for each dot if all dots contributed equally to participants' responses and no other source of noise or bias were present. It is clear that participants did not weigh each of the ten dots evenly in either feature domain. Instead, we found *primacy* for participants' mean position judgments, with dots that appeared early in the sequence contributing more to the perceived mean position than later dots. The pattern was reversed for judgments of mean dot size, where participants showed clear *recency*; later dots contributed more to the perceived mean size. The size of the effect was considerable for both feature domains, where the most-weighted dots contributed approximately two to three times as much as the least-weighted dots. Additionally, the effect was relatively smooth in both cases, with weights gradually increasing or decreasing as the dot number increased, though pooling across many trials may have averaged out more discrete effects. A two-way ANOVA on the weight data showed a significant dot number x trial type interaction in a two-way ANOVA, $F(9,216) = 10.40$, $p < 0.001$, $\eta_p^2 = .30$, and one-way ANOVAs performed separately on location, $F(9,216) = 23.15$, $p < 0.001$, $\eta_p^2 = .49$, and size trials, $F(9,216) = 3.10$, $p = 0.002$, $\eta_p^2 = .11$, showed significance in both cases.

A set of five additional experiments were conducted separately in the two feature domains to test for the replicability of these findings. Under a variety of experimental manipulations, we consistently found primacy for mean position judgments and recency for mean size judgments (see Appendix A for methods and results). Together, these results indicate that summary representations of mean size and mean position extracted from items presented across time do not incorporate all presented items equally. Instead, we find that perception of mean size and position across time favor different portions of the information stream, with mean size favoring more recently presented items and mean position favoring earlier presented items.

2.4 Experiment 2: Averaging facial expression across time

Results from Experiment 1 suggest that how summary representations are computed across time may differ across feature domains. Experiments 2 and 3 explored this possibility by extending the method to two new feature domains. In Experiment 2, we used emotive face stimuli created by Haberman & Whitney (2007, 2009) to explore how summary representations of mean facial expression are generated across time.

2.4.1 *Method*

Participants, apparatus, stimuli, and procedure in general mirrored those of Experiment 1, except where otherwise noted.

2.4.1.1 Participants and apparatus

A new group of twenty participants was recruited from the University of Washington undergraduate student body using different participants from the same Department of Psychology participant pool as before. The decision to use this smaller number of participants than used in Experiment 1 was made ahead of data collection since in Experiment 2 all trials (rather than half) would contribute to a single set of weights per participant. The apparatus as in Experiment 1 was used, except that stimuli were presented on a medium gray background instead of black, and no grid was present.

2.4.1.2 Stimuli and procedure

Each trial began with the presentation of a black central cross approximately 0.4° in width and height, though participants were free to look wherever they wished over the course of the trial. After 500 ms, the cross changed color to red as a trial start warning and was present for a random period of time between one and two seconds. A series of eight faces was then presented.

Face stimuli consisted of eight human faces, presented one after another in the center of the screen. Each face was present for 252 ms and was followed by an 82 ms blank inter-face interval, resulting in a presentation rate of approximately 3 Hz. The set of faces used was a subset of stimuli used by Haberman & Whitney (2007, 2009) in a series of experiments showing that humans can accurately and efficiently extract the mean emotional expression contained in a set of human faces, and consisted of fifty faces from the same person. The extreme two faces were actual photos, one showing a happy expression and one showing a sad one. The remaining 48 faces were regularly-spaced interpolations between the two emotionally extreme ones, created with image morphing software. See Haberman & Whitney (2009) for further details on the generation and properties of the faces, and see Figure 2-3 for example faces. Just as was done by Haberman & Whitney, we assigned each face in the stimulus set a number from 1 to 50 for the purposes of stimulus sampling and modelling, such that the distance between sequential faces is arbitrarily defined as one “eu”, or *emotional unit*. When presented, the face stimuli subtended

6.0° vertically and 4.5° horizontally. The exact faces shown on each trial were chosen by sampling eight times from a Gaussian distribution (rounding to the nearest eu) with a standard deviation of 6 eu and a center that was itself drawn on each trial from a uniform distribution over face space between faces 11 and 39, inclusive. Any face that was sampled outside of the range 1 to 50 was resampled until it fell within the showable range.

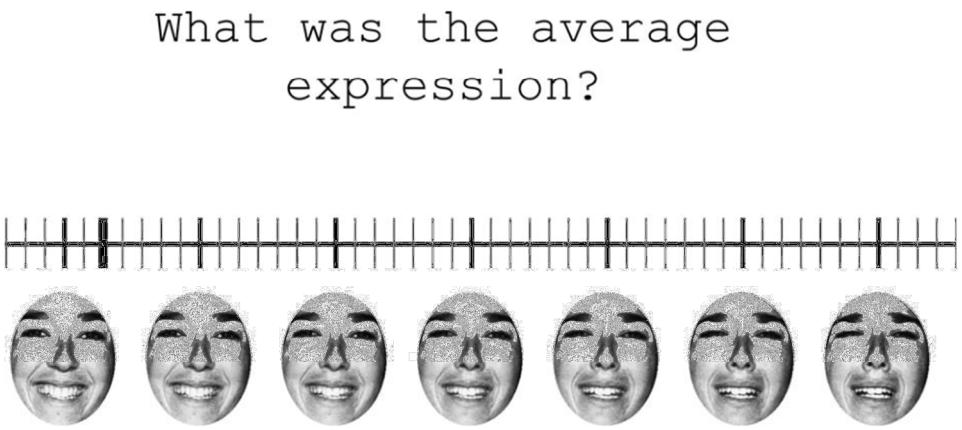

Figure 2-3. Response screen used in Experiment 2.

The series of faces was followed by a blank period of 500 ms, followed by a response screen (see Figure 2-3) containing a horizontal line with fifty evenly spaced vertical tick marks in it, corresponding to the fifty possible response faces. Seven of the tick marks (numbers 4, 11, 18, 25, 32, 39, and 46) were larger than the others and had images of their corresponding face displayed below them. The purpose of these ticks and faces was to act as landmarks that participants could use to base their responses on. Once the response screen appeared, participants were asked to report the "average expression" in the faces shown on that trial by clicking the mouse cursor on the tick mark that most closely matched their estimate. Participants were explicitly instructed to not limit themselves to the landmark ticks or faces and to select one of the ticks in between if they thought the correct answer was between two landmark faces. The initial cursor position was chosen randomly at the start of each response period. The response period

ended when the participant clicked on a tick mark, and was immediately followed by the start of the next trial.

Each participant completed 320 trials in blocks of 40 trials. The break, instruction, and practice procedures used in Experiment 1 were also used here. A full experimental session lasted about fifty minutes.

2.4.2 *Results*

The same weighted average model used in Experiment 1 was fitted to the present data and a set of eight weights was obtained for each participant that quantified the average contribution of the eight temporal positions to the participants' responses. The model once again described a significant portion of the variance, though the amount explained was in general lower and more variable than seen in Experiment 1 (model R^2: mean = 0.63, SD = 0.17; all $p < 0.001$). There are a number of possible reasons for this. Both the stimulus and response space resolution were lower in Experiment 2, where only fifty unique values were possible compared to the virtually unconstrained response spaces of Experiment 1. Another potential source of noise is the fact that participants were presented only a subset of all possible face images while making their selection. Indeed, inspecting the frequency of participant responses across the fifty possible faces reveals that participants tended to choose ticks at the seven sample faces more often than ticks in between. While this tendency may add to the overall variability of the results, it should not affect the profile of weights over time. Finally, it is possible that the weighted average model simply describes the cognitive operation underlying facial expression averaging less well than for location or size.

Mean weights as a function of temporal position across all participants are shown in Figure 2-4, with error bars depicting 95% confidence intervals. Just as with estimates of mean size and mean location, not all faces contributed equally to the participants' responses. Instead, as in the size domain, estimates of mean facial expression exhibited clear recency, where the later faces influenced participants' reports of the mean more than the earlier ones. Again, the size of the effect was relatively large, with each of the last two faces contributing, on average, between 1.5 and 4 times more than each of the first two faces. The effect was once more relatively smooth, with the average weight increasing gradually from face number one to eight. A one-way

ANOVA confirmed the statistical significance of the effect, $F(7,133) = 9.12$, $p < 0.001$, $\eta_p^2 = .32$.

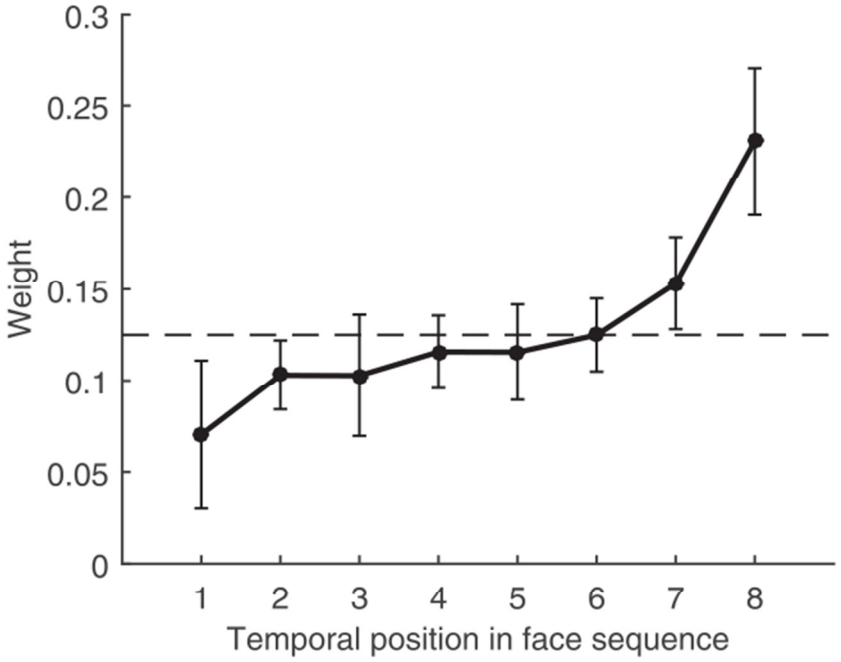

Figure 2-4. Results from Experiment 2.

2.5 EXPERIMENT 3: AVERAGING MOTION DIRECTION ACROSS TIME

Results from Experiment 2 suggest that computation of a summary representation for facial expression uses a temporal weighting profile similar to that used in computation of mean dot size, and distinct from that used in computation of mean dot location. Experiment 3 extended the method from Experiment 2 into a fourth feature domain: motion direction.

2.5.1 *Method*

A new group of twenty participants was recruited just as in Experiments 1 and 2, and the apparatus, stimuli, and procedure used closely mirrored those of Experiment 2, except where otherwise noted.

2.5.1.1 Stimuli and procedure

Each trial began with the presentation of a central red square marker approximately 0.5° in width and height, though once again participants were free to look wherever they wished over the course of the trial. After 200 ms, the square disappeared as a trial start warning. After a random period of between one and two seconds, a moving dot fields was then presented.

The moving dot field consisted of a square region (width = 8°) with a circular aperture (diameter = 8°) overlaid on it, both centered on the screen. Ten small (width = 0.12°) black dots moved within a square field, though only those inside the circular aperture were visible to the participant. The ten dots always moved with 100% coherency at 10°/s. Each dot had a maximum lifetime of 200 ms (12 frames at a 60 Hz monitor refresh rate) and was initialized with a random starting "age" between 0 and 11 monitor refresh frames. Initial dot location in the field was also random. When a dot reached the maximum age, it was destroyed and a new dot with age zero was created at a random point in the field. Whenever a dot's motion carried it outside the square field, its location was wrapped around to the opposite edge of the field.

Over the course of a single trial, the dot field moved in eight distinct directions, one after another, for 333 ms in each direction. Transitions between motion directions were abrupt, though dots persisted through the transition if their age permitted it. The set of eight motion directions were chosen on each trial by sampling from a Gaussian distribution with a standard deviation of 30° and a center that was itself sampled on each trial from a uniform distribution across all possible directions.

After the sequence of eight dot fields (lasting 2667 ms in total) there was a blank period of 300 ms, followed by a response screen. The response screen contained a central red square marker with a red line of length 2° radiating from it in a random initial direction. Once the response screen appeared, participants were asked to report the "average" or "overall" direction of motion present during that trial by using the mouse to adjust the radial direction of the red line until it pointed in the direction of the perceived average motion. The response period ended when the participant submitted their response with a mouse click and was followed by a 1000 ms inter-trial interval.

Each participant completed 320 trials in blocks of 40 trials. The break, instruction, and practice procedures used in Experiments 1 and 2 were also used here. A full experimental session lasted about forty minutes.

2.5.2 *Results*

Though the dot motion parameters were chosen to maximize, for any given motion direction, the likelihood of perceiving coherent motion in that direction, some participants found the task very difficult, apparently due to the motion reversal illusion. Of the 20 initial participants, data from 3 were discarded due to clearly outlying mean absolute response errors (MAEs). The discarded participants' MAEs were 24.5°, 33.5°, and 39.0°, compared to an average MAE of 14.4° (*SD* = 3.5°) in the rest of the participants. Within the remaining 17 participants, high error trials were removed from the data, where high error was defined as a response error greater than three standard deviations from the mean response error for that participant. This resulted in discarding, on average, 1.2% of the data (~3.7 trials) from each participant, with 2.2% of the data (7 trials) being removed in the most affected participant.

Best-fitting weights describing the average influence that each of the eight motion direction epochs had on participants' responses were obtained by fitting the same model described in Experiments 1 and 2 to the remaining data. Just as in Experiment 1, the data were described very well by the model (model R^2: mean = 0.98, *SD* = 0.01; all $p < 0.001$).

Mean weights from the 17 included participants are shown in Figure 2-5, with error bars depicting 95% confidence intervals. As was seen in Experiments 1 and 2, weights deviated from even weighting across temporal position. A one-way ANOVA confirmed this, $F(7,112) = 2.90$, $p = 0.008$, $\eta_p^2 = .15$. In particular, the last motion direction appeared to contribute somewhat more (~1.5 times) than the rest of the individual motion directions, in an apparent recency effect. Interestingly, and in contrast to our findings in other domains, the average weight given to the first motion direction seen on each trial also appeared to be relatively large, suggesting some degree of primacy. However, pairwise t-tests between weights for the first motion direction and motion directions two through seven showed that no pairs were statistically different, with all six *p*s > 0.13 before multiple comparisons correction. These results suggest that summary

representations for average motion direction over time (at least when the motion directions are discrete and separated in time) are computed more like those of size than location.

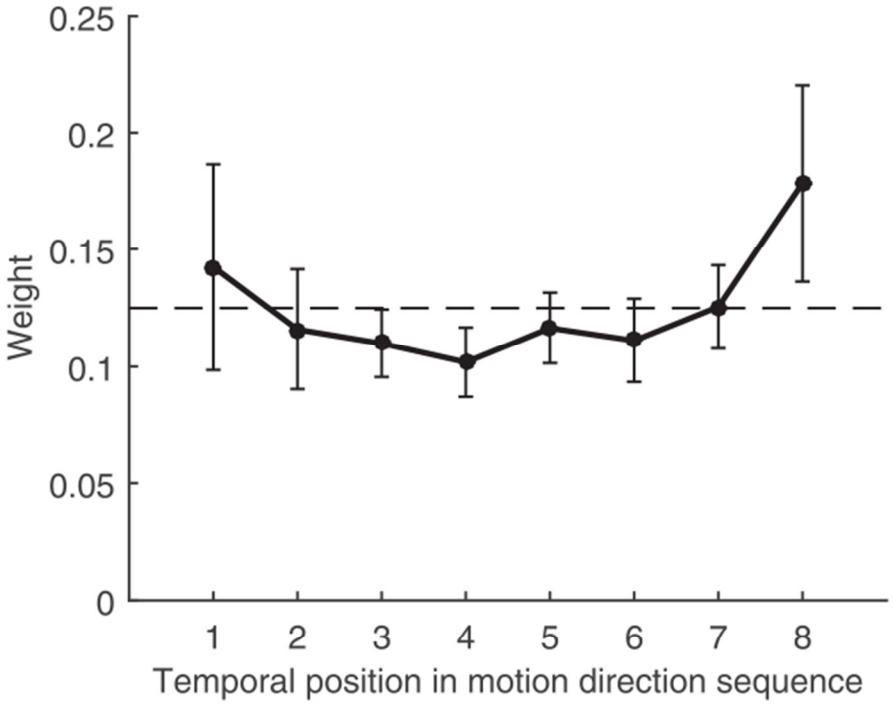

Figure 2-5. Results from Experiment 3.

2.6 General Discussion

Our primary finding is that the influence an item has on a summary representation depends on its temporal position in the information stream. Specifically, summary representations of mean position were more strongly influenced by earlier items (primacy) and summary representations of mean size, mean facial expression, and mean motion direction were more strongly influenced by later items (recency). Across the experiments, the effect was reasonably large, with the most influential items in the stream contributing on average about 1.5 to 3 times more than the least influential items. Below, we consider a variety of explanations for the particular pattern of results we observed.

2.6.1 *Strategies for summary computation*

One argument for why primacy was observed in the mean location task is that it can be a highly functional strategy. Kiani and colleagues (Kiani, Hanks, & Shadlen, 2008) found a form of primacy in a task where nonhuman primates had to integrate motion information across time. The dot motion stimulus they used moved with 0% coherence on average, but over the course of each individual trial there were fluctuations in the moment-to-moment motion direction. The investigators found that motion information in the early portion of the stimulus influenced the monkey's eventual decision more than that in the later portion. They and others went on to show that monkey and human behavior in this task is consistent with a model of information collection where the cost of sampling additional information increases as more and more evidence is accrued (Drugowitsch, Moreno-Bote, Churchland, Shadlen, & Pouget, 2012). In other words, if an observer is optimizing for energy or time expenditure in addition to accuracy, primacy might be an optimal strategy. Later information in a stream might not be worth integrating into the summary if the benefit to the estimate is outweighed by the cost of collecting it. This framework might explain why we found primacy for judgments of mean location, but it does not explain why we found recency in the other feature domains.

However, recency can also be understood as a functional behavior. Recency might reflect the way the brain makes predictions about what the next item in a series will be based on previous items. For example, if previously shown facial expressions predict the facial expression that is likely to come next, then it makes sense to be most sensitive to the predicted incoming expression and less sensitive to unpredicted expressions. It has been shown that this type of adaptive gain control leads to recency when observers are asked to integrate samples over time (Cheadle et al., 2014). In this account recency is produced because, by the time the last sample arrives, observers have a strong prediction about its value and are thus highly sensitive to even small deviations from that prediction. Again however, such an account cannot explain the totality of our results, since the model Cheadle et al. describe cannot naturally produce the primacy we observed for mean location.

It should also be noted that recency becomes an attractive strategy if the underlying process generating the samples is non-stationary. Many things in the real world that humans might want

to summarize, such as a conversation partner's face or a moving car, change in their properties over time. In this case, a summary representation that discards old information and reflects the most up-to-date state of the world is obviously valuable. However, two problems exist for this as a satisfying explanation of our recency findings. First, the underlying processes generating the stimuli in our experiments were stationary over the course of a trial in all cases. Second, it is unclear why the presumption of non-stationarity would apply to size, facial expression, and motion direction but not to location, where primacy was found.

Finally, related recent work on serial dependence has shown that the perceived orientation of a stimulus is systematically biased toward the orientation of recently seen stimuli (J. Fischer & Whitney, 2014). This result has been interpreted as evidence of a "continuity field" that promotes visual stability over time by effectively acting as a low-pass temporal filter that biases the current perception of the world towards recent events (Liberman, Fischer, & Whitney, 2014). On its face, the serial dependence effect is reminiscent of our recency effect, but it seems instead to predict primacy for judgments of the mean in our tasks, since the perception of the later samples would in theory be pulled towards those of the early samples. This would presumably result in earlier samples being represented more than later ones in the overall mean judgment. So it is possible that this newly-documented perceptual bias from recent events affects or supports the perception of the mean in our tasks, but more work would be needed to determine the exact nature of its role.

2.6.2 *Visual memory*

All of the present experimental tasks involve maintaining visual memory representations, either of specific items presented in the sequence or a running belief of the average, over the course of a trial. Given this, it is important to consider explanations for primacy and recency that involve characteristics of or limitations in visual working memory.

Though not unopposed, the traditional view of visual working memory is that it is composed of at least partially separated subsystems that are involved in storing different types of visual information. Specifically, considerable behavioral (Hyun & Luck, 2007; Logie & Marchetti, 1991; Woodman & Luck, 2004; Woodman, Vogel, & Luck, 2001), electrophysiological (Goldman-Rakic, 1996), and neuroimaging work (Courtney, Ungerleider, Keil, & Haxby, 1996,

1997; Smith & Jonides, 1997; Smith et al., 1995) supports the existence of subsystems for spatial- and object-based working memory representations. If the spatial working memory subsystem is primarily recruited in our mean location task and the object-based working memory subsystem is recruited in our mean size and mean facial expression tasks, then it is possible that primacy and recency are the result of differential characteristics of those systems. Motion direction as a visual feature, however, is inherently spatial and thus poses a problem for this account, since we found recency in that domain. However, since our motion stimulus was 100% coherent and changed abruptly from one direction to the next, the motion sequences in our task may have been encoded and summarized as a series of orientations, a feature more strongly associated with object-based processing systems than spatial-based systems. While there is some alignment between working memory subsystems and our findings across domains, this alone does not constitute a satisfying explanation for our effects. The relative lack of information about the differential properties and functioning of spatial- and object-based working memory subsystems makes it difficult to explain why one would lead specifically to primacy and the other to recency in a summarization task, for example. This problem is exacerbated by the fact that the most popular visual working memory tasks measure memory for objects that are defined by a binding of location and either color or orientation (Luck & Vogel, 1997; Zhang & Luck, 2008), confounding spatial- and object-based memory.

