Model comparison of the effects of stimulus structure on visual working memory recall

Madeleine Grunde-McLaughlin and Cheng Qiu and Alan A. Stocker

Department of Psychology, University of Pennsylvania

Goddard laboratories, Rm 421
3710 Hamilton walk
Philadelphia, PA 19106

Abstract

Multiple models attempt to represent visual working memory by reflecting biases in visual recall. Models that store visual memories independently account for the limited resources of working memory, but do not account for the attraction bias that pulls the recall of different stimuli towards their mean. Models that group stimuli into hierarchies account for the attraction bias, but do not account for the repulsion bias in which stimuli are remembered as further away from a category boundary. Models that are hierarchical and