Could primacy and recency in our summarization task be driven by serial position effects in short-term memory? As in, do our results reflect more about memory quality for items presented in sequences than summarization processes themselves? There are indications of serial position effects in short term memory in at least two of the domains that we investigated here: location (Farrand & Jones, 1996; Farrand, Parmentier, & Jones, 2001; Guérard & Tremblay, 2008; Jones, Farrand, Stuart, & Morris, 1995) and faces (Hay, Smyth, Hitch, & Horton, 2007; Ward, Avons, & Melling, 2005). In these tasks, participants were shown a series of to-be-memorized objects (usually 5-12 items at 0.5 Hz) and are asked to reconstruct the sequence after some retention interval (usually 0-30 seconds). However, little evidence of domain differences in recall as a function of position in the sequence is noted. Instead, both primacy and recency of recall are seen in nearly all face and location experiments, with several of the researchers noting that the strength of primacy or recency seems to depend more on the specific testing or recall method used than on the feature domain (Farrand, et al., 2001; Jones, et al., 1995; Ward, et al., 2005).

While these findings are clearly related to the present results, this lack of domain differences makes it difficult to conclude that serial position effects in memory are solely responsible for our summarization findings, in particular the differences we observed across domains.

Finally, a related explanation for our findings of primacy and recency in summary computation is that one or both of them are due to capacity limitations in visual working memory. In this hypothesis, primacy is produced when the limited capacity of visual working memory is filled with memory representations from the early items and few to no resources are left to store the later items. Alternatively, recency might be produced by the same limited capacity if later items push representations of earlier items out of memory. The effect of memory capacity limitations on summary computation could in theory be tested by adding a sequence length manipulation to the experiments we report on here. If either or both of primacy and recency significantly diminish or disappear with shorter sequences, this would constitute evidence that capacity limits play a strong role in our findings. But if either primacy or recency are still observed with sequence lengths below the traditional visual working memory capacity (about 3-4 items according to Luck & Vogel, 1997), then the contribution of memory capacity to either effect is likely limited. It should be noted, though, that even if capacity limitations are involved in our findings, it is not immediately clear how this explanation alone would produce primacy in one feature domain and recency in the others.

2.6.3 *Implications for summary representation across domains*

Since it was reported in 2001 that human observers appeared to extract the mean size of an array of discs with surprising speed, precision, and automaticity, the concept of summary statistics has been invoked to describe reports of perceptual averaging across a wide variety of stimulus domains. Size, orientation, position, brightness, color, motion direction, speed, facial expression, facial identity, biological motion, and frequency of tones have all been discussed as features across which summary statistics might be computed. However, a priori, it seems unlikely that averaging in all of these domains shows the same characteristics that made summary perception of mean size so intriguing when it was first reported, especially considering that the physiology underpinning representation of these various features and objects in the brain is very different, and in some cases still not well understood. Despite this, the extent to which averaging across

these domains reflects the same computation has not been well-studied. In fact, to the best of our knowledge, the experiments reported here are the first to directly compare how summary representations are computed across a set of different feature domains.

Here we provide evidence that summary representation *behaves* differently in different feature domains, but do our results constitute evidence for distinct *mechanisms* for summary computation across time in different feature domains? As in, are summaries in different feature domains computed via a shared or similar process, or is each computed using anatomically or functionally distinct processes? Summaries that look different do not necessarily come from distinct mechanisms. A single mechanism that combines perceptual or memory representations into a summary could in theory produce different-looking results given different input. This is implied in our discussion of the role of visual memory above; perhaps the quality of early versus late item representations that are fed into a summarization mechanism simply varies across domains. This possibility combined with the lack of previous comparative work in summary representation domains makes it difficult to argue conclusively that distinct mechanisms are involved in summarizing in location and non-location domains, even if the summary computed is meaningfully different across domains.

In conclusion, it perhaps should not be surprising that the way in which summary representations are computed varies across feature domains. Just as other perceptual judgments fall along a continuum from low to higher-level processing, summary computation may do the same. Some feature domains are summarized at the sensory level. For example, a photoreceptor computes a weighted average of the spectrum of incoming light over a fixed period of time and space. Other feature domains are likely to require different, higher level cognitive processes, as in judging someone's overall moral character by his or her deeds. By understanding what similarities and differences exist between different types of summary representation, we will be better equipped to search for their underlying mechanisms, which is a major goal for this promising area of research. But until then, we conclude that not all summary statistics are created equal.

Chapter 3. CAUSES OF AVERAGING DIFFERENCES ACROSS VISUAL FEATURE DOMAINS

3.1 INTRODUCTION

In the previous chapter, we documented marked serial order effects for computing summary statistics on various visual features when the stimuli to be summarized arrive over time. In particular, we noted *primacy* in summaries of mean location for a series of briefly flashed dots. Earlier dots contributed two to three times more to participants' mean location estimates than later dots. On the other hand, we noted *recency* in summaries of mean dot size, the mean expression in a series of faces, and the mean motion direction of an animated dot field. In these domains, later information influenced mean estimates more than early information.

While these effects were notable for what they tell us about how summarization behaviors and mechanisms differ across visual feature domains, it remains unclear why exactly we observed them. After all, both primacy and recency will reliably result in incorrect estimates of means. In a perfect average, all items contribute equally regardless of the order in which they appeared. Furthermore, it was not apparent from those data why primacy in particular would operate in location and recency would operate in the other domains. In the discussion of Chapter 2, we considered several reasons why primacy or recency might be desirable. For example, primacy reflect a loss of interest in accruing additional information once we have an estimate that is "good enough," freeing us to attend to or process other information, and recency might be useful if the process generating the information to be summarized moves or changes in its properties over time. However, these accounts are more justifications than explanations, and they do not address how the effects might come to be. Here we consider and test three possible explanations for the temporal order effects that we observed in Chapter 2's mean location and mean size tasks.

3.1.1 *Memory*

Either primacy or recency could be the result of capacity limitations in visual working memory (VWM). One strategy that participants might have used in the mean location and mean size tasks described in Chapter 2 is to store each dot shown in the sequence individually in VWM. Then,

once the series is complete, the individual memory representations can be combined together into an estimate of the average. The downside of such a strategy is that each sequence in our task contained ten dots and the capacity of VWM is generally thought to be only about 3-4 objects or features (Luck & Vogel, 1997; Zhang & Luck, 2008). While the exact capacity seems to vary depending on what is being stored (Alvarez & Cavanagh, 2004; Brady & Alvarez, 2011), the number of dots that would need to be stored is beyond most estimates of VWM capacity.

In this situation, either primacy or recency could arise from the inability of VWM to precisely store every dot's location or size for a given trial. For example, it is possible that the early dots are stored precisely in memory as they appear, since there are few other memory demands early in a trial. However these early dots may consume most or all available storage resources, leaving few resources to precisely store the later dots. If, at the end of the trial, some dot memory representations are more precise than others but all are to be combined into an estimate of an underlying group average, then best results will be achieved by giving the more precisely represented dots more weight in the averaging process. This would result in the primacy of weights that we observed in the mean location task. However, a similar situation could also produce recency if the early dots are pushed out of memory by the later ones, "stealing" resources from them. If at the end of the trial the later dots are more precisely represented than the earlier ones, they will be more heavily weighted in the average.

This explanation for our effects is attractive in that it can naturally produce primacy or recency using the same framework, but a significant disadvantage is that it is not immediately clear how it would produce primacy in one feature domain (location) and recency in another (size). On one hand, it is possible that something about how object size is stored in memory makes those representations more susceptible to interference or disruption by subsequently presented items. Meanwhile, perhaps memory representations of object location are less fragile and can survive the arrival of new information. This could in theory explain how VWM capacity limitations produce both primacy and recency depending on the domain. Alternatively, it is possible that VWM capacity produces only one of primacy or recency, and that some other mechanism is responsible for the other.

In either case, the memory capacity hypothesis can be tested by manipulating the length of the dot sequences that need to be averaged such that some sequences are at or below the traditional VWM capacity limit of 3-4 items. If primacy or recency or both disappear or are significantly attenuated with short sequences, that would constitute evidence in favor of a memory-based explanation for that effect.

3.1.2 *Attention*

A second account for primacy involves the dynamics of visual attention in our task. In Experiment 1 of Chapter 2, ten bright dots were flashed in quick succession on an otherwise dark background, and the time and location of the first dot's arrival was relatively unpredictable. Because of this, it is highly likely that each trial involved a succession of visual attention shifts by the participant, both endogenous (based on predictions about where the next dot will appear) and exogenous (driven by the sudden appearances of salient objects). Thus, it is possible that patterns of attentional allocation could have produced either the primacy or recency that we observed.

The benefits of visual attention to perceptual processing are well known. Among them is increased spatial resolution at attended locations, leading a better ability to localize objects separately even when they are close together (Carrasco, Williams, & Yeshurun, 2002; He, Cavanagh, & Intriligator, 1996; Yeshurun & Carrasco, 1998). While participants in our mean location task are likely voluntarily attending to each dot as it appears, the first dot likely also benefits from attentional capture, since it constitutes a sudden bright object on an otherwise dark field. If the first dot receives more attention due to its salience, it presumably is processed and represented with a high degree of precision relative to the later dots. As is mentioned in the memory hypothesis above, items represented with high precision should contribute more to a group average than those represented with low precision. This could explain why the first dot is so heavily weighted in judgments of mean location. But as was seen in the weight profiles in Chapter 2 and Appendix A, the first several dots exhibited higher weights, not just the first. At least two explanations for how attention captured by the first dot could result in larger weights for subsequent dots are possible.

One possibility is if the perceived locations of dots after the first were "pulled" towards the locus of attention as summoned by the first dot. Spatial attention has been shown to warp visual space by altering the tuning properties (J. Fischer & Whitney, 2009) and receptive field locations of early visual neurons for the attended portions of the visual field, apparently shifting them towards the attended location (Connor, 1994; Womelsdorf, Anton-Erxleben, Pieper, & Treue, 2006). However, this receptive field shifting predicts not a contraction but an expansion of visual space near the attended location, which presumably allows for greater spatial resolution at attended locations. Behavioral paradigms where participants are asked to localize objects near attended locations have documented this "attentional repulsion" effect, wherein briefly flashed objects are perceived as appearing further away from an attended location than they actually are (Binda, 2011; Pratt & Turk-Browne, 2003; Suzuki & Cavanagh, 1997). Thus, it seems unlikely that attention's influence on visual space could have produced the patterns of primacy we observed for judgments of mean location.

A second and perhaps more likely possibility for how attention as captured by the first dot could result in overweighting of subsequent dots is that once attention is captured by the sudden appearance of the first dot, its benefits linger long enough such that subsequent dots in the area also receive them. Since all dots on a given trial are sampled from the same Gaussian distribution, early dots do predict the location of later ones. However, it may be unlikely that the spatial extent of attention after it has been captured is large enough to reliably benefit subsequent dots (Henderson & Macquistan, 1993; Shulman, Wilson, & Sheehy, 1985), which appear on average 4° away from their predecessors. Furthermore, most research has shown that the benefits of exogenously summoned attention peak around 100-200 ms after the capturing event and are usually gone once 300 ms have passed (Egeth & Yantis, 1997; Müller & Rabbitt, 1989; Nakayama & Mackeben, 1989; Posner & Cohen, 1984). Since a new dot appeared every 200 ms in our method, it is possible that the second dot received residual attentional capture benefits, but they were likely gone by the arrival of the third dot. So this account could possibly explain large (and unequal) weights for the first two dots, but it does not easily account for the continued decline in dot weights across the rest of the sequence that we often observed (see Appendix A).

One way to test whether attentional capture is responsible for primacy in mean location judgments is to employ a salient, attention-capturing precue before the dot sequences. Such a

cue, especially one far from the location of subsequent dots, would prevent attentional benefits from affecting the first dots. If primacy disappears in the presence of such a cue, then it is highly likely that capture produced by the first dot is at least partially responsible for it. While such an experiment is entirely feasible, it should be noted that the attention-related accounts for serial order effects discussed here predict only primacy, and cannot reasonably produce recency. Thus, if the attentional capture explanation is accurate, another account would be needed to explain recency in the domains where we found it.

3.1.3 *Eye movements*

A final explanation for serial order effects in estimating the mean of a set presented over time involves systematic mislocalizations of dots resulting from eye movements that presumably occur over the course of a trial. In Experiment 1 described in Chapter 2, participants were free to look wherever they wished during stimulus presentation. It is logical to assume that participants began each trial with their gaze in the center of the screen, since it represented the average dot location across many trials. However, since the first dot to appear on each trial was predictive of where the following dots would appear, participants likely foveated it once it appeared in order to better process subsequent dots. If participants continued this strategy of moving their gaze to the expected location of the next dot after each dot appeared, then at least two known biases in object localization could have operated to produce primacy in their estimates of the mean.

The first bias involves systematic mislocalizations of objects near the time of saccadic eye movements. Many researchers have documented this phenomenon, wherein objects flashed between about 100 ms before and 100 ms after a saccade initiation are perceived as appearing closer to the saccade destination than they actually are (Mateeff, 1978; Matin & Pearce, 1965; Morrone, Ross, & Burr, 1997; Ross, Morrone, & Burr, 1997; Schlag & Schlag-Rey, 1995). Morrone and colleagues (1997) showed that the effect can be drastic, with objects being mislocalized by up to 10° in the direction of the saccadic target for large (20°) eye movements. These mislocalizations are considered to be a consequence of the predictive remapping of visual space that occurs near the time of saccades: part of the reason why we don't experience drastic visual motion when we move our eyes. However, if participants in our task are moving their eyes after each dot appears, it's very possible that next dot (200 ms later) is being mislocalized as a

consequence, since these effects are present for up to 130 ms after saccade initiation and normal saccadic latencies are 80-200 ms (Carpenter, 1977; B. Fischer & Boch, 1983). Importantly, these mislocalizations could produce primacy, since the perceived location of each dot would be pulled towards that of the previous one.

Interestingly, briefly flashed objects, especially on dark backgrounds, can be systematically mislocalized even without eye movements (Mateeff & Gourevich, 1983; Osaka, 1977; Sheth & Shimojo, 2001; van der Heijden, van der Geest, de Leeuw, Krikke, & Müsseler, 1999). Participants in these experiments, when asked to determine the position of briefly occurring stimuli on a sparse visual background, consistently estimate their position as closer to the fovea than their actual position. Assuming that participants in our dot task are gazing at their current belief about the mean location of the group by the time that the next dot arrives, this effect could mean that later dot locations are again pulled towards those of early dots, or beliefs based on them. Just as in saccade-induced mislocalizations, this could produce primacy for estimates of the mean of the group.

These eye movement-related explanations for primacy could be tested by simply asking participants to hold fixation over the course of dot presentation on each trial. Even if dots are consistently mislocalized towards the fovea in this case, the effect would be constant across early and late dots and could not reasonably produce primacy by itself. If primacy for mean location is diminished or eliminated when fixation is enforced, then it is likely that biases related to eye movements are responsible for the effect.

3.1.4 *The present study*

Our previous experiments revealed but did not explain the presence of serial order effects in summary computation of mean location and mean size for a series of items presented over time. Thus, the following three behavioral experiments, all employing extensions of the basic method used in Experiment 1 of Chapter 2, were conducted to test the three explanations for primacy and recency introduced above.

3.2 EXPERIMENT 1: MANIPULATING SEQUENCE LENGTH

To understand whether capacity limitations in visual short-term memory might be responsible for primacy and recency for judgments of mean size and mean location, we introduced a sequence length manipulation into the location and size task used in Experiment 1 of Chapter 2. If primacy and/or recency are evident in longer sequences, but disappear or are attenuated for shorter sequence lengths, this would constitute evidence that memory limitations are at least partially responsible for the serial order effects we observed in Chapter 2.

3.2.1 *Method*

3.2.1.1 Participants

Twenty six students at the University of Washington were recruited from the Department of Psychology's undergraduate participant pool, where students may volunteer as participants for studies in exchange for course credit. All participants had normal or corrected-to-normal vision. Recruitment and study procedures in all experiments presented here were conducted in accordance with the ethical policies set forth by the University of Washington's Human Subjects Division, and those in the Declaration of Helsinki.

3.2.1.2 Apparatus

All study procedures took place in a dimly lit room, with the participants seated 50 cm from a CRT monitor subtending 40.4° x 30.8° of visual angle. A chinrest was used to ensure constant viewing distance to the monitor, on which all stimuli were shown against a black background. A faint grey grid pattern was always present as part of the background in order to provide spatial reference. All stimuli were generated by custom software written using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) for MATLAB (The Mathworks, Natick, MA).

3.2.1.3 Stimuli & procedure

The stimuli and procedure used here was nearly identical to that described in Experiment 1 of Chapter 2. On each trial, participants saw a series of dots that varied in their location and in size and then were asked to report either the average location or size of the group. However, on a given trial in this experiment, the series shown could consist of one, two, four, or eight dots.

On each trial, participants were precued with the word “Location” or “Size” presented slightly above a white central cross approximately 0.8° in width and height. Despite the presence of the central cross, no instructions about fixation were given and participants were free to look wherever they wished over the course of the experiment. This word precue was present for a random period of time between one and two seconds and indicated what the participant would be asked to report at the end of the trial.

Each trial consisted of a series of white dots that varied in their position and size. Each dot was shown for 150 ms and was followed by a 50 ms blank inter-dot interval, resulting in a dot presentation rate of 5 Hz. On a given trial, dot locations were chosen by sampling ten times from a bivariate Gaussian probability distribution with a standard deviation of 2.3°. The center of the bivariate Gaussian distribution was drawn randomly on each trial from a square-shaped uniform probability distribution centered on the middle of the screen and subtending 21.1° x 21.1°, corresponding to 70% of the vertical height of the screen. If a dot’s sampled location was outside the borders of the screen, it was moved to the point on the screen nearest to its originally sampled location. Since the perceptual size of 2-D discs has been shown to vary approximately with the radius raised to the 1.52 power (Teghtsoonian, 1965), dot size was sampled in this perceptual space, then converted to radius space for display on the screen. The dot sizes used on each trial were sampled from a Gaussian distribution with a standard deviation of 0.4 of these perceptual units. The center of the Gaussian itself was sampled on each trial from a uniform distribution covering perceptual sizes 0.6 to 2.5, corresponding to dot radii of 0.7° to 1.8°. If any dot size was sampled such that its radius was less than 0.22° the trial was resampled using the same Gaussian distribution until no dot was below this size. The series of dots was followed by a blank period of 300 ms, followed by a response period.

During the response period, participants were reminded what to report by text appearing near the middle of the screen. On location trials participants reported the “average location” or “center” of the dots seen on that trial by moving the mouse cursor and clicking on their perceived center. The mouse cursor was visible to the participants only during the response period of location trials. On size trials participants reported the “average size” of the dots seen on that trial by adjusting the size of a centrally-presented test dot, the radius of which varied with the horizontal location of the mouse. Participants clicked to submit their response when the perceived mean

size was obtained. The test dot's initial radius was chosen randomly on each trial from 0.2° to 2.9°, which also served as the limits of possible responses. The response period ended when the participant submitted his or her response, and was followed by a 1500 ms inter-trial interval.

Each participant received full instructions from an experimenter and then completed approximately sixteen practice trials (from a variety of set size and response type conditions) in view of the experimenter before beginning the experimental trials. Each participant completed 16 blocks of 40 trials, with 2 blocks allocated for sequence length one, 2 blocks for sequence length two, 4 blocks for sequence length 4, and 8 blocks for sequence length eight. Within each sequence length, half of the blocks asked the participant to report the average size and half asked the participant to report the average location. Participants were made aware of what set size and response type would be used ahead of each block, and the order of blocks shown was randomly determined for each participant. Each participant completed the 640 experimental trials in two sessions on separate days, with each session lasting approximately 50 minutes. The participants were free to take breaks after each block, but could also do so at any point during the experiment by simply waiting to submit their response for a given trial.

3.2.2 *Results & Discussion*

Just as in Chapter 2, weights that quantified the relative influence of each dot number on participants' responses were obtained by fitting a weighted average model (as used in Juni, et al., 2012) to each participant's data separately for size and location trials, and also separately for each sequence length. The model took the form

$$R_j = \sum_{i=1}^{10} w_i x_{ij}$$

where R_j is the participant's response for trial j, x_{ij} is the position or radius of the dot at temporal position i and trial j, and w_i is the weight for temporal position i. Linear regression was used to obtain the least-squares best fitting set of weights for each participant for each task. Data from one participant was thrown out due to excessive response error; his mean error was approximately three times that of all other participants across all conditions. The model fit the data well in the remaining 25 participants in both location (model R^2: mean = 0.99, SD = 0.01; all $p < 0.001$) and size (model R^2: mean = 0.73, SD = 0.14; all $p < 0.001$) trials. Mean weights

across participants for both trial types and for sequence lengths two, four, and eight are shown in Figure 3-1, with error bars depicting 95% confidence intervals.

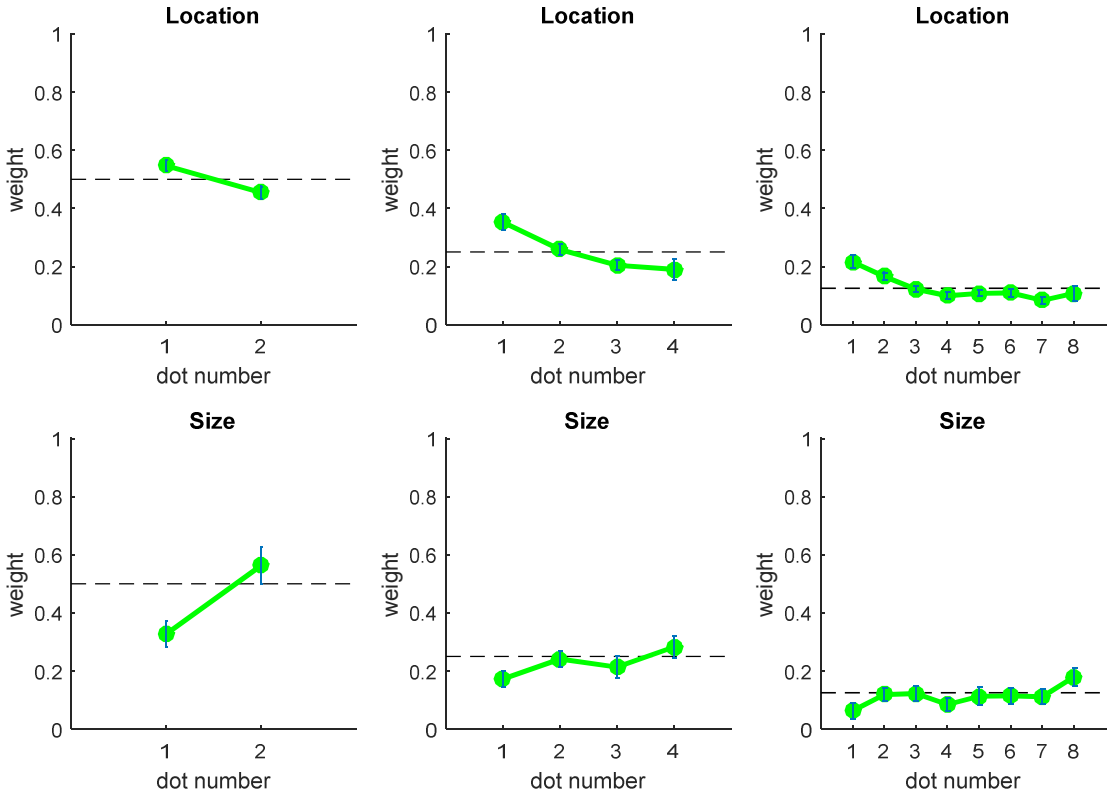

Figure 3-1. Results from Experiment 1.

The dashed lines in Figure 3-1 show the weight that would be obtained for each dot in that condition if all dots contributed equally to participants' responses and no other source of noise or bias were present. Just as in Experiment 1 in Chapter 2, primacy was apparent for judgments of mean dot location, while recency was apparent for judgments of mean dot size. The sizes of the effects were considerable, with the most weighted dots often contributing 1.5 to 2.0 times as much as the least weighted dots. Notably, the effects appeared not only in the eight-item sequences, but also in the four- and two-item sequences, with two-way ANOVAs on the weights showing significant dot number x response type interactions for sequence length eight, $F(7,336) = 15.30$, $p < 0.001$, sequence length four, $F(3,144) = 22.32$, $p < 0.001$, and even sequence length two, $F(1,48) = 36.28$, $p < 0.001$. Furthermore, six one-way ANOVAs testing for the effect of dot

number on the computed weights in each condition were all significant, all ps ≤ 0.001. Together these results indicate that primacy was observed for all sequence lengths in the mean location task and recency was observed for all sequence lengths in the mean size task.

Even if temporal order effects are still statistically observed for short sequences, it is possible that primacy or recency is attenuated compared to longer sequences. Since our obtained weights are not directly comparable across different sequence lengths, we quantified the strength of primacy and recency in each sequence length by obtaining weights for comparable parts of the sequences. For example, the strength of primacy in the location task can be compared across sequence lengths by obtaining weights for only the first two dots in each of the lengths. The same can be done for the first four dots in sequence lengths four and eight. The results of this analysis for both the location task and the size task are shown in Figure 3-2.

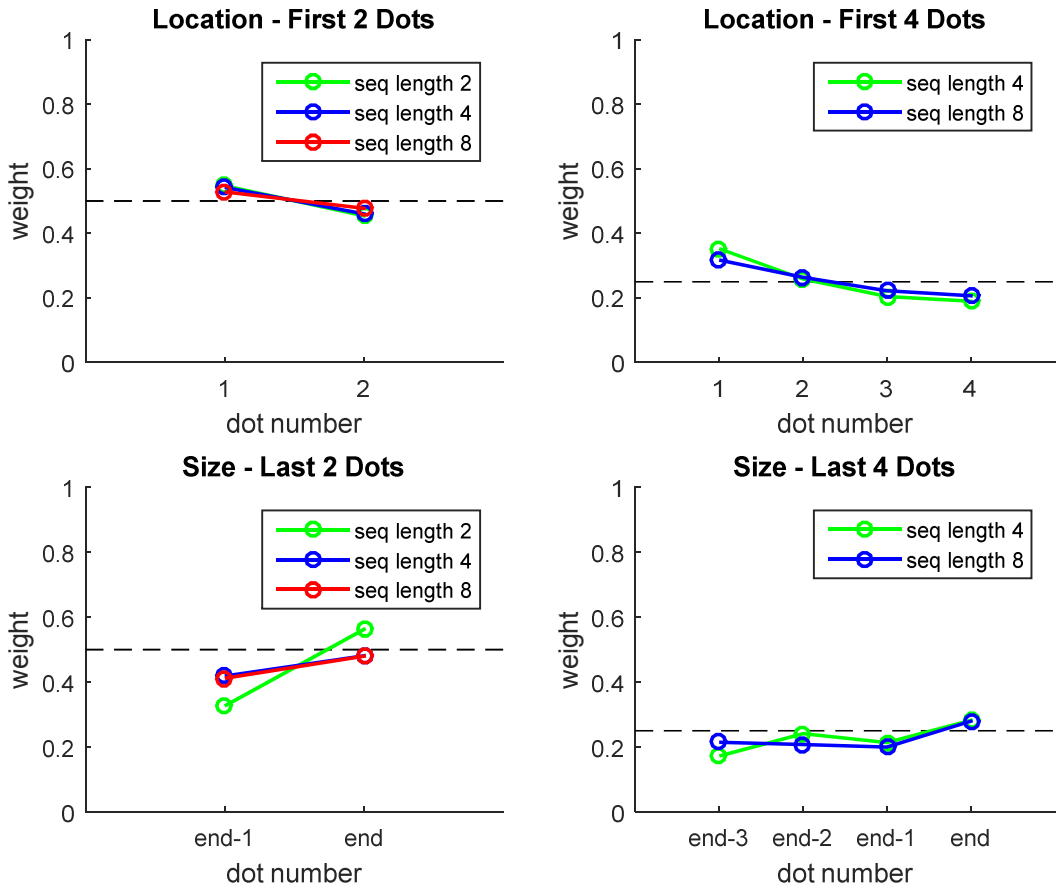

Figure 3-2. Comparing primacy and recency across sequence lengths.

If primacy were attenuated in the shorter sequence lengths for the location task, we would expect to see that in weights computed for the first two dots of each sequence length. However, the difference between the first two weights appears to be comparable across all sequence lengths, and a two-way ANOVA supports this, revealing no significant dot number x sequence length interaction, $F(2,72) = 1.55$, $p = 0.22$. Similarly, no significant difference in primacy is found when comparing weights for the first four dots in sequence lengths four and eight, $F(3,144) = 2.05$, $p = 0.11$. In the size task, comparing weights for the first four dots between sequence lengths four and eight also reveals no differences in the strength of recency, $F(3,144) = 1.59$, $p = 0.19$. Interestingly, comparing weights for the first two dots between sequence lengths two, four, and eight in the mean size task does reveal a significant dot number x sequence length interaction, $F(2,72) = 7.19$, $p = 0.001$, but the effect appears to be opposite of what might be expected: recency actually appears to be stronger for shorter sequences in this case.

Together, these results suggest that visual short term memory capacity is not the primary cause of primacy or recency in mean judgments from stimuli presented over time. If primacy in mean location judgments was caused by the early items in a sequence taking up all available memory resources and leaving none for the later items, for example, then we would expect to find either no primacy or attenuated primacy for cases where only two or four items were presented. This is because storing each item in those sequences should require considerably fewer resources than in eight- or ten-item sequences. However, neither primacy nor recency appeared to diminish in the shorter sequences, suggesting limited short term memory capacity is not directly responsible for either effect.

3.3 Experiment 2: Manipulating attention

In order to test whether attentional capture by the first dot in each sequence produced primacy for judgments of mean location, we preceded some trials with salient but irrelevant precue dots and compared weights obtained for these trials to weights obtained for trials with no precue dot.

3.3.1 *Method*

The method used here was nearly identical to the mean location tasks described previously in this document; participants saw a series of dots on each trial and judged the mean location of the group. The primary exception to the normal method was that in two thirds of blocks, an irrelevant dot with unique color preceded the main dot sequence. On these trials, participants were told to average only the main sequence of dots, and to ignore the cue.

3.3.1.1 Participants

Twenty two students at the University of Washington were recruited from the Department of Psychology's undergraduate participant pool, where students may volunteer as participants for studies in exchange for course credit. All participants had normal or corrected-to-normal vision. Recruitment and study procedures in all experiments presented here were conducted in accordance with the ethical policies set forth by the University of Washington's Human Subjects Division, and those in the Declaration of Helsinki.

3.3.1.2 Apparatus

All study procedures took place in a dimly lit room, with the participants seated 50 cm from a CRT monitor subtending 40.4° x 30.8° of visual angle, on which all stimuli were shown against a black background. A faint grey grid pattern was always present as part of the background in order to provide spatial reference to the participants. All aspects of the experiment were controlled by custom MATLAB (The Mathworks, Natick, MA) software, making use of routines included in the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997). A chinrest was used to ensure constant distance between the participant and the monitor.

3.3.1.3 Stimuli & procedure

Just as in Experiment 1, each trial began with the presentation of a white central cross approximately 0.8° in width and height. The cross was present for a random period of time between 1 and 2 s before the dot sequence for that trial was shown. Each dot sequence consisted of a series of six small (radius 0.4°) colored dots that varied in their position. Each dot was shown for 150 ms and was followed by a 50 ms blank inter-dot interval, resulting in a dot presentation rate of 5 Hz. On a given trial, dot locations were chosen by sampling six times from a bivariate Gaussian probability distribution with a standard deviation of 2.3°. The center of the

bivariate Gaussian distribution was drawn randomly on each trial from a square-shaped uniform probability distribution centered on the middle of the screen and subtending 21.1° x 21.1°, corresponding to 70% of the vertical height of the screen. If a dot's sampled location was outside the borders of the screen, it was moved to the point on the screen nearest to its originally sampled location.

Unlike previous dot location tasks described in this dissertation, three precueing conditions were used here. In the first, no precue dot was used and the sequence of six dots was presented as described above. In the second and third conditions, a precue dot that differed in color from the rest was presented immediately before each sequence. In a random half of participants, the precue dot was pink and the main sequence dots were green, while in the other half the colors were reversed. Just as with the main sequence dots, the precue dot had radius 0.4° and was present for 150 ms, followed by a 50 ms blank inter-dot interval. There was no additional interval between the precue dot and the start of the main sequence. In the second precue condition (the 'near precue' condition), the location of the precue dot was sampled from the same distribution used to obtain dot locations for that trial's main sequence. In the third precue condition (the 'far precue' condition), the precue dot's location was determined by choosing a random point on the screen at least 8° from the sampling distribution center for that trial. Precue condition was blocked such that one third of blocks contained only trials with no precue and the remaining two thirds of blocks contained trials from both precue conditions randomly interleaved within each block. Thus, from the participants' perspective there were precue blocks and no-precue blocks. This was done to equate the allocation of willful attention on the dots across all trials and examine the effect specifically of exogenous attention on dot weights. This can only be done if participants know whether they will need to willfully attend to the first dot that appears or not.

The series of dots was followed by a blank period of 300 ms, followed by a response period. During the response period, participants reported the "average location" or "center" of the main dot sequence seen on that trial (ignoring any precue dot) by moving the mouse cursor and clicking on their perceived center. The mouse cursor was visible to the participants only during the response period. The response period ended when the participant submitted his or her response, and was followed by a 1500 ms inter-trial interval.

Each participant received full instructions from an experimenter and then completed approximately twelve practice trials (including trials from all precue conditions) in view of the experimenter before beginning the experimental trials. As in other experiments where salient exogenous attention cues are used, participants were told to ignore the precue dot and to only average the main dot sequence on each trial. Each participant completed 9 blocks of 40 trials, with precue and no-precue blocks appearing in random order. Participants were made aware of whether precues would be present ahead of each block, and were reminded what color the precue would be and what color the main sequence would be. Each participant completed the 360 experimental trials in approximately 50 minutes. The participants were free to take breaks after each block, but could also do so at any point during the experiment by simply waiting to submit their response for a given trial.

3.3.2 *Results & Discussion*

Weights describing the relative influence of each dot number on participants' responses were obtained for each participant for each of the three cue conditions via the same method used in Experiment 1 above and in Chapter 2. Mean weights across all participants for each of the three precue conditions are plotted in Figure 3-3 below, with error bars depicting 95% confidence intervals.

Visual inspection of the weights shows that primacy is still clearly present in all three conditions. On average, the early dots contribute more to the mean location judgments, with the first dot influencing the response approximately 1.5 to 2 times as much as the last dot. A second observation is that the weighting profile seen as a function of dot number is very similar across the three cueing conditions, with no obvious effect of the precue on the dot weights seen in the main sequence. In order to verify this, a two-way ANOVA was conducted on the weights. Unsurprisingly given the clear primacy effect, a significant main effect of dot number was found, $F(5,105) = 18.43$, $p < 0.001$. However, crucially, no dot number x cue condition interaction was found, $F(10,210) = 1.03$, $p = 0.42$, providing no evidence of an influence of cueing condition on weights obtained in the three cueing conditions. The precue dot appeared to have little effect on how the dots in the main sequence were combined into an average.

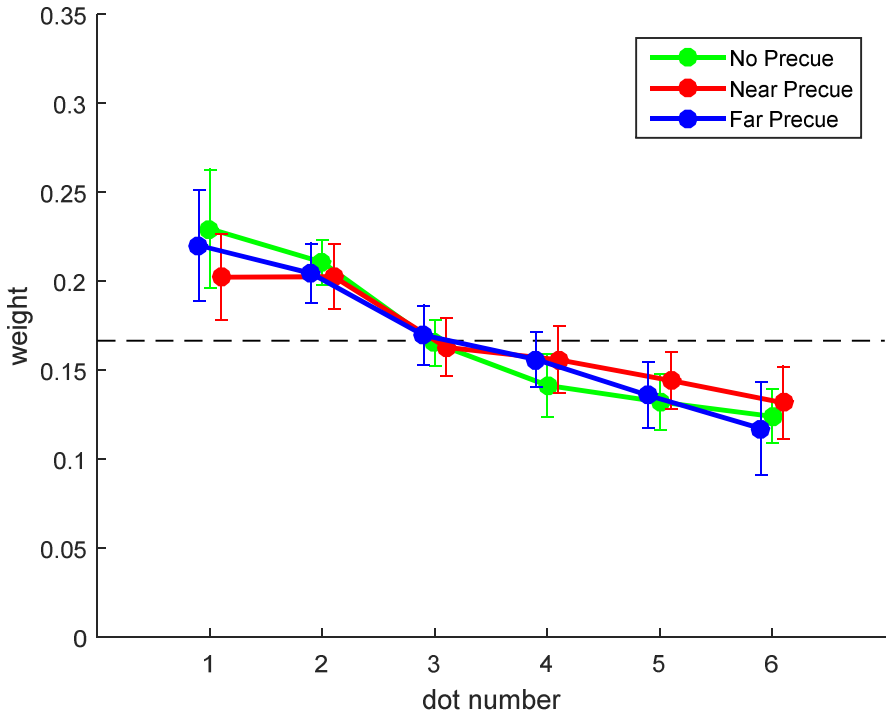

Figure 3-3. Results from Experiment 2.

It is possible that the precue dot did have a small effect on the main sequence weights, but that the relatively small number of trials (120 per participant) in each of the cue conditions made it difficult to observe. In order to increase the chances of observing a true effect of cue condition, we combined the near and far precue condition trials into a single precue condition and plotted the resulting weights against those from the no cue condition in Figure 3-4 below, with error bars depicting 95% confidence intervals. Once again, however, no significant dot number x cue condition interaction was found, $F(5,105) = 1.59$, $p = 0.17$. Again, no evidence was found that the precue dots affected the weighting of the relevant dots.

Together, these results suggest that primacy remained present and undiminished despite the dot sequence being preceded by a salient precue dot designed to capture attention in the same way that the first dot in a normal dot sequence would. If processing benefits associated with exogenously-attracted attention were responsible for primacy for mean location, then we would have expected to see a flatter set of weights in the precue conditions than in the no precue

condition. Thus, we conclude that primacy for judgments of mean location is not primarily caused by attention being attracted to the first dots in each sequence.

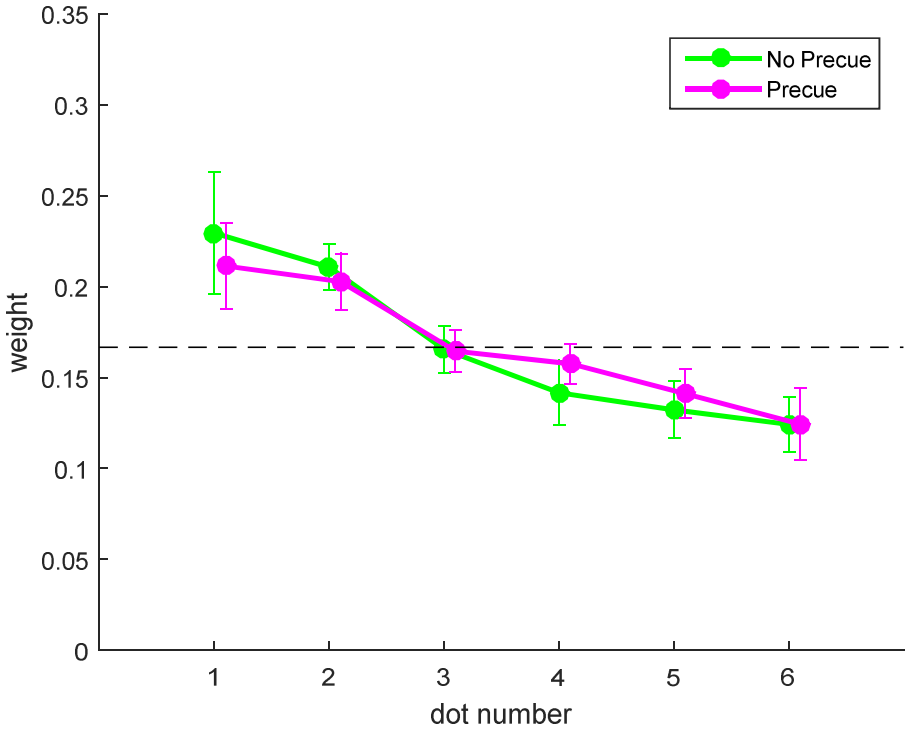

Figure 3-4. Results from no precue versus precue trials.

3.4 Experiment 3: Enforcing Fixation

All previous experiments described in this thesis have allowed free viewing of the stimuli as they are presented. In tasks where dots appeared away from the center of the screen (a very common occurrence for the mean location tasks), it is logical to assume that participants adjusted their gaze over the course of each trial. In Experiment 3, we investigated whether eye movements during stimulus presentation contributed to the primacy effect we have consistently observed for judgments of mean location of stimuli presented over time. We did so by using eye tracking to enforce fixation throughout the stimulus presentation phase of each trial.

3.4.1 *Method*

The method used here was nearly identical to mean location tasks described previously in this document; participants saw a series of dots on each trial and judged the mean location of the group. The primary exception to the normal method was that participants here were asked to fixate before and during stimulus presentation, while their eye position was recorded with a desk-mounted eye tracker.

3.4.1.1 Participants

Nine members of the University of Washington's Department of Psychology were recruited for participation, for which they were compensated at $10/hour. All participants had normal or corrected-to-normal vision. Recruitment and study procedures in all experiments presented here were conducted in accordance with the ethical policies set forth by the University of Washington's Human Subjects Division, and those in the Declaration of Helsinki.

3.4.1.2 Apparatus

All study procedures took place in a dimly lit room, with the participants seated 70 cm from a CRT monitor subtending 29.2° x 22.4° of visual angle, on which all stimuli were shown against a black background. A faint grey grid pattern was always present as part of the background in order to provide spatial reference to the participants. Stimulus presentation and eye tracker data collection was performed by custom MATLAB (The Mathworks, Natick, MA) software, making use of routines included in the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997)

An ASL D6 desktop eye tracker running at 60 Hz sat below and in front of the stimulus presentation monitor, approximately 55 cm from the participant. The eye tracker and stimulus presentation monitors were controlled by separate computers connected on a local network. During the recording portions of the experiment, the stimulus presentation computer obtained and stored the current eye position from the eye tracking computer on average every 11 ms. A chinrest was used to ensure constant distance between the participant and the monitor and eye tracker. The experimenter was present in the room for all experimental procedures.

3.4.1.3 Stimuli & procedure

Data collection was performed over 10 blocks of 40 trials for each participant, with each block lasting approximately six minutes. Prior to each block, participants were asked to position themselves in the chinrest and to minimize head movement until the block was complete. A short nine-point eye tracking calibration procedure was performed immediately before each block, after which participants began the trials.

Each trial consisted of pre-trial, stimulus, and response phases. A white fixation cross, approximately 0.8° in width and height, was present in the center of the screen during the pre-trial and stimulus phases. At the start of the experiment, participants were instructed to keep their gaze on the fixation cross and minimize blinking whenever it was present. The pre-trial phase lasted between 1.7 and 2.2 s, and eye position was recorded starting 1 s into the phase. In the stimulus phase, a series of eight small (radius = 0.4°) white dots were presented one at a time. Each dot was shown for 150 ms and was followed by a 50 ms blank inter-dot interval, resulting in a dot presentation rate of 5 Hz. On a given trial, dot locations were chosen by sampling eight times from a bivariate Gaussian probability distribution with a standard deviation of 2.3°. The center of the bivariate Gaussian distribution was drawn randomly on each trial from a square-shaped uniform probability distribution centered on the middle of the screen and subtending 15.3° x 15.3°, corresponding to 70% of the vertical height of the screen. If a dot's sampled location was outside the borders of the screen, it was moved to the point on the screen nearest to its originally sampled location. Eye position was recorded throughout the stimulus phase. Once the last dot disappeared, the fixation cross was removed, eye position recording stopped, and the response phase began.

The response phase began with a blank period of 300 ms, after which the mouse cursor appeared in a random screen position. Participants then reported the "average location" or "center" of the dots seen on that trial by moving the cursor and clicking on their perceived center. Since responses were unspeeded and the fixation cross was not present, participants were encouraged to rest, blink, and look wherever they liked during this phase. The response period ended when the participant submitted his or her response, and was followed by a 1500 ms inter-trial interval.

Prior to any data collection, participants were briefed on the eye tracking procedures, including instructions related to maintaining fixation, blinking, and minimizing head movement in the chinrest. Participants also received instructions about the experimental task and completed ten to twenty practice trials without eye tracking prior to the experimental trials. Each participant completed the 400 experimental trials in two sessions on separate days, with each session lasting approximately 40 minutes. The participants were encouraged to take breaks after each block at their discretion. If necessary, the experimenter reminded participants between blocks to maintain fixation and minimize blinking whenever the fixation cross was present.

3.4.2 *Results & Discussion*

Since the purpose of Experiment 3 was to determine whether primacy is still present for mean location judgments when participants are fixating, we identified and discarded trials where eye movements were made during stimulus presentation according to the following procedure. First, all eye position samples from the pre-trial and stimulus phase eye traces that indicated a gaze outside the borders of the computer monitor were removed and treated as signal losses. Then, pre-trial eye position baselines were computed as the average x and y eye positions over the 300 ms preceding stimulus presentation on each trial. Finally, any trial where even a single eye position sample from the stimulus phase eye trace (when the dots were being presented) fell more than 1.5° from the pre-trial baseline was discarded from further analysis. All other trials were considered valid fixation trials. On average, this resulted in 5.7% of the data (~23 trials out of 400) being discarded in each participant, with 14.8% of trials being discarded from the most affected participant.

Weights quantifying the influence of each temporal position in the dot sequences were computed for each participant just as in previous experiments in this chapter and in Chapter 2. Mean weights for the group are shown in Figure 3-5, with error bars depicting 95% confidence intervals. As can be seen, primacy is still clearly evident despite measures taken to enforce fixation throughout stimulus presentation, with the first dot contributing considerably more to responses than any other dot. A significant effect of dot number in a one-way ANOVA on the weights confirms this, $F(7,56) = 9.77$, $p < 0.001$. However, informal comparisons to weighting profiles obtained in previous mean location tasks suggests that primacy is diminished in the

present data, and may only be present in the overweighting of the first dot. In tasks where participants have been free to move their eyes as the dots appear, primacy has tended to appear throughout the whole weight profile, with weights gradually decreasing to a minimum at the last dots. The present pattern appears notably different. A second, post hoc one-way ANOVA reveals no significant effect of dot number when only considering weights from dots two through eight, $F(6,48) = 1.76$, $p = 0.13$, consistent with this interpretation. These results suggest that primacy for judgments of mean location is significantly diminished, though not totally eliminated, when eye movements are minimized during the presentation of the objects to be summarized.

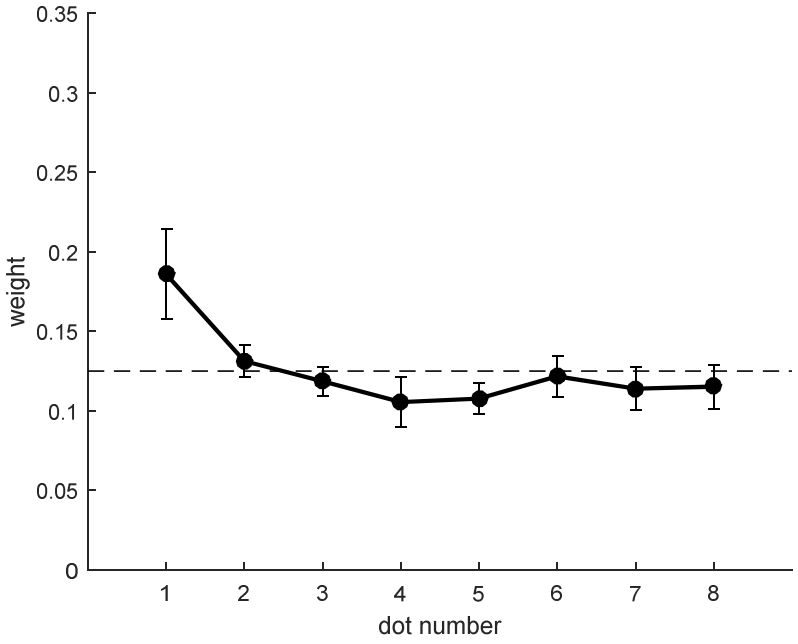

Figure 3-5. Results from Experiment 3.

3.5 General Discussion

Experiments reported on in Chapter 2 revealed prominent serial order effects in estimates of the mean for visual objects presented across time. Specifically, primacy was found for estimates of the mean location of a series of dots, and recency was found for estimates of their mean size. Here we developed and tested three specific hypotheses about the cause of these primacy and recency effects: visual working memory limitations, attentional capture, and localization biases introduced by eye movements. Experiment 1 used a sequence length manipulation to show that

neither primacy nor recency was diminished when summarizing short dot sequences that should not have taxed memory capacity. Experiment 2 used an attention capturing but irrelevant precue dot to show that primacy for judgments of mean location was not caused by exogenous attention benefits for the early dots.

However, in Experiment 3, participants showed greatly diminished primacy for estimates of mean location when they were forced to fixate during dot presentation. In our data, the first dot was still overweighted compared to the rest, but unlike in previous experiments, weights for dots after the first were not statistically distinguishable from each other. These results suggest that primacy as observed in Chapter 2 was at least partially caused by eye movements that likely occurred while the dot sequence was being presented. The most plausible mechanism for this would appear to be one (or both) of the two object localization biases discussed in the introduction. In the first, compression of visual space occurring near the time of saccades (Morrone, et al., 1997; Ross, et al., 1997) could have pulled the perceived location of newly arriving dots towards previous dots, resulting in higher weights for early dots. In the second, object mislocalizations toward the fovea (Mateeff & Gourevich, 1983; Osaka, 1977; Sheth & Shimojo, 2001; van der Heijden, et al., 1999) could have produced higher weights for early dots if participants followed logical eye movement patterns, which is likely. The current data cannot distinguish between these two specific causes, but our primary conclusion is that primacy for judgments of mean location is largely caused by eye movements during item presentation.

3.5.1 *An insistent first dot*

In Experiment 3, we found that mean location weights for dots after the first were flat when fixation was enforced, but the first dot's weight still remained high. Why? One possibility is that the peri-saccadic compression of space discussed in the introduction was triggered without the actual saccade being executed. To the best of our knowledge no test for the presence of peri-saccadic compression in the absence of the saccade itself has been reported, but related work suggests that it is possible.

One important aspect of peri-saccadic compression that suggests it could operate independently of saccade execution is that the compression actually begins 100 ms or more before the onset of the saccade itself (Honda, 1993; Morrone, et al., 1997). It is thought to be a consequence of

predictive remapping of neurons in lateral intraparietal cortex (Duhamel, Colby, & Goldberg, 1992), superior colliculus (Walker, Fitzgibbon, & Goldberg, 1995), and frontal eye fields (Umeno & Goldberg, 1997), the purpose of which is to anticipate the new visual input that will be caused by the saccade. Importantly, however, this remapping is triggered by saccadic intention, and occurs before the onset of the saccade itself. A clever saccadic adaptation experiment conducted by Bahcall & Kowler (1999) further supports this possibility. Participants in their experiment made saccades to peripheral targets that surreptitiously moved from their original location during the saccade, such that the originally planned saccade systematically undershot the target. Continued exposure to this paradigm causes saccadic adaptation, where saccades consistently overshoot targets that do not move during the saccade. By measuring peri-saccadic mislocalizations in this situation, the researchers showed that the mislocalizations varied as a function of the planned saccade, but not the one that was actually executed. This indicates that peri-saccadic compression of space is not tied to saccade execution, but rather to saccadic intention.

Assuming that peri-saccadic compression can operate without saccade execution, then it is possible that it was triggered by a reflexive eye movement intention that was inhibited before execution. Participants in Experiment 3 often reported difficulty in maintaining fixation during trials, suggesting that reflexive saccades were being actively inhibited. The first dot in particular likely elicited the start of a reflexive saccade response, since it was a highly anticipated, salient visual object that appeared at an unpredictable time against a sparse visual background. If peri-saccadic compression occurred in response to the appearance of the first dot, then it likely affected the perceived location of the second dot, pulling it towards the first. This, in turn, would produce an increased weight for the first dot but flat weights for the rest, which is exactly the pattern we observed in Experiment 3.

This account is admittedly hypothetical, since it is not known whether peri-saccadic compression of visual space can occur for inhibited saccades, but this possibility is attractive in its ability to explain the insistently high weight of the first dot that we found in Experiment 3. Empirical testing for peri-saccadic compression for planned but not executed saccades would allow for more firm conclusions.

3.5.2 *Elusive recency*

Despite further investigation, the causes of recency for summaries of mean object size, facial expression, and motion direction remain unclear. Here we briefly discuss two possible routes forward for understanding why recency is observed in certain feature domains.

One possibility is that sequences of samples in recency-exhibiting domains interfere with each other retroactively. In other words, the arrival of new samples interferes with either the perception or storage of preceding samples. Backwards masking of samples is an example of how visual items presented in a rapid sequence could perceptually interfere with each other. This could produce recency if the later items mask or perceptually degrade the earlier items, but are not themselves degraded since no items follow them. This hypothesis could be tested by adding irrelevant masking stimuli to the end of sequences and examining the resulting weights for reduced recency. Backwards interference could also occur in memory. This type of explanation is similar to one given in the memory capacity hypothesis developed in the introduction, where later items "steal" limited memory resources from earlier items. However, the difference is that limitations in storage capacity play no role here. The first item might have all the memory resources it needs, but if a second item demands to be stored before consolidation of the first is complete, then backwards interference might result and produce recency. In this hypothesis, the limiting factor that leads to poor memory representations for early items is a bottleneck in the storage process and not the in the storage resources. This hypothesis was partially tested in Experiment 5 of Appendix A, where two dot presentation rates were used. In theory, a slower presentation rate would relieve a storage process bottleneck, but no effect of presentation rate on recency was observed. Further testing of this hypothesis is needed.

One possible mechanism for recency was discussed in Chapter 2 but was not tested here. Adaptive gain control refers to adjusting the response range of a signal-receiving system such that most of the dynamic range is situated at the value of the signal that is predicted to come next. In this way, even small differences between the predicted and arriving signal are amplified. In cases where the signal that arrives next is predictable, this strategy is an effective way to increase the overall sensitivity of a system with limited dynamic range. Cheadle and colleagues (2014) found behavioral, pupillometric, and neuroimaging evidence that human participants

apply a process very like this one when asked to compute the average of a series of oriented gratings presented serially across time. The amount of influence a particular sample had on the reported average was largely determined by how close in orientation it was to the previous sample. Samples consistent with previous ones were most diagnostic of responses, the hallmark of an adaptive gain control. One interesting consequence of adaptive gain control is that it leads to recency when a series of samples come from the same distribution and are therefore predictive of each other. This is because by the time that the later samples arrive, the system has developed an accurate prediction about what their values will be and is thus highly sensitive to even small deviations from the prediction. Thus the later samples exert more influence over the final belief. Determining whether this model fits the present data well, or designing further experiments that would allow application of the model, would help determine whether adaptive gain control could have caused our findings of recency for mean size, mean facial expression, and mean motion direction.

3.5.3 *Belief updating*

One notable aspect of the results from Experiment 1's sequence length manipulation is what it indicates about how participants are completing our tasks. We have borrowed the terms primacy and recency from memory research, where they describe particular patterns of performance for recalling lists of to-be-memorized items, usually words. This evokes the idea that participants in our tasks are attempting to store each of the dots in a given series in memory as they appear, then are computing the average afterwards based on those memory representations. This is explicitly part of the memory capacity hypothesis developed and tested here. However, the fact that primacy and recency are both still quite recognizable for sequence lengths of two dots suggests that participants are not at all using a memorize-then-average approach.

A different strategy that participants might have been using is a belief updating strategy. In this strategy, a belief about the group average is developed after (or even before) the first dot appears, and is then updated after each new dot until the sequence ends, at which point the current belief is reported as the participant's response. From the participant's perspective, this strategy is attractive for sequences longer than a few of items since only three memory representations need to be intact at any given moment: that of the current belief, that of the current dot, and the

number of dots incorporated into the current belief. If the most recent dot's representation is discarded after it is used to update the belief, then theoretically, no additional memory is needed to produce perfect averages for sequences up to infinite length. In real-world settings where the exact number of items that need to be averaged is not always known ahead of time, this strategy seems particularly appealing since a current belief is always ready for report or decision-making.

From a researcher's perspective, a belief updating framework is useful because it brings with it a variety of tools that can be used to model participant responses in specific ways. For example, the rules of Bayesian inference specify exactly how a new dot and the current belief should be combined based on the statistical uncertainty associated with each. This type of model is appealing for the present circumstances since primacy, recency, and even weighting have clear analogues in it. If early dots have low uncertainty associated with them compared to later dots (if participants are paying more attention to them, for example), the rules of Bayesian inference dictate that the final belief will represent those dots more, producing primacy. If all dots are represented with the same amount of uncertainty, then they all contribute equally to the final belief, producing even weighting. In a sense, the weighted average model that we have used throughout these experiments to analyze our data does represent this kind of approach, since the weights associated with each dot can be shown to be mathematically related to the precision of those dot representations in a Bayesian belief updating framework. A final advantage of the belief updating approach over the weighted average model, though, is that it naturally models the dynamics of the estimate of the mean over the course of each trial. The belief is available at any given point throughout the trial rather than only at the end. Though we have not applied this approach rigorously here, it does represent a promising way to move forward in trying to understand the present data, and might more closely reflect how participants actually complete this type of task.

Chapter 4. WHAT MAKES PEOPLE SEE RANDOMNESS?

4.1 INTRODUCTION

How do we determine what has a cause and what is simply the result of randomness? Answering this question is one of the core functions of human psychology, with much of our perceptual and cognitive machinery devoted to it. Randomness detection and its inverse, pattern detection, are vital human behaviors, since siphoning out regularities from random noise allows us to discern the underlying rules and systems that govern the world around us. This knowledge, in turn, allows us to adapt productively to our environment and maximize our chances of living happy, successful lives.

Given this, it is not surprising that understanding how we decide that something is random or patterned has implications for the full breadth of psychological study. Much of the physiology of human vision, for example, is devoted to detecting structure such as edges (Hubel & Wiesel, 1968), shapes (Pasupathy & Connor, 2002), and objects (Kanwisher, McDermott, & Chun, 1997) in visual noise. In early language learning, word segmentation depends crucially on recognizing close temporal associations between spoken sounds (Saffran, Aslin, & Newport, 1996), while grammar and syntax learning relies on discerning regularities in how words are arranged into sentences (Gomez & Gerken, 1999; Redington, 1998). Research in social psychology shows that effective social functioning depends highly on understanding relationships and associations between behaviors and the underlying emotions and motivations that generated them (S. C. Johnson, 2003; Warneken & Tomasello, 2006), and may be driven by processes that extract these patterns from the social noise over the course of development (Meltzoff, Kuhl, Movellan, & Sejnowski, 2009). Furthermore, how well people understand randomness has even been shown to underpin clinically-relevant behaviors, such as pathological gambling (Ladouceur, Sylvain, Letarte, Giroux, & Jacques, 1998; Ladouceur & Walker, 1996; Michalczuk, Bowden-Jones, Verdejo-Garcia, & Clark, 2011; Toneatto, Blitz-Miller, Calderwood, Dragonetti, & Tsanos, 1997), hallucinatory tendencies (Jakes & Hemsley, 1986), and belief in conspiracies and supernatural forces (Blackmore, 1985; Brugger, 1990; Kay, Moscovitch, & Laurin, 2010).

Finally and most practically, misunderstanding how randomness behaves can negatively affect important real-world decisions made based on data. For example, investors misattributing cause to randomness in stock market data stand to lose large amounts of personal wealth (Barberis, 2003; J. T. Johnson, G. J., 2005). Similar misinterpretations of patterns and noise made on data relevant to public policy, such as crime rates, economic indicators, educational test scores, or public opinion, could potentially be even more devastating.

4.1.1 *Perceiving & generating randomness*

Interestingly, despite its importance to many aspects of human functioning, many studies of randomness perception over the past half century show that humans seem to systematically misunderstand how random processes behave, at least when considering simple devices like a roulette wheel or a flipped coin. Evidence documenting our seemingly skewed idea of randomness is reviewed extensively elsewhere by Wagenaar (1972), Bar-Hillel & Wagenaar (1991), and Oskarsson et al. (2009), but see work by Nickerson & Butler for a counterargument (Nickerson, 2002; Nickerson & Butler, 2009). Importantly, however, two types of tasks are commonly used: perception tasks and generation tasks. In perception tasks, participants are shown sequences of outcomes, usually as if they were the result of a random Bernoulli process such as flipping a coin (e.g., HHHTTHTTHTTTH), and are asked to rate how "random" the sequence looks on some arbitrary scale. In generation tasks, participants are asked to imitate randomness by producing a set of outcomes that looks like the product of true randomness. The primary finding from both types of tasks is that participants seem to believe that a random device produces outcomes that alternate more than an actual random device does; participants seem to believe that true randomness does not often produce "streaky" outcomes where there are repeated events of one state, as in the sequence HHHHTTTHHH.

This general preference for over-alternation suggests that perception and generation tasks measure an at least partially overlapping conception of how randomness behaves. However, the two tasks are rarely used in the same study, and when they are, their results are almost never directly compared. Thus, this idea has so far gone mostly untested. One goal of the present study was to obtain extensive measurements for a particular type of stimulus from both randomness perception and randomness generation tasks. If there is a consistent internal concept of

randomness, it should express itself similarly in both. Observing such a result would help to bring together over fifty years of literature on these topics that so far have not been directly tied together. It would also help us to understand exactly what our idea of randomness is, and what sorts of judgments it governs.

4.1.2 *Predicting randomness perceptions*

A separate issue in randomness perception has to do with what influences our decision to identify a particular stimulus or series of events as "random." We currently know very little about what specific factors cause people to judge sets of outcomes as random-seeming. There is some indication that feeling a lack of control increases the perception of structure in random noise (Pittman, 1980; Whitson & Galinsky, 2008). Other work has shown that attributions of randomness to a sequence of events are affected by how people conceptualize the generating device (Ayton & Fischer, 2004). For example, participants evaluate a single set of coin flip outcomes very differently if they are asked to attend to the coin versus attending to the hand flipping it (Roney, 2009).

These influences on randomness perception come from factors external to the sequence of outcomes being judged. But what properties of the sequences themselves lead to the perception of randomness? The *representativeness account*, described in highly influential heuristics and biases research from the 1970s, argues that the extent to which a sequence of binary outcomes will be labelled as random depends on to what extent the sequence is "representative" of typical random generating processes (Kahneman & Tversky, 1974; Tversky & Kahneman, 1974). For example, those authors argue that a long series of heads in a set of coin flips is not representative of the fact that a fair coin is supposed to produce heads and tails with equal frequency in the long run. The representativeness account is a useful theoretical tool in that it predicts certain factors (e.g., how often a sequence streaks versus alternates) to be involved in randomness perception. But as noted by Falk & Konold (1997), the total "representativeness" of a sequence defies easy quantification.

Various other characteristics of a sequence can be easily quantified, however. Can the subjective impression of randomness be predicted based on simple numerical descriptors of a sequence of events? Previous work hints at factors contributing to the perception of randomness, but no

attempt has yet been made to use them to predict how random specific sequences might look to people. Thus, a second goal of the present study is to develop a model that will predict what people will see as random or non-random. Being able to successfully predict subjective judgments of randomness could have implications for a wide variety of important behaviors, such as investing, gambling, cryptography, as well as public policy decisions.

4.1.3 *The present study*

We conducted a series of experiments to a) compare results from perception and generation tasks, and b) to develop a model to predict perceived randomness for a given sequence. All experiments reported here used short binary sequences intended to represent hypothetical outcomes from flipping a coin five times, such as HHTHT. Using these stimuli, we found a very high degree of correspondence between perception and generation tasks, suggesting that people possess a consistent (if erroneous) concept of randomness that expresses itself similarly across different tasks. We also developed and fitted a model to people's randomness judgments and found that the perceived randomness of event sequences is highly predictable from simple statistical descriptors of the sequences themselves, and is generalizable to both longer and other types of sequences.

4.2 EXPERIMENT 1: PERCEIVING & GENERATING RANDOM SEQUENCES

4.2.1 *Experiment 1A: Perception task*

Typical experiments designed to estimate the perceived randomness of binary sequences use subjective reports such as rating scales. To avoid problems associated with subjective measures, we measured the perceived randomness of coin flip sequences using a two-alternative forced choice procedure and a subsequent ranking analysis to place each sequence on a dimension of perceived randomness.

4.2.1.1 Participants

Participants were recruited on Amazon's Mechanical Turk service (http://www.mturk.com), an online marketplace where users can browse, sign up for, and complete small paid jobs posted by other users through their internet browser. The number of participants in each experiment was

not predetermined; instead a set number of trials were made available for participants to complete at their leisure (see Procedure below). In total, 104 users were recruited and participated in Experiment 1A, earning on average $0.025 per completed trial, equivalent to an effective hourly rate of $17.17. All recruitment and study procedures in all experiments presented here were conducted in accordance with the ethical policies set forth by the University of Washington's Human Subjects Division, and those in the Declaration of Helsinki.

Since workers on Mechanical Turk are by design completely anonymous to the job posters, demographic information such as participant age and sex is not known for the present samples. However, our experiments were only accessible to workers who, at the time of participation, were U.S.-based and whose quality of work was rated at 95% or higher by other job posters. Generally, research on the demographics of Mechanical Turk workers shows that those based in the U.S. are fairly representative of the general population, resembling them more closely in major demographics than do the university undergraduate samples traditionally used in psychology experiments (Buhrmester, Kwang, & Gosling, 2011; Paolacci, Chandler, & Ipeirotis, 2010).

4.2.1.2 Stimuli & procedure

We restricted our stimulus space to all 32 possible outcomes from flipping a coin five times in a row. On each trial, participants were presented with two five-element sequences defined by a string of Hs and Ts. Accompanying the two sequences were the instructions "Both of the below sequences are possible results from flipping a coin 5 times in a row. Please choose the one that looks more 'random' to you." Participants indicated their answer by clicking on one of two onscreen buttons and then submitting the trial as complete. Participants were free to respond as quickly or slowly as they wished. The experiment comprised 9,920 trials, with the number of trials chosen in order to obtain 20 responses for each of the 496 unique sequence pairs. Since participants were free to complete as many or as few trials as they wished, the number completed by each participant ranged from 1 to 550, with a mean of 95 trials completed per participant. In total, 36 participants completed 50 or more trials, while 51 participants completed 20 or more trials.

4.2.2 *Experiment 1B: Generation task*

In this experiment, a separate group of participants were asked to generate their own 5-element coin flip sequences, to allow direct comparison with the results of the perception task.

4.2.2.1 Participants

Participants were again recruited on Amazon's Mechanical Turk service. In total, 900 participants signed up and earned $0.03 per trial for an effective hourly rate of $9.61.

4.2.2.2 Stimuli & procedure

On each trial, participants were presented with the instructions "Please use the buttons below to make up a sequence of coin flips. Remember, make the sequence look like it came from flipping a real coin five times in a row." For each position in the initially empty five-element sequence, one button was available that would create an H in that position and another was available that would create a T in that position. Each selection within a trial required a distinct mouse movement and click, a method of responding chosen to avoid repetitions of the same selection (e.g., HHHHH) being considerably easier than alternating selections, a concern documented in previous generation studies (Bar-Hillel & Wagenaar, 1991; Wagenaar, 1972). Once a sequence was created, participants clicked another button to submit the completed trial. As in Experiment 1A, participants were free to respond as quickly or slowly as they wished.

The experiment comprised 9,600 trials, with the number of trials chosen to create an expected value of 300 generations for each of the 32 possible sequences. Since participants were free to complete as many trials as they wished (up to a maximum of 40), the number completed by each participant ranged from 1 to 40, with a mean of 11 trials completed per participant. In total, 269 participants contributed at least 20 trials, and 405 participants contributed at least 10 trials.

4.2.3 *Results & Discussion*

4.2.3.1 Experiment 1A

In order to obtain an estimate of perceived randomness for each of the 32 coin flip sequences from participants' forced-choice data, we employed (with permission from the author) Colley's Bias-Free Matrix Rankings method (Colley, 2005). This algorithm, which is famously used to

help determine which two American college football teams will play in the yearly Bowl Championship Series National Championship game as well as several other bowl games, takes in win-loss data for each team and outputs continuous scale ratings describing their relative strengths. Since the sequences in our experiment can be thought of as "teams" that win or lose the randomness judgment against other sequences, we can apply this method to obtain "perceived randomness" ratings (in arbitrary units) for each sequence based on participant responses. The defining feature of the Colley algorithm is that the effect of each win and loss on a sequence's rating is scaled by the win-loss record of the sequence it faced. So choosing the sequence HHTTT as more random-looking than TTHTH is more meaningful than choosing HHTTT over HHHHH, which was very rarely chosen as more random-looking.

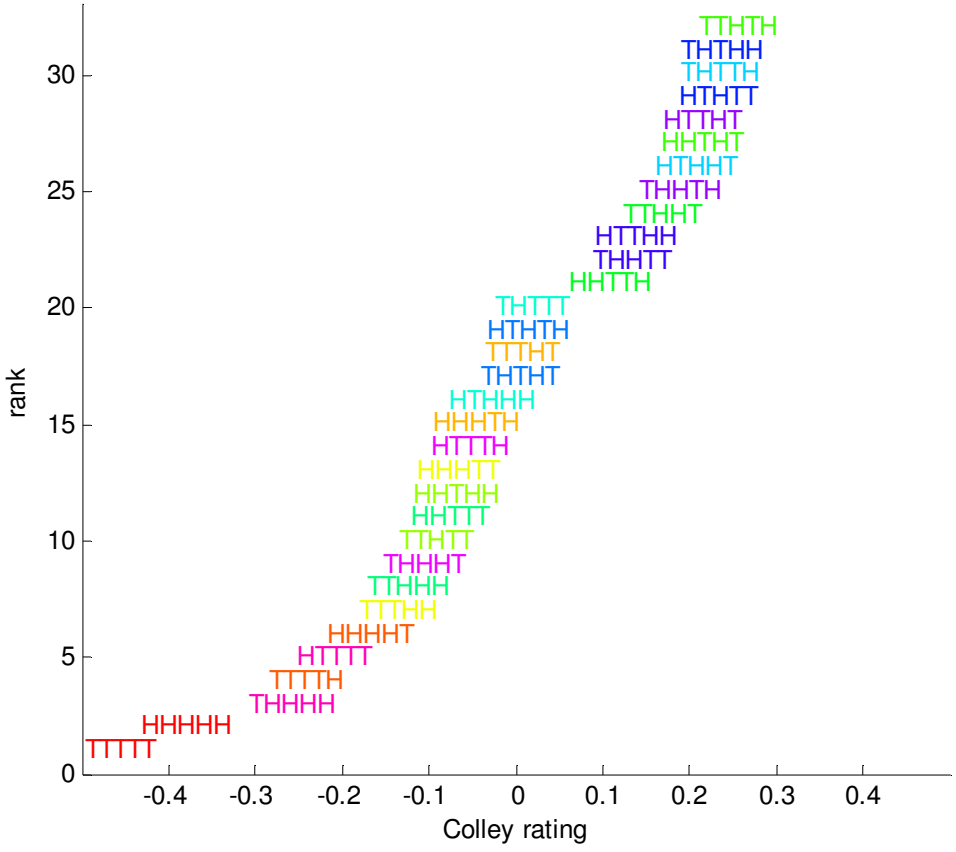

Figure 4-1. Results from Experiment 1A's randomness perception task.

All 32 sequences are plotted in order of their computed Colley rating in Figure 4-1. Sequences are paired with their inverses (e.g., HHTTT and TTHHH) in the figure by color. Notably, some sequences were consistently rated as more random-looking than those they faced, and some were consistently rated as less random-looking. This is consistent with previous studies demonstrating that participants readily rate some sequences of events like these as more or less random than others, despite the mathematical fact that no one sequence is any more likely to be generated by a fair coin than any other sequence (Falk, 1981; Falk & Konold, 1997). What is surprising is the degree of systematicity in the Colley ratings. For example, the eight sequences rated as most random-looking all share nearly identical forms: they all contain exactly three of one result and two of the other and they all contain only one streak, which is always of length two. Furthermore, no streak of length three or more is seen until close to the average Colley rating. Finally, severe heads/tails imbalances and streaks of four or more were assigned low Colley ratings, being consistently identified as less random-looking by participants. Together, this is evidence that even though no specific sequence is any more likely to be produced by a truly random generating process than any other sequence, participants responded meaningfully when asked to rate how random-looking specific sequences are. In fact, they seemed to do so with reference to a clear and consistent internal concept of what is and is not random-looking.

Since the Colley method does not naturally produce a measure of the ratings' reliability, we estimated it by re-calculating Colley ratings based on the subset of trials where the 16 sequences starting with heads were matched up against each other, comprising 25% of all trials. We also re-computed Colley ratings for a separate 25% of trials where the 16 sequences starting with tails were matched up against each other. We found that Colley ratings of sequences and their inverses (e.g., HHTTT and TTHHH) obtained in this way correlated very well ($r = 0.97$, $p < 0.001$) despite coming from totally independent data sets. More informally, we can assess reliability of the ratings simply by visually comparing Colley ratings of sequences against those of their inverses in Figure 1. Visual inspection shows that sequences are rated similarly to their inverses, again suggesting that the ratings are reliable estimates of perceived randomness.

4.2.3.2 Experiment 1B

The frequency with which participants generated sequences across the 9,600 experimental trials is shown in Figure 4-2. Subjects generated some sequences more than others. The most

frequently-generated sequences were generated over four times more than the least-generated sequences. This is despite the fact that a true random generator is expected to produce each of these sequences with equal frequency. Regardless of this seeming misunderstanding of how randomness behaves, it is clear that there is a high degree of consistency in the data, with sequences and their mirror images (e.g., HHTTH and TTHHT) being generated with highly correlated frequencies (r = 0.98, p < 0.001). Additionally, even though participants under-generated them compared to an actual fair coin, participants did seem to understand that even "non-random" sequences such as HHHHH and TTTTT are produced some of the time.

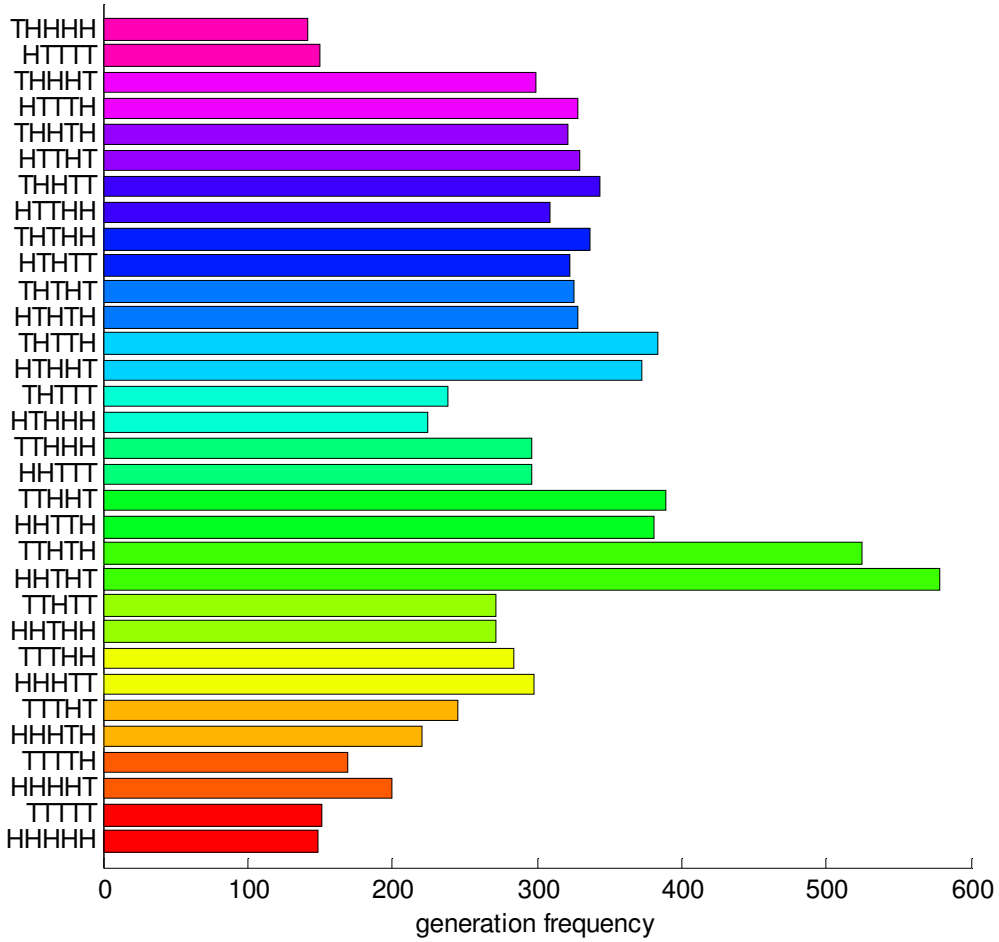

Figure 4-2. Results from Experiment 1B's randomness generation task.

Interestingly, sequences that were generated the most seemed to obey some of the same rules as sequences with high Colley ratings in Experiment 1A. For example, the two most-generated sequences TTHTH and HHTHT had the highest and fifth-highest Colley ratings. On the other hand, sequences with low Colley ratings were among the least-generated sequences. This is notable given that generation frequencies and Colley ratings were obtained using independent participant groups and very different tasks. In order to test for a consistent concept of randomness governing responses to both the perception and generation tasks, we plotted all 32 sequences' frequency of generation against their Colley ratings in Figure 4-3. We found a very high degree of correspondence between the two tasks (r = 0.80, p < 0.001), with sequences perceived as more random being generated more frequently. Interestingly, the two outlier sequences TTHTH and HHTHT were generated about 1.5 times more than predicted by their Colley ratings. This close correspondence between results E1A and E1B suggests that an underlying concept of randomness determines not only which sequences appear to look more random but also which sequences are more often randomly generated.

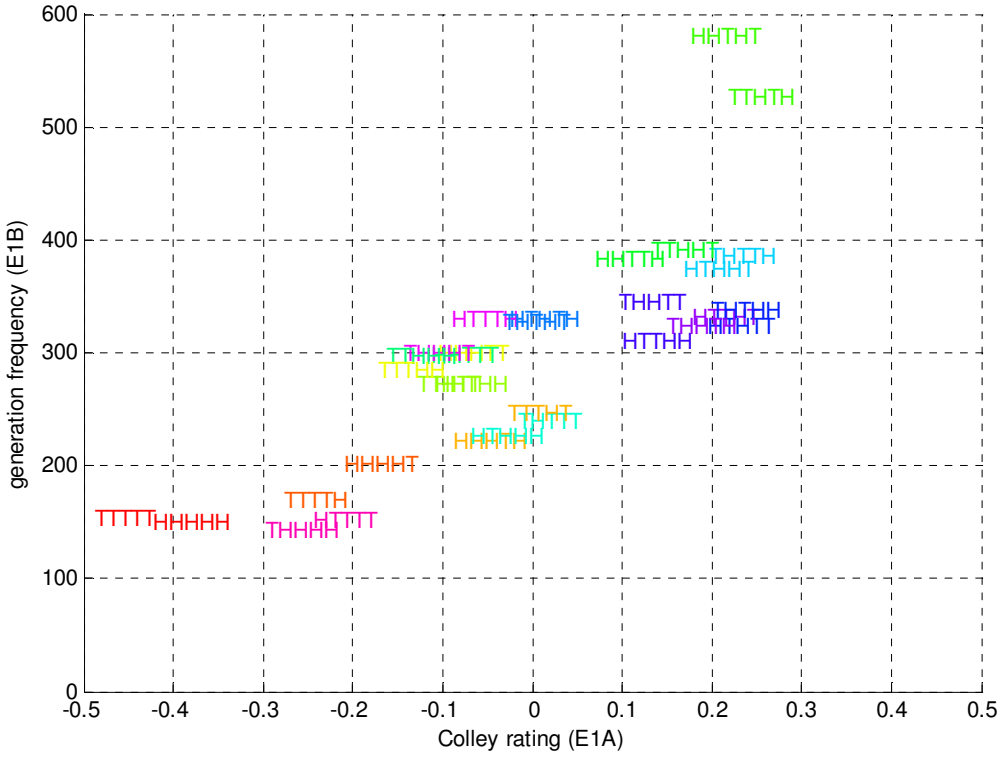

Figure 4-3. Sequences from Experiment 1A and Experiment 1B plotted together.

4.2.3.3 Predicting perceptions of randomness

Given the high degree of systematicity participants displayed in judging the perceived randomness of sequences, we attempted to identify what properties of the sequences themselves led to these judgments. Specifically, we tested the influence of the sequences' *entropy*, *probability of alternation*, and *symmetry* on their Colley ratings from Experiment 1A. Entropy, also known as Shannon entropy, quantifies the balance of Hs and Ts in the sequences. Entropy is defined as:

$$-P_H \log(P_H) - P_T \log(P_T)$$

Where P_H and P_T are the proportions of heads and tails in the sequence, respectively, and *P log(P)* is set to zero when *P* is zero. Thus, entropy is maximal when an equal number of Hs and Ts are present and decreases to zero if sequence contains either all Hs or all Ts. For a five element sequence, entropy will take on one of three values depending on whether there are zero, one, or two heads (or tails).

We define the probability of alternation (or pA) as the proportion of times a sequence alternates rather than repeats states. For example, HHTHH alternates twice out of the four possible opportunities for alternation and thus has a pA of 0.5.

Finally, symmetry here quantifies how symmetrical a sequence is about its center. Symmetry is defined as:

$$\frac{1}{n_2}\sum_{i=1}^{n_2} S_i S_{n-i+1}$$

Where S_i is 1 if the i_{th} member of the sequence is H and -1 if it is T. n is the length of the sequence, and n_2 is n/2 rounded down to the nearest integer. Thus, for the sequence HHTTH, the symmetry is [(1)(1) + (1)(-1)]/2 = 0. A symmetric sequence has symmetry of 1. For sequences of length 5, symmetry takes on the values of -1, 0 or 1. The three properties, entropy, probability of alternation and symmetry, were chosen based on previous work suggesting that they may contribute to the conception of randomness for binary sequences (Falk & Konold, 1997; Oskarsson, et al., 2009; Wagenaar, 1972).

Zero-order correlations between these properties and sequences' Colley ratings are shown in Figure 4-4. All three properties predicted a significant amount of variance in the perceived randomness of sequences. Specifically, Colley ratings decreased with increasing symmetry (r = -0.38, p = 0.03), increased with increasing entropy (r = 0.74, p < 0.001), and increased with increasing numbers of alternations (r = 0.80, p < 0.001). Since previous work has shown that perceived randomness is usually maximal for intermediate pA values (Falk & Konold, 1997; Gilovich, 1985), an additional curvilinear fit was computed for pA predicting Colley ratings. The results were similar (r = 0.88, p < 0.001), with predicted Colley ratings maximal for a pA of about 0.8.

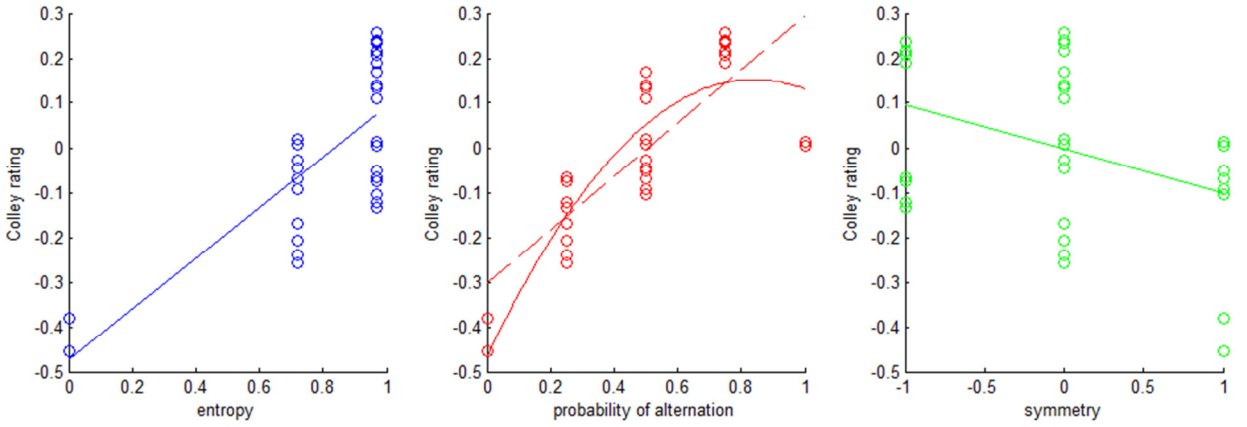

Figure 4-4. Correlating sequence properties and their perceived randomnesses.

Since the predictive power of these three factors individually on perceived randomness of sequences was surprisingly high, we tested whether the perceived randomness of our sequences could be successfully predicted based only on these simple statistical descriptors. To do this, we fit a Thurstonian scaling model (Maloney & Yang, 2003; Thurstone, 1927) directly to our participants' responses from Experiment 1A. The model employs a weighted sum of entropy (E), probability of alternation (A), and a quadratic function of symmetry (S) to place each sequence on an equal-discriminability scale according to Equation 1 below. Based on the correlations reported above, the effect of entropy and symmetry is treated as linear by the model, but the effect of pA is treated as quadratic:

$$R = k_S S + k_E E + k_A (A - c_A)^2$$

Each sequence's location on this scale is its predicted perceived randomness, *R*. The model predicts that for any two sequences, the greater the difference in their *R* values, the more likely a participant will chose the sequence with the greater *R* as appearing more random. Specifically, if the two sequences differ by ΔR, then the probability that the sequence with the higher *R* will be chosen by a participant as more random-looking is determined by a cumulative normal function of ΔR. The model therefore can predict the probability of our observed choice data in Experiment 1A for any set of model parameters. We used a standard optimization routine to fit the model, finding the four model parameters that maximized the probability of observing our data. We tested the quality of fit of the model by comparing the predicted randomness values it produced to the Colley ratings initially obtained for each sequence, with this comparison shown in Figure 4-5. The resulting correlation was very strong, $r = 0.91$, $p < 0.001$. Higher predicted randomness by the model was coincident with high Colley ratings of perceived randomness.

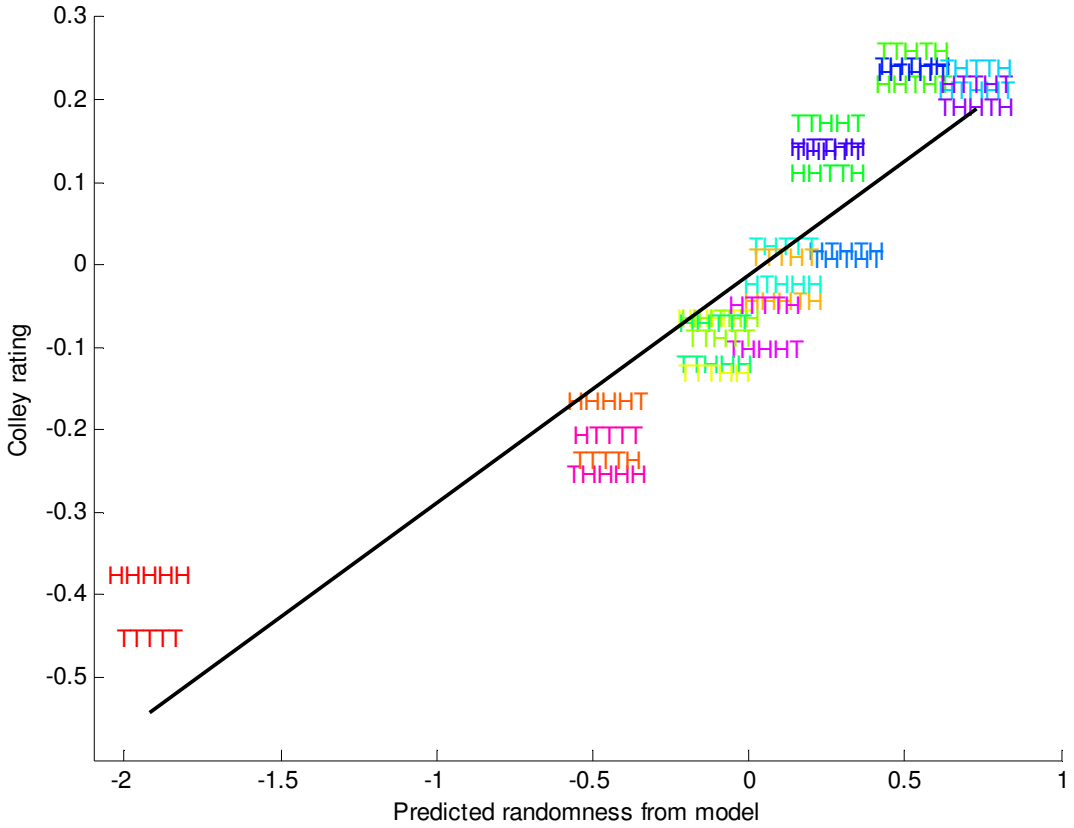

Figure 4-5. Predicting perceived randomness of Experiment 1 sequences.

It should be noted that Colley ratings were determined by an algorithm completely separate from the model, but they both rely on the same win and loss data to produce their results. So it is possible that, even given a hypothetical data set with arbitrary responses, the Colley ratings and model-predicted randomness produced from them would correlate well due entirely to them using the same underlying data. In order to avoid this issue, we split the Experiment 1A data into two random halves, obtained Colley ratings from one half, obtained predicted randomness scores from the other half, then measured the correlation between the two. After conducting this procedure 1,000 times, the resulting correlations between predicted randomness scores and Colley ratings were consistently high (mean $r = 0.91$, range = [0.87, 0.93]) and statistically significant in all cases (all $p < 0.001$). These validation procedures indicated that the model's quality of fit was very good, and that the perceived randomness of the five-element coin flip sequences could be predicted extremely well based only on a few simple statistical descriptors of the sequences themselves. Crucially, our model does this successfully without knowing anything about the participants, their cognitive state, their education level, the experimental context, or even the specific type of stimuli used.

4.3 Experiments 2 & 3: Predicting longer and different types of sequences

Since the predicted randomness model takes in only general statistical descriptors of the sequences and not the sequences themselves, the model as fit to five-element coin flip sequences might be generalizable to other types of sequences. In other words, is the fitted model specific only to short symbolic coin flip sequences or is it generalizable? In Experiment 2, we tested the generality of the model by applying it to judgments of basic visual patterns. In Experiment 3, we tested whether the model can predict the perceived randomness of longer sequences.

4.3.1 *Method*

The methods in Experiment 2 and 3 were the same as in Experiment 1A, except where noted below.

4.3.1.1 Participants

Participants were again recruited on Amazon's Mechanical Turk service, with participation available only to U.S.-based workers with high quality-of-work ratings. In total, 61 users were recruited and participated in Experiment 2, earning $0.02 per completed trial, equivalent to an effective hourly rate of $10.37. A separately-recruited sample of 66 participants contributed to Experiment 3, earning $0.02 per trial for an effective hourly rate of $10.01.

4.3.1.2 Stimuli & procedure

On each trial in Experiment 2 and 3, participants were presented with two images each consisting of a sequence of vertically-arranged Gabor patterns, as seen in Figure 4-6. Each Gabor in a given sequence was oriented either vertically or horizontally. Accompanying the two sequences on each trial were the instructions "Both of the below sequences are possible arrangements of five horizontal and vertical patches. Please choose the sequence that looks the most 'random' to you." In Experiment 2, each sequence contained five Gabor patterns, and so all 32 possible sequences of horizontal and vertical Gabors were used to create 496 unique sequence pairs. In Experiment 3 each sequence contained eight Gabor patterns, but since using all of the 256 possible sequences would result in a prohibitively large number of sequence pairs, a subset of 32 eight-element sequences was used. The specific sequences used were determined by computing predicted randomness scores for all 256 possible sequences using the model fitted in Experiment 1A, then sampling evenly across the range of R values. This was done in order to achieve a broad spread of expected perceived randomness values in the stimulus set.

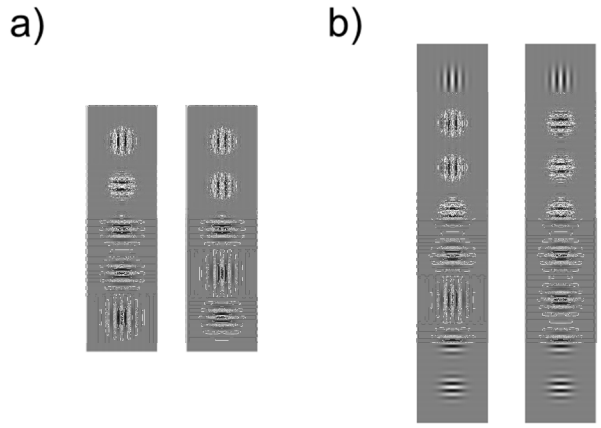

Figure 4-6. Sample stimuli used in Experiments 2 and 3.

Experiments 2 and 3 each consisted of 4,960 trials, with the number chosen in order to obtain 10 responses for each of the 496 unique sequence pairs in each experiment. Since participants were free to complete as many or as few trials as they wished, the number completed by each participant in Experiment 2 ranged from 1 to 324, with a mean of 81 trials completed per participant. In total, 25 participants contributed 50 or more trials and 34 participants contributed 20 or more trials. In Experiment 3, the number of trials completed by each participant ranged from 1 to 834, with 23 participants contributing 50 or more trials and 34 participants completing 20 or more trials.

4.3.2 *Results & Discussion*

The goal of Experiments 2 and 3 was to test whether the model of predicted randomness generalizes from sequences of Hs and Ts to other types of binary sequences. To do this, we allowed the model as fitted to coin flip sequences in Experiment 1A to predict the perceived randomness of the sequences used in Experiments 2 and 3 from their entropy, probability of alternation, and symmetry. Importantly, we did not re-fit the model, but rather used the model parameters obtained in Experiment 1A. Then, to test the validity of the predicted randomness values, we compared them to Colley ratings obtained directly from participants' responses in Experiments 2 and 3. The results of these comparisons are shown in Figure 4-7.

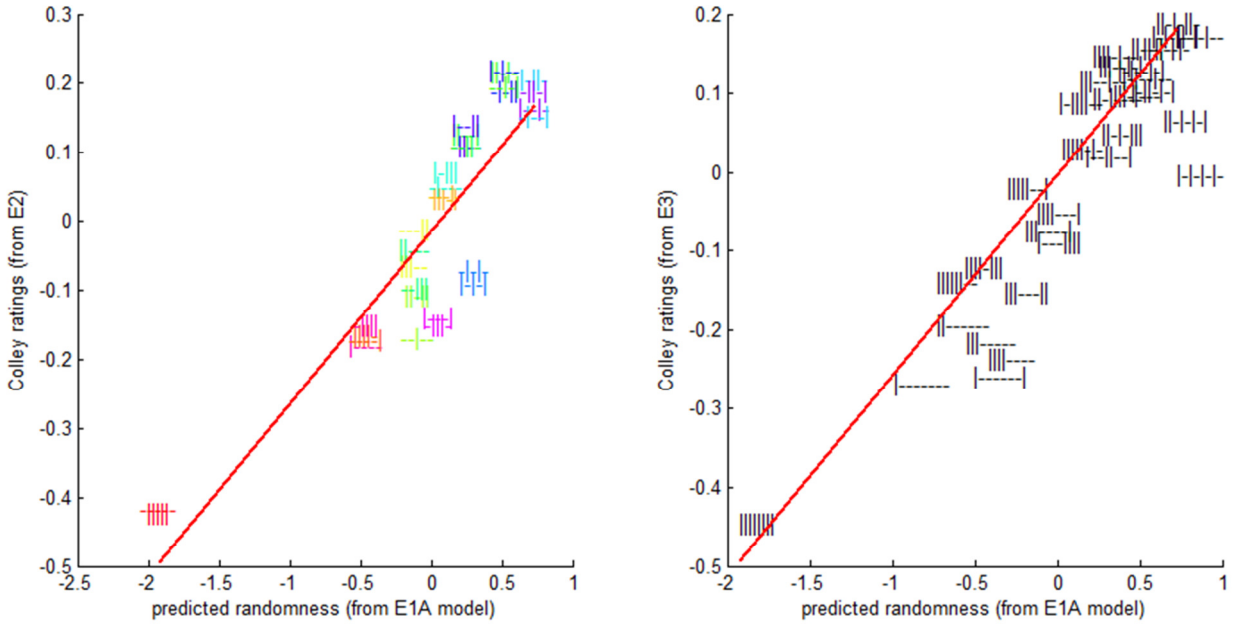

Figure 4-7. Predicting perceived randomness of Experiment 2 and 3 sequences.

As can be seen, high correlations were found between predicted randomness values and Colley ratings in both Experiment 2, $r = 0.89$, $p < 0.001$, and Experiment 3, $r = 0.91$, $p < 0.001$. Despite not being fitted to the types of sequences used here, the model predicted the perceived randomness of our five- and eight-element Gabor sequences very well.

4.4 General Discussion

We are pattern detectors. Much of our perceptual and cognitive machinery is devoted to detecting patterns in noise, from the earliest stages of sensory processing to the highest levels of decision making. All processes of pattern detection necessarily require some internal model of randomness from which to discriminate from. Here, we've shown that people appear to have a consistent, coherent internal conception of what random outcomes looks like. We found that this concept is expressed very similarly whether people are asked to judge a series of outcomes or produce them. Furthermore, we found that this internal model is highly predictable: participant judgments of randomness were predicted very well based only on a few simple statistical properties of the stimulus being judged. Finally, we found that the model, without modification, generalized well to both longer and qualitatively different types of sequences.

4.4.1 *Potential applications*

Accurately predicting how random or non-random specific sequences will appear to us has significance for behaviors as disparate as gambling choices, investing, belief in the supernatural, sports strategy, and high-impact data interpretation. For example, public officials must often make important policy decisions based on a series of discrete outcomes, such as quarterly or yearly changes in crime rate. Seeing four out of five months of downticks in violent crime rates may convince a city council (or the public) that an expensive crime crackdown has "worked," when in fact the downticks are simply the result of random fluctuations and the policy itself is not particularly effective. Misapprehensions about randomness in a case like this can cause significant waste of public resources and real harm to individuals and the community. Similar situations can be easily imagined for data on global temperature, disease prevalence rates, curriculum evaluation, and more.

Another possible application of models of the type developed here concerns cryptography, or hiding a pattern or signal in a context. The standard approach would involve embedding the signal in truly random noise. But as we know, true randomness does not appear random to humans, and could arouse suspicion in a human observer. Since our model can predict the perceived randomness of any binary sequence, it provides a way to determine what context would result in the most random-looking sequence for any set of givens. As can be seen in Figure 4-8, a pre-specified pattern can be made to "disappear" into the context using this technique. Rather than signals embedded in noise, signals embedded in subjective randomness might be uniquely effective in communicating hidden information past human observers. This principle and model could in theory be extended to two-dimensions in order to mask regularities or patterns in geographical spaces. Such a strategy could be used in a military setting to arrange a variety of random-seeming checkpoints to intercept intelligence or materials at a specific location, or to mask operations on specific targets (e.g., nuclear centrifuges or extremist cells) without it appearing that the nature of the target was known ahead of time.

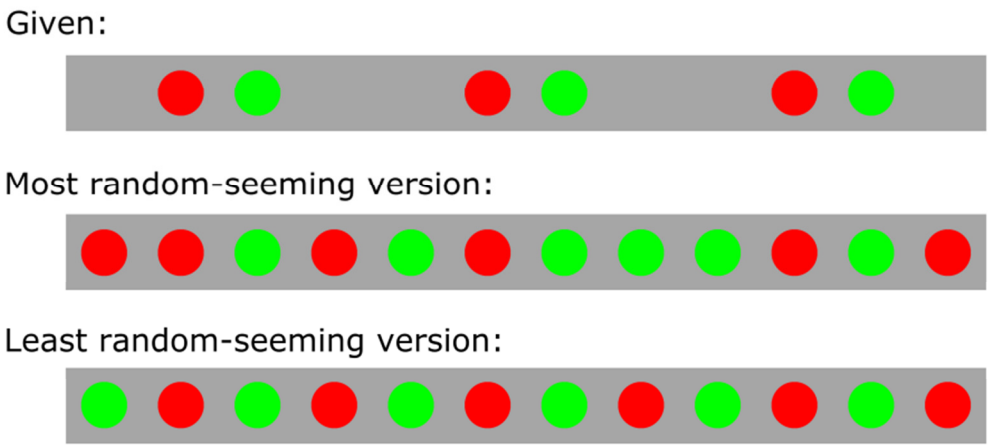

Figure 4-8. Use of the model to embed signals in random-seeming contexts.

4.4.2 *Limitations*

Currently, the primary limitation of our model is that it is fit to and predicts only binary, one-dimensional, visual sequences. We have shown that the model predicts randomness judgments very well for five-element coin flip sequences, five-element Gabor patterns, and eight-element Gabor patterns, but it would be useful to be able to predict randomness judgments for longer sequences, sequences with more than two alternatives, two-dimensional grid-type stimuli, or auditory sequences, for example. The model as fitted to five-element sequences predicted eight-element sequences successfully, but in longer sequences the influence of each characteristic changes, as does the correlation between entropy and pA specifically. Very little is known about randomness perception in sequences with more than two alternatives (but see Wagenaar, 1972), but each sequence characteristic used in our model can be calculated for sequences with three, four, or more alternatives. Thus there is no *a priori* reason why the model could not be generalized to those sequences given data to fit it with. Some researchers have studied randomness perception in two-dimensional grids (Falk, 1975; Zhao, Hahn, & Osherson, 2014), but no work tightly links characteristics of those grids to their judged randomness. Furthermore, pA is difficult to conceptualize for this type of stimuli, so it is not clear how predictable perceived randomness is in this case. Finally, while our auditory systems have sophisticated pattern detection abilities, it is entirely unclear how randomness is detected or judged in this domain, but our model could theoretically be applied here as well.

Another potential limitation of the present work is how participants interpreted the instructions in our task. We intentionally left the definition of randomness up to the participants since it was exactly their conception of randomness that we were trying to quantify, but we have no direct knowledge of their interpretations. In a sense, the question posed to participants in the perception tasks is meaningless, since randomness is not an attribute of specific sequences, but of their generating processes. Furthermore, no single sequence used here is more likely to be generated by a fair coin than any other sequence, assuming five independent flips. Yet despite this, participants responded very systematically, resulting in strong and sensible model fits and a close correspondence between perception and generation tasks. Furthermore, our results fit generally with previous findings suggesting that people prefer over-alternating and well-balanced sequences when thinking of randomness. This suggests that our instructions elicited their internal concept of randomness well. However, it remains possible that the explicit concept of randomness invoked by our instructions does not match perfectly with that used naturalistically by people to make decisions about, for example, gambling or stock market outcomes. This would be an important connection to make, but it remains for future work.

4.4.3 *Are we bad at understanding randomness?*

The present work is concerned primarily with identifying a consistent internal idea of randomness and predicting its outputs. But the idea of randomness that we characterized here seems to be flawed, too. Participants showed prevalent, consistent biases in perceiving and imitating randomness. Does this mean that humans are bad at thinking about randomness? That has historically been the prevailing view (Bar-Hillel & Wagenaar, 1991; Oskarsson, et al., 2009; Wagenaar, 1972), but more recent work suggests this may not be a fair characterization. For example, researchers have recently found that after beginning to flip a fair coin, an observer will, on average, have to wait longer before encountering some specific strings of outcomes than they will for others (Hahn & Warren, 2009; Sun & Wang, 2010). Interestingly, they found that an observer would need to wait longer in order to see sets of streaky outcomes like HHHH than they would for non-streaky sets of outcomes, like HTTH. Other work argues that even random-looking sequences constitute only weak evidence for a random generating process, and that we are right to discount that possibility (Williams & Griffiths, 2013). Finally, Miller & Sanjurjo (2015) show that our concept of randomness is a natural consequence of how we experience

sequences of outcomes. This more positive view of randomness perception is more consistent with findings elsewhere that suggest humans are in fact very well-tuned to the statistics of our environment (Fiser & Aslin, 2001; Saffran, et al., 1996; Zhao, Al-Aidroos, & Turk-Browne, 2013).

Regardless of whether our notion of randomness is statistically correct, it may indeed be adaptive. Identifying regularities and laws in our environment is the only way we can make better decisions or optimize our behavior. In this sense, concluding that some stimulus or set of outcomes was randomly produced is giving up on being able to understand and take advantage of it. And given that the natural world does contain structure and obey laws, even fair coins obeying the laws of physics, it is understandable that we are reluctant to conclude that some specific part of it is unknowable. Seen in this light, our aversion to labeling randomness as such may be viewed not as stupidity or deficiency, but perseverance in our goal of understanding the cosmos.

Chapter 5. FINAL REMARKS

Here I have presented and discussed scientific investigations into two specific aspects of how we deal with our noisy world. In a sense, both investigations have involved documenting deficiencies in our ability to deal with variability in our environment. Chapters 2 and 3 focused on the discovery and causes of seemingly suboptimal biases in how we compute the mean of a series of visual objects. Chapter 4 demonstrated how people readily label (and generate) sequences of outcomes as random and non-random despite the fact that each sequence was equally likely to be generated by a truly random process. The model described in that chapter even went so far as to quantify and codify exactly what goes into these seemingly nonsensical decisions.

However, this should not be taken to mean that we are bad in general at functioning in a noisy environment. It is true that in specific situations, such as in gambling or other evaluations of chance, we often have trouble making effective decisions. But in the vast majority of other domains we are quite exquisite pattern-detecting machines. For example, we have no trouble whatsoever in very accurately interpreting nearly any visual scene we are likely to come across despite the huge amount of ambiguity and informational chaos that exists in the patterns of light they produce on our retinas. From this array of light we easily detect and recognize any of a vast number of objects in our surrounds despite nearly infinite variation in lighting, perspective, and position. Similarly, from highly variable and often degraded sound patterns we can recognize hundreds of thousands of words as well as the language or accent of the speaker. This is in addition to understanding the language's literal and inferred meanings, as well as the speaker's age, sex, and even mood. In comparison, tens of thousands of highly accomplished scientists and engineers attempting to develop the same abilities in automated systems have, over half a century, only been able to obtain a fraction of human performance. Such is the extent of our ability to take noisy input and extract fruitful meaning.

Thus, the purpose of eliciting, exploring, and documenting seeming deficiencies or failures of our perception and cognition is not to lament our abilities, but rather to understand them. Study of a system that works perfectly tells us little, but study of a system stressed or pushed to the point of failure gives us the knowledge needed to improve it, or to apply its principles

productively elsewhere. This is the endeavor of perceptual and cognitive science, and with it, we are improving ourselves all the time.

BIBLIOGRAPHY

Albrecht, A. R., & Scholl, B. J. (2010). Perceptually averaging in a continuous visual world: extracting statistical summary representations over time. *Psychological science, 21*(4), 560-567.

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

Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychol Sci, 15*(2), 106-111.

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

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

Ayton, P., & Fischer, I. (2004). The hot hand fallacy and the gambler's fallacy: two faces of subjective randomness? *Mem Cognit, 32*(8), 1369-1378.

Bahcall, D. O., & Kowler, E. (1999). Illusory shifts in visual direction accompany adaptation of saccadic eye movements. *Nature, 400*(6747), 864-866.

Bar-Hillel, M., & Wagenaar, W. A. (1991). The perception of randomness. *Advances in applied mathematics, 12*(4), 428-454.

Barberis, N. T., R. (2003). A survey of behavioral finance. In G. M. H. Constantinides, M.Stulz, R. M. (Ed.), *Handbook of the economics of finance* (1 ed., Vol. 1, pp. 1053-1128): Elsevier.

Binda, P. M., M. C.Boynton, G. M.Murray, S. O. (2011). Spatial attention affects perceived stimulus position. *Journal of Vision, 11*(11), 229-229.

Blackmore, S. T., T. (1985). Belief in the paranormal: Probability judgements, illusory control, and the 'chance baseline shift'. *British journal of psychology, 76*(4), 459-468.

Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: ensemble statistics bias memory for individual items. *Psychological science, 22*(3), 384-392.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spat Vis, 10*(4), 433-436.

Brugger, P. L., T.Regard, M. (1990). A 'sheep-goat effect' in repetition avoidance: Extra-sensory perception as an effect of subjective probability? *British journal of psychology, 81*(4), 455-468.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's mechanical turk: A new source of inexpensive, yet high-quality, data? *Perspectives on psychological science, 6*(1), 3-5.

Carpenter, R. H. S. (1977). *Movements of the eyes*. London: Pion.

Carrasco, M., Williams, P. E., & Yeshurun, Y. (2002). Covert attention increases spatial resolution with or without masks: support for signal enhancement. *J Vis, 2*(6), 467-479.

Cheadle, S., Wyart, V., Tsetsos, K., Myers, N., de Gardelle, V., Herce Castañón, S., et al. (2014). Adaptive gain control during human perceptual choice. *Neuron, 81*(6), 1429-1441.

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

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

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

Colley, W. N. (2005). *The Colley Matrix System for Ranking College Football.* Paper presented at the Society for Industrial and Applied Mathematics Conference on Applications of Applied Dynamical Systems, Snowbird, UT.

Connor, C. E. G., A. L.Van Essen, D. (1994). Modulation of receptive field profiles in area V4 by shifts in focal attention. *Investigative ophthalmology & visual science, 35*(4), 2147.

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

Courtney, S. M., Ungerleider, L. G., Keil, K., & Haxby, J. V. (1996). Object and spatial visual working memory activate separate neural systems in human cortex. *Cereb Cortex, 6*(1), 39-49.

Courtney, S. M., Ungerleider, L. G., Keil, K., & Haxby, J. V. (1997). Transient and sustained activity in a distributed neural system for human working memory. *Nature, 386*(6625), 608-611.

Dakin, S. C. (2001). Information limit on the spatial integration of local orientation signals. *Journal of the Optical Society of America. A, Optics, image science, and vision, 18*(5), 1016-1026.

de Fockert, J., & Wolfenstein, C. (2009). Rapid extraction of mean identity from sets of faces. *Quarterly journal of experimental psychology (2006), 62*(9), 1716-1722.

de Gardelle, V., & Summerfield, C. (2011). Robust averaging during perceptual judgment. *Proceedings of the National Academy of Sciences of the United States of America, 108*(32), 13341-13346.

Drugowitsch, J., Moreno-Bote, R., Churchland, A. K., Shadlen, M. N., & Pouget, A. (2012). The cost of accumulating evidence in perceptual decision making. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 32*(11), 3612-3628.

Duhamel, J. R., Colby, C. L., & Goldberg, M. E. (1992). The updating of the representation of visual space in parietal cortex by intended eye movements. *Science, 255*(5040), 90-92.

Egeth, H. E., & Yantis, S. (1997). Visual attention: control, representation, and time course. *Annu Rev Psychol, 48*, 269-297.

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

Falk, R. (1975). Perception of randomness. Hebrew University.

Falk, R. (1981). *The perception of randomness.* Paper presented at the Fifth international conference for the psychology of mathematics education.

Falk, R., & Konold, C. (1997). Making sense of randomness: Implicit encoding as a basis for judgment. *Psychological Review*.

Farrand, P., & Jones, D. (1996). Direction of report in spatial and verbal serial short-term memory. *Quarterly Journal of Experimental Psychology, 49A*(1), 140-158.

Farrand, P., Parmentier, F. B. R., & Jones, D. M. (2001). Temporal-spatial memory: Retrieval of spatial information does not reduce recency. *Acta Psychologica, 106*, 285-301.

Fischer, B., & Boch, R. (1983). Saccadic eye movements after extremely short reaction times in the monkey. *Brain Res, 260*(1), 21-26.

Fischer, J., & Whitney, D. (2009). Attention narrows position tuning of population responses in V1. *Curr Biol, 19*(16), 1356-1361.

Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Nat Neurosci, 17*(5), 738-743.

Fiser, J., & Aslin, R. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science, 12*, 499-504.

Gilovich, T. V., R.Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive psychology, 17*(3), 295-314.

Goldman-Rakic, P. S. (1996). Regional and cellular fractionation of working memory. *Proc Natl Acad Sci U S A, 93*(24), 13473-13480.

Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition, 70*(2), 109-135.

Greenwood, J. A., Bex, P. J., & Dakin, S. C. (2009). Positional averaging explains crowding with letter-like stimuli. *Proceedings of the National Academy of Sciences of the United States of America, 106*(31), 13130-13135.

Guérard, K., & Tremblay, S. (2008). Revisiting evidence for modularity and functional equivalence across verbal and spatial domains in memory. *J Exp Psychol Learn Mem Cogn, 34*(3), 556-569.

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

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current biology : CB, 17*(17), 3.

Haberman, J., & Whitney, D. (2009). Seeing the mean: ensemble coding for sets of faces. *Journal of experimental psychology. Human perception and performance, 35*(3), 718-734.

Hahn, U., & Warren, P. A. (2009). Perceptions of randomness: why three heads are better than four. *Psychological review, 116*(2), 454-461.

Hay, D. C., Smyth, M. M., Hitch, G. J., & Horton, N. J. (2007). Serial position effects in short-term visual memory: a SIMPLE explanation? *Mem Cognit, 35*(1), 176-190.

He, S., Cavanagh, P., & Intriligator, J. (1996). Attentional resolution and the locus of visual awareness. *Nature, 383*(6598), 334-337.

Henderson, J. M., & Macquistan, A. D. (1993). The spatial distribution of attention following an exogenous cue. *Percept Psychophys, 53*(2), 221-230.

Honda, H. (1993). Saccade-contingent displacement of the apparent position of visual stimuli flashed on a dimly illuminated structured background. *Vision Res, 33*(5-6), 709-716.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *J Physiol, 195*(1), 215-243.

Hyun, J. S., & Luck, S. J. (2007). Visual working memory as the substrate for mental rotation. *Psychon Bull Rev, 14*(1), 154-158.

Jakes, S., & Hemsley, D. R. (1986). Individual differences in reaction to brief exposure to unpatterned visual stimulation. *Personality and individual differences, 7*(1), 121-123.

Johnson, J. T., G. J. (2005). Blowing bubbles: Heuristics and biases in the run-up of stock prices. *Journal of the academy of marketing science, 33*(4), 486-503.

Johnson, S. C. (2003). Detecting agents. *Philos Trans R Soc Lond B Biol Sci, 358*(1431), 549-559.

Jones, D., Farrand, P., Stuart, G., & Morris, N. (1995). Functional equivalence of verbal and spatial information in serial short-term memory. *J Exp Psychol Learn Mem Cogn, 21*(4), 1008-1018.

Juni, M. Z., Gureckis, T. M., & Maloney, L. T. (2012). Effective integration of serially presented stochastic cues. *Journal of vision, 12*(8).

Kahneman, D., & Tversky, A. (1974). Subjective probability: A judgment of representativeness. *The Concept of Probability in Psychological ….*

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci, 17*(11), 4302-4311.

Kay, A. C., Moscovitch, D. A., & Laurin, K. (2010). Randomness, attributions of arousal, and belief in god. *Psychol Sci, 21*(2), 216-218.

Kiani, R., Hanks, T. D., & Shadlen, M. N. (2008). Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. *The Journal of neuroscience : the official journal of the Society for Neuroscience, 28*(12), 3017-3029.

Ladouceur, R. W., Sylvain, C., Letarte, H., Giroux, I., & Jacques, C. (1998). Cognitive treatment of pathological gamblers. *Behav Res Ther, 36*(12), 1111-1119.

Ladouceur, R. W., & Walker, M. (1996). A cognitive perspective on gambling. *Trends in cognitive and behavioral therapies*, 89-120.

Liberman, A., Fischer, J., & Whitney, D. (2014). Serial dependence in the perception of faces. *Curr Biol, 24*(21), 2569-2574.

Logie, R. H., & Marchetti, C. (1991). Visuo-spatial working memory: Visual, spatial or central executive? *Advances in Psychology, 80*, 105-115.

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

Maloney, L. T., & Yang, J. N. (2003). Maximum likelihood difference scaling. *J Vis, 3*(8), 573-585.

Mateeff, S. (1978). Saccadic eye movements and localization of visual stimuli. *Percept Psychophys, 24*(3), 215-224.

Mateeff, S., & Gourevich, A. (1983). Peripheral vision and perceived visual direction. *Biol Cybern, 49*(2), 111-118.

Matin, L., & Pearce, D. G. (1965). Visual Perception of Direction for Stimuli Flashed During Voluntary Saccadic Eye Movements. *Science, 148*(3676), 1485-1488.

Meltzoff, A. N., Kuhl, P. K., Movellan, J., & Sejnowski, T. J. (2009). Foundations for a new science of learning. *Science, 325*(5938), 284-288.

Michalczuk, R., Bowden-Jones, H., Verdejo-Garcia, A., & Clark, L. (2011). Impulsivity and cognitive distortions in pathological gamblers attending the UK National Problem Gambling Clinic: a preliminary report. *Psychol Med, 41*(12), 2625-2635.

Miller, J. B., & Sanjurjo, A. (2015). Surprised by the Gambler's and Hot Hand Fallacies? A Truth in the Law of Small Numbers.

Morgan, M. J., & Glennerster, A. (1991). Efficiency of locating centres of dot-clusters by human observers. *Vision research, 31*(12), 2075-2083.

Morrone, M. C., Ross, J., & Burr, D. C. (1997). Apparent position of visual targets during real and simulated saccadic eye movements. *J Neurosci, 17*(20), 7941-7953.

Müller, H. J., & Rabbitt, P. M. (1989). Reflexive and voluntary orienting of visual attention: time course of activation and resistance to interruption. *J Exp Psychol Hum Percept Perform, 15*(2), 315-330.

Nakayama, K., & Mackeben, M. (1989). Sustained and transient components of focal visual attention. *Vision Res, 29*(11), 1631-1647.

Nickerson, R. S. (2002). The production and perception of randomness. *Psychol Rev, 109*(2), 330-357.

Nickerson, R. S., & Butler, S. F. (2009). On producing random binary sequences. *Am J Psychol, 122*(2), 141-151.

Osaka, N. (1977). Effect of refraction on perceived locus of a target in the peripheral visual field. *J Psychol, 95*(1st Half), 59-62.

Oskarsson, A. T., Van Boven, L., McClelland, G. H., & Hastie, R. (2009). What's next? Judging sequences of binary events. *Psychol Bull, 135*(2), 262-285.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on amazon mechanical turk. *Judgment and decision making, 5*(5), 411-419.

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

Pasupathy, A., & Connor, C. E. (2002). Population coding of shape in area V4. *Nat Neurosci, 5*(12), 1332-1338.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis, 10*(4), 437-442.

Piazza, E. A., Sweeny, T. D., Wessel, D., Silver, M. A., & Whitney, D. (2013). Humans use summary statistics to perceive auditory sequences. *Psychological science, 24*(8), 1389-1397.

Pittman, T. S. P., N. L. (1980). Deprivation of control and the attribution process. *Journal of personality and social psychology, 39*(3), 377.

Posner, M., & Cohen, Y. (1984). Components of visual orienting. In H. Bouma & D. Bouwhuis (Eds.), *Attention and Performance X* (pp. 531-556). London: Erlbaum.

Pratt, J., & Turk-Browne, N. B. (2003). The attentional repulsion effect in perception and action. *Exp Brain Res, 152*(3), 376-382.

Redington, M. C., N.Finch, S. (1998). *Cognitive science, 22*(4), 425-469.

Robitaille, N., & Harris, I. M. (2011). When more is less: extraction of summary statistics benefits from larger sets. *Journal of vision, 11*(12).

Roney, C. J. T., L. M. (2009). Sympathetic magic and perceptions of randomness: The hot hand versus the gambler's fallacy. *Thinking & reasoning, 15*(2), 197-210.

Ross, J., Morrone, M. C., & Burr, D. C. (1997). Compression of visual space before saccades. *Nature, 386*(6625), 598-601.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926-1928.

Schlag, J., & Schlag-Rey, M. (1995). Illusory localization of stimuli flashed in the dark before saccades. *Vision Res, 35*(16), 2347-2357.

Sheth, B. R., & Shimojo, S. (2001). Compression of space in visual memory. *Vision Res, 41*(3), 329-341.

Shulman, G. L., Wilson, J., & Sheehy, J. B. (1985). Spatial determinants of the distribution of attention. *Percept Psychophys, 37*(1), 59-65.

Smith, E. E., & Jonides, J. (1997). Working memory: A view from neuroimaging. *Cognitive Psychology, 33*(1), 5-42.

Smith, E. E., Jonides, J., Koeppe, R. A., Awh, E., Schumacher, E. H., & Minoshima, S. (1995). Spatial versus object working memory: PET investigations. *Journal of Cognitive Neuroscience, 7*(3), 357-375.

Solomon, J. A. (2010). Visual discrimination of orientation statistics in crowded and uncrowded arrays. *J Vis, 10*(14), 19.

Solomon, J. A., Morgan, M., & Chubb, C. (2011). Efficiencies for the statistics of size discrimination. *Journal of vision, 11*(12), 13.

Spencer, J. (1961). Estimating averages. *Ergonomics*.

Spencer, J. (1963). A further study of estimating averages. *Ergonomics*.

Sun, Y., & Wang, H. (2010). Perception of randomness: On the time of streaks. *Cognitive psychology, 61*(4), 333-342.

Suzuki, S., & Cavanagh, P. (1997). Focused attention distorts visual space: an attentional repulsion effect. *J Exp Psychol Hum Percept Perform, 23*(2), 443-463.

Teghtsoonian, M. (1965). The judgment of size. *American journal of psychology, 78*, 392-402.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological review, 34*(4), 273.

Toneatto, T., Blitz-Miller, T., Calderwood, K., Dragonetti, R., & Tsanos, A. (1997). Cognitive distortions in heavy gambling. *J Gambl Stud, 13*(3), 253-266.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science, 185*(4157), 1124-1131.

Umeno, M. M., & Goldberg, M. E. (1997). Spatial processing in the monkey frontal eye field. I. Predictive visual responses. *J Neurophysiol, 78*(3), 1373-1383.

van der Heijden, A. H., van der Geest, J. N., de Leeuw, F., Krikke, K., & Müsseler, J. (1999). Sources of position-perception error for small isolated targets. *Psychol Res, 62*(1), 20-35.

Wagenaar, W. A. (1972). Generation of random sequences by human subjects: A critical survey of literature. *Psychological Bulletin, 77*(1), 65.

Walker, M. F., Fitzgibbon, E. J., & Goldberg, M. E. (1995). Neurons in the monkey superior colliculus predict the visual result of impending saccadic eye movements. *J Neurophysiol, 73*(5), 1988-2003.

Ward, G., Avons, S. E., & Melling, L. (2005). Serial position curves in short-term memory: functional equivalence across modalities. *Memory, 13*(3-4), 308-317.

Warneken, F., & Tomasello, M. (2006). Altruistic helping in human infants and young chimpanzees. *Science, 311*(5765), 1301-1303.

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

Whitney, D., Haberman, J., & Sweeny, T. D. (2013). From textures to crowds: Multiple levels of summary statistical perception. In J. S. Werner & L. M. Chalupa (Eds.), *The New Visual Neurosciences*. Cambridge, MA: MIT Press.

Whitson, J. A., & Galinsky, A. D. (2008). Lacking control increases illusory pattern perception. *Science, 322*(5898), 115-117.

Williams, J. J., & Griffiths, T. L. (2013). Why are people bad at detecting randomness? A statistical argument. *Journal of experimental psychology: Learning, memory, and cognition, 39*(5), 1473.

Womelsdorf, T., Anton-Erxleben, K., Pieper, F., & Treue, S. (2006). Dynamic shifts of visual receptive fields in cortical area MT by spatial attention. *Nat Neurosci, 9*(9), 1156-1160.

Woodman, G. F., & Luck, S. J. (2004). Visual search is slowed when visuospatial working memory is occupied. *Psychon Bull Rev, 11*(2), 269-274.

Woodman, G. F., Vogel, E. K., & Luck, S. J. (2001). Visual search remains efficient when visual working memory is full. *Psychol Sci, 12*(3), 219-224.

Yeshurun, Y., & Carrasco, M. (1998). Attention improves or impairs visual performance by enhancing spatial resolution. *Nature, 396*(6706), 72-75.

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature, 453*(7192), 233-235.

Zhao, J., Al-Aidroos, N., & Turk-Browne, N. B. (2013). Attention is spontaneously biased toward regularities. *Psychol Sci, 24*(5), 667-677.

Zhao, J., Hahn, U., & Osherson, D. (2014). Perception and identification of random events. *J Exp Psychol Hum Percept Perform, 40*(4), 1358-1371.

APPENDIX A – SUPPLEMENTAL EXPERIMENTS FOR CHAPTER 2

Here are presented five additional behavioral experiments that further measured the temporal weighting profile of mean position and mean size judgments. The primary purpose of the experiments was to test the reproducibility of the findings from Experiment 1 in Chapter 2. Namely, finding primacy for judgments of mean dot location and finding recency for judgments of mean dot size. The secondary purpose of the experiments was to measure the effect of various experimental manipulations on these temporal weighting profiles. The method in each experiment was very similar to that used in Experiment 1 in Chapter 2, and so only deviations from that method are noted here.

SUPPLEMENTAL EXPERIMENT 1: AVERAGING POSITION WITH FEEDBACK

To understand how estimates of mean dot location are affected by the presence of corrective feedback, we presented participants with a series of small white dots and asked them to report the mean position of the group. Some participants were presented with the actual correct answer alongside their answer after their report, while others were not. We then quantified the influence of the temporal position of each dot on the participants' location estimates across many trials in each group of participants.

Method

Participants

Two groups of sixteen undergraduate students at the University of Washington were recruited and compensated in the same manner as in Experiment 1 in Chapter 2. Participants were assigned to their group at the time of the experimental session by referring to a randomly pre-shuffled condition list.

Stimuli

Each trial consisted of a series of ten small white dots that varied in their position. Stimuli were the same as in Experiment 1 in Chapter 2 with the exception that here all dots were the same size, their radii equal to 0.3° of visual angle. Dot locations were sampled as previously described, except that here the standard deviation of the sampling distribution on each trial was 3.0°.

Procedure

Experimental procedure was the same as in Experiment 1 in Chapter 2 with the following exceptions. While the central cross that signaled the start of each trial was the same as it was previously, no verbal precue or postcue was used here since only one stimulus property (mean location) was being reported here. Additionally, a feedback period of 1500 ms was added after the response period of each trial. In one group of participants (the no-feedback group), the feedback period consisted of only a green dot indicating the participant's response. In the other group (the feedback group), the feedback period contained a green dot indicating the participant's response and also a red dot indicating the actual, correct answer. After the feedback period, the next trial began immediately.

Each participant completed 400 trials in ten blocks of 40 trials each. In addition to the normal task instruction, the experimenter also explained the dots contained in the feedback period to participants in both groups, and encouraged both groups to try to answer as accurately as possible on each trial.

Results

Weights describing the relative influence of each temporal position in the sequence of dots (one through ten) on each participant's responses were obtained by fitting the weighted average model described in Chapter 2. Average weights by group are shown in Figure S1.

As in Experiment 1 in Chapter 2, primacy was found for judgments of mean dot location. Earlier dots appeared to contribute more to participant responses than later dots did, a finding supported by a significant main effect of temporal position in a two-way repeated measures ANOVA, $F(9,270) = 20.27$, $p < 0.001$, $\eta_p^2 = .40$. Though there appeared to be no overall different in the

magnitude of weights in the feedback and no-feedback group, the main effect of feedback condition was statistically significant, $F(1,30) = 6.57$, $p = 0.02$, $\eta_p^2 = .18$. However, there was no significant interaction between temporal position and feedback condition, $F(9,270) = 1.54$, $p = 0.14$, $\eta_p^2 = .05$, suggesting that the presence of feedback after each trial did not affect the shape of the temporal weighting profile compared to no feedback.

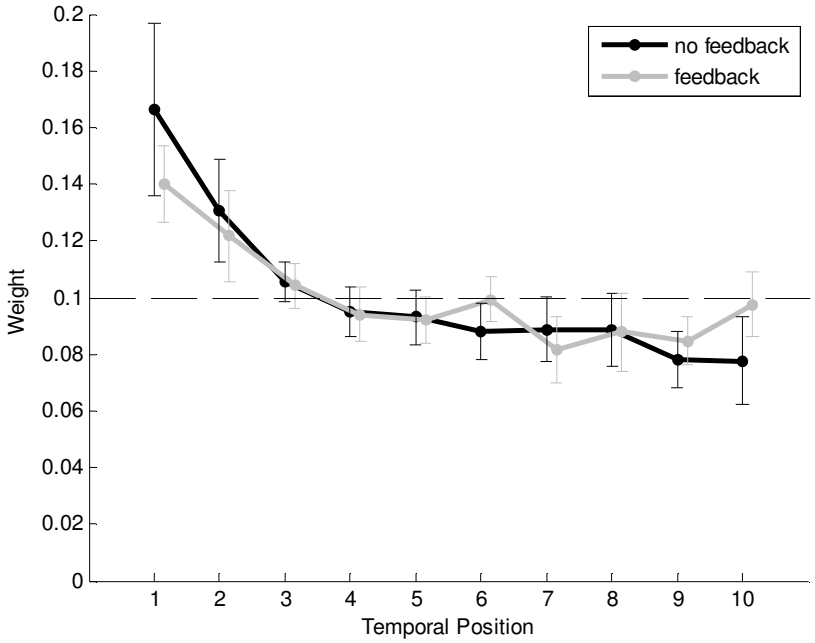

Figure S1. Results from Supplemental Experiment 1.

SUPPLEMENTAL EXPERIMENT 2: AVERAGING POSITION AT DIFFERENT PRESENTATION RATES

Here we manipulated the rate at which dots were presented within each participant, adding in trials with half the presentation rate as that used in Experiment 1 in Chapter 2.

Method

A new group of sixteen participants was recruited as before. Apparatus, stimuli, and procedure were as described in Supplemental Experiment 1 with the following exceptions.

Stimuli

The only difference in stimuli between Supplemental Experiment 1 and this one was that here dots were shown at two different presentation rates across trials. In "fast" trials, each dot was present for 150 ms and was followed by a 50 ms blank inter-dot interval. In "slow" trials, each dot was present for 300 ms and was followed by a 150 ms blank inter-dot interval. It should be noted that "fast" trials used the same presentation rate (5 Hz) as in Experiment 1, but that "slow" trials used half that rate (2.5 Hz).

Procedure

There was no feedback to participants in this experiment, with the feedback period being replaced by a 1500 ms blank inter-trial interval. Participants completed 400 trials each, in ten blocks of 40. Each block contained either all fast trials or all slow trials, with fast and slow blocks alternating over the course of the experiment. First block type encountered was random and counterbalanced across participants. Participants were explicitly told about the different presentation rates, and encountered the same number of both types during the practice phase, which consisted of about ten trials.

Results

Weights were obtained for each participant for each condition in the manner described previously. Mean weights across participants in each presentation rate condition are shown in Figure S2.

Again an unmistakable pattern of primacy was found for estimates of mean dot location, evidenced by a significant main effect of temporal position, $F(9,135) = 25.56$, $p < 0.001$, $\eta_p^2 = .63$. Interestingly, the pattern of weights was indistinguishable across the two presentation rates, with no significant main effect of presentation rate, $F(1,15) = 0.78$, $p = 0.3$, $\eta_p^2 = .05$, or interaction between presentation rate and temporal position, $F(9,135) = 0.53$, $p = 0.85$, $\eta_p^2 = .03$. Giving the participants more time per dot to update their belief about the mean position of the dots did not appear to affect the temporal weighting profile of their responses at all. Primacy was still found.

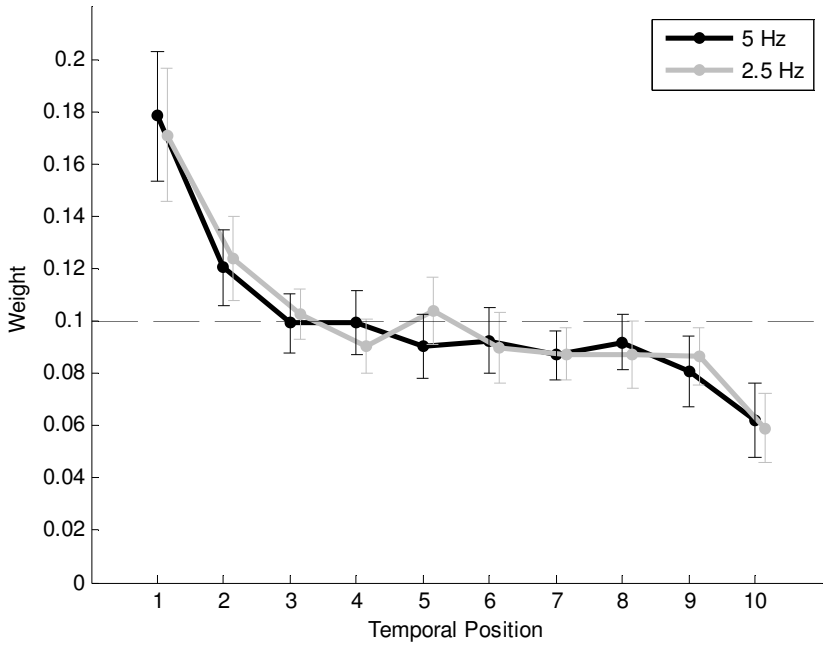

Figure S2. Results from Supplemental Experiment 2.

SUPPLEMENTAL EXPERIMENT 3: AVERAGING POSITION WITH MORE OR LESS VARIABLE DOTS

Instead of manipulating presentation rate within participants, in this experiment we manipulated the average spatial spread of the dots to see if this would the temporal weighting profile the participants applied to the sequence of dots in judging mean position.

Method

A new group of seventeen participants was recruited as before. Apparatus, stimuli, and procedure were as described in Supplemental Experiment 2 with the following exceptions.

Stimuli

Two trial types were used in this experiment. All stimulus parameters were the same between trial types except for the standard deviation of the sampling distribution that dot positions were sampled from. “Narrow” trials used a standard deviation of 1.5° of visual angle and “wide” trials

used a standard deviation of 4.8°. All trials used a presentation rate of 5 Hz, just as in Supplemental Experiment 1 and in Experiment 1 in Chapter 2.

Procedure

Participants completed 400 blocks of 40 trials each, with the two trial types randomly intermixed throughout all blocks. Participants saw an equal number of narrow and wide trials in the practice phase, which consisted of about ten trials.

Results

Weights were obtained for each participant for each condition in the manner described previously. Mean weights across participants in each location variance condition are shown in Figure S3.

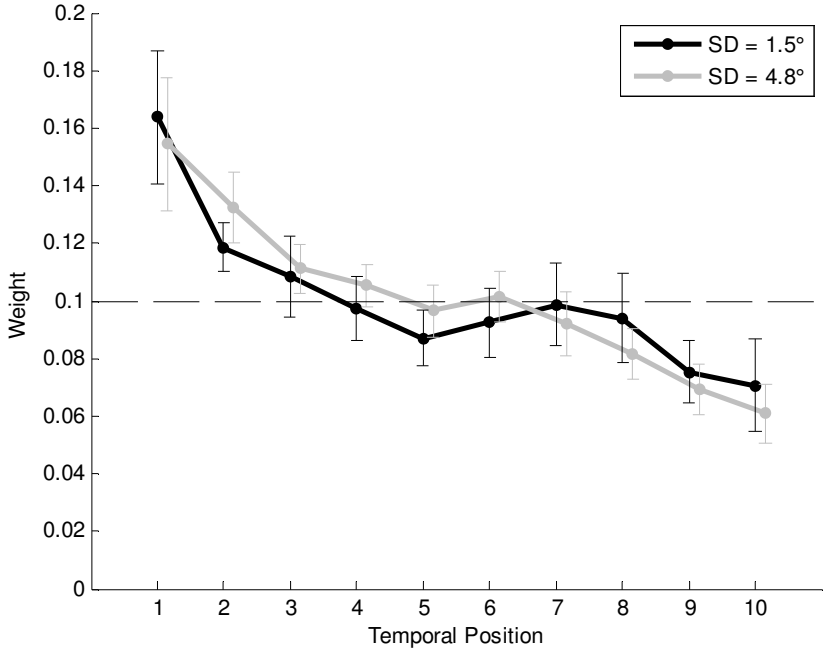

Figure S3. Results from Supplemental Experiment 3.

Once again, an overall pattern of primacy was found, with earlier dots contributing more to participants' responses than later dots. This observation was supported by a significant main

effect of temporal position in a two-way ANOVA, $F(9,144) = 21.02$, $p < 0.001$, $\eta_p^2 = .57$. However, changing the mean spatial spread of the dots shown on each trial did not appear to affect the weights participants applied to the dots, with no significant main effect of the condition variable, $F(1,16) = 0.31$, $p = 0.58$, $\eta_p^2 = .02$, and no significant interaction between variables, $F(9,144) = 1.67$, $p = 0.10$, $\eta_p^2 = .09$.

SUPPLEMENTAL EXPERIMENT 4: AVERAGING SIZE

The purpose of this experiment was to test for the presence of recency in computing summaries for mean size across time when dot position did not vary and only size judgments were being made.

Method

A new group of fifteen undergraduate students at the University of Washington was recruited and compensated as described in Experiment 1 of Chapter 2.

Stimuli

On each trial in this experiment, participants viewed a series of eight white dots that varied only in their size. Each dot was shown for 200 ms and was followed by a 133 ms blank inter-dot interval, resulting in a dot presentation rate of 3 Hz. On a given trial, dot radii were chosen by sampling eight times from a Gaussian probability distribution, the center of which was sampled on each trial from a uniform distribution ranging from 0.6° to 2.3° of visual angle. The standard deviation of the Gaussian distribution was always set to 0.3 times the center of the same distribution in order to minimize the possibility of sampling dot radii below zero. If a dot's radius was sampled to be below zero, it was resampled until it was above zero.

The stimulus area and the response area were separated in distance, with the stimulus dots appearing 5° to the left of the center of the screen and the response dot (which only appeared after the stimulus dots) appearing 5° to the right of the center of the screen. A pair of grey vertical hash marks marking the horizontal position of these locations was visible throughout the experiment so that participants knew where to expect stimulus and response dots. Each hash

mark in a pair was located 5° vertically away from the horizontal meridian of the screen such that even large dots rarely came within 1° of the mark.

Procedure

Each trial began with the presentation of a white cross approximately 0.8° in width and height at the dot presentation location. After 250 ms, the cross became red to signal the imminent arrival of the dot sequence. After a random period between one and two seconds, the cross disappeared and was immediately followed by the dot sequence. The dot sequence was followed by a 600 ms blank period, which was followed by the response period. Here the participants were expected to report the "average size" of the dots shown on that trial by adjusting a sample response dot. The response dot's initial radius was chosen randomly on each trial from 0.1° to 3.2°, which also served as the limits of possible responses. The response period ended when the participant submitted his or her response, and was followed by a 500 ms inter-trial interval.

Each participant completed 350 trials and was given opportunity to take breaks after every 40 trials, though participants were also advised that they could take a break at any time by waiting to submit their response for a given trial. Each participant received full instructions from an experimenter and then completed approximately ten practice trials in view of the experimenter before beginning the experimental trials. A full experimental session lasted about an hour.

Results

Weights were obtained for each participant for each condition in the manner described previously. Mean weights across participants are shown in Figure S4.

As in Experiment 1 in Chapter 2, we found recency for computing mean size across time, where the later dots appeared to contribute more to responses than earlier dots. This was supported by a significant effect of temporal position in a one-way repeated measures ANOVA, $F(7,98) = 29.08$, $p < 0.001$, $\eta_p^2 = .68$. Though the overall pattern of recency likely contributed significantly to this outcome, the statistical significance was likely also partially driven by the effect seen in the first three dots. Interestingly, the first dot appeared to suffer less downweighting than the second, giving the temporal weighting profile a checkmark-like appearance that was not observed in Experiment 1 in Chapter 2.

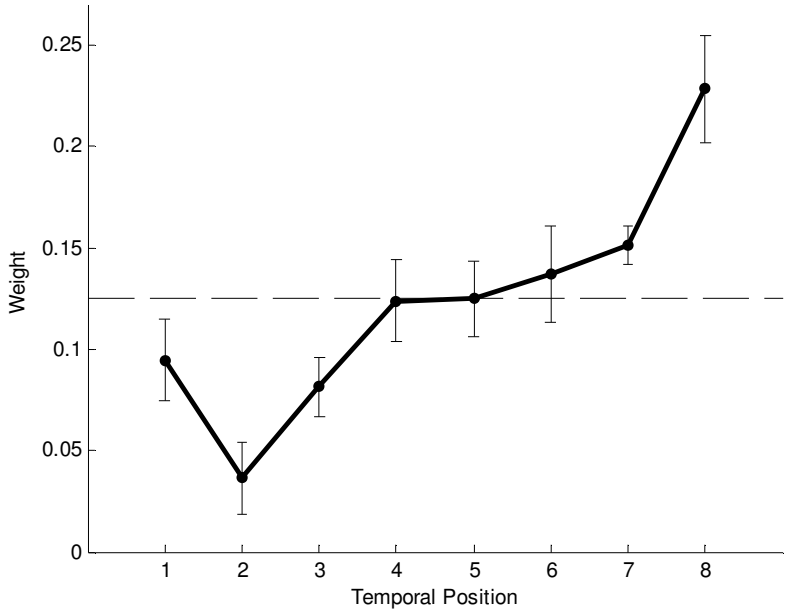

Figure S4. Results from Supplemental Experiment 4.

SUPPLEMENTAL EXPERIMENT 5: AVERAGING SIZE AT DIFFERENT PRESENTATION RATES

In this final experiment, we manipulated the rate of dot presentation and measured the effect on temporal weighting profiles for judgments of mean size.

Method

A new group of twenty participants were recruited and compensated as described previously. Apparatus, stimuli, and procedures were the same as in Supplemental Experiment 4 with the following exceptions.

Stimuli

The stimuli used here were exactly the same as used in Supplemental Experiment 4, excepting that there were two types of trials in this experiment. In "fast" trials, each dot was present for 133 ms and was followed by a blank inter-dot interval of 67 ms, resulting in a presentation rate of 5

Hz. In "slow" trials, each dot was present for 333 ms and was followed by a blank inter-dot interval of 167 ms, resulting in a presentation rate of 2 Hz.

Procedure

Participants each completed 320 trials in eight blocks of 40. Each block contained either all fast trials or all slow trials, with fast and slow blocks alternating over the course of the experiment. First block type encountered in a given session was random and counterbalanced across participants. Participants were explicitly told about the different presentation rates, and encountered the same number of both types during the practice phase, which consisted of about ten trials.

Results

Weights were obtained for each participant for each condition in the manner described previously. Mean weights across participants in each presentation rate condition are shown in Figure S5.

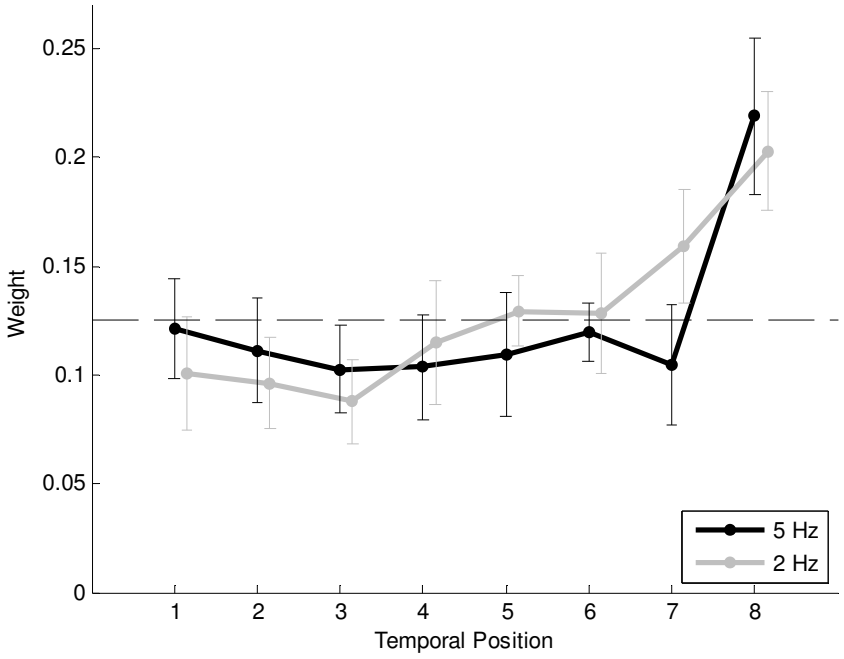

Figure S5. Results from Supplemental Experiment 5.

As in Supplemental Experiment 4 and Experiment 1 in Chapter 2, we found a general pattern of recency for estimates of mean dot size across time. This was supported by a significant main effect of temporal order in a two-way repeated measures ANOVA, $F(7,133) = 13.77$, $p < 0.001$, $\eta_p^2 = .42$. Additionally, a significant main effect of presentation rate was also found, $F(1,19) = 9.80$, $p < 0.01$, $\eta_p^2 = .34$. Interestingly, inspection of the temporal weighting profiles suggests that weights increased gradually over the course of a trial in the 2 Hz presentation rate condition, while weights remained relatively constant until the last dot in the 5 Hz presentation rate condition, where the weight increased abruptly. However, the interaction between temporal order and presentation rate only approached significance, $F(7,133) = 1.96$, $p = 0.07$, $\eta_p^2 = .09$, providing no evidence that varying presentation rate affects the shape of the temporal weighting profile.

General Discussion

The primary purpose of the experiments presented here was test the reproducibility of the basic findings of Experiment 1 in Chapter 2. Results from all five supplemental experiments did exactly that, consistently finding primacy for estimating mean dot position and recency for estimating mean dot size across a variety of conditions.

The secondary purpose of these experiments was to measure the effect of various stimulus manipulations on the temporal weighting profiles in both position and size domains. In the position domain, we manipulated presentation rate, the variance of the underlying sampling distribution, and incorporated corrective feedback, but none of these conditions lead to a change in the temporal weighting profile participants applied to the stimuli, at least to the extent discernable by our analysis method and statistical power. In the size domain, manipulating presentation rate produced two weighting profiles that appeared to differ in shape, but the effect was not statistically significant. In Chapter 2, we raise the point that summary representations computed across time may differ across feature domains, so the possibility that weighting profiles in the size domain might be more malleable than those in the position domain is intriguing. However, our supplemental results do not allow strong conclusions to be made on this point, since we were not able to significantly or consistently perturb the pattern of weights participants applied to the sequences of dots in either position or size domains. Further study will

be needed to explore how manipulable temporal weighting profiles are in various feature domains.

VITA

Bjorn Hubert-Wallander was born to Nancy Hubert and Jan Wallander in January of 1986 in Los Angeles, California. The family moved to Birmingham, Alabama later that year, and Bjorn graduated from Mountain Brook High School in 2004. He subsequently enrolled at Vanderbilt University in Nashville, Tennessee, completing his Bachelor of Arts in 2008, with majors in Psychology and English. While there, he worked as an undergraduate research assistant in Dr. Randolph Blake's vision research lab in the Department of Psychology. From 2008 to 2010, Bjorn was employed as a lab manager for Dr. Daphne Bavelier's lab in the Department of Brain & Cognitive Sciences at the University of Rochester in Rochester, New York. In the fall of 2010, Bjorn entered the graduate program in Psychology at the University of Washington in Seattle, Washington under the advisement of Drs. Geoffrey Boynton and Scott Murray. He had a pretty good time there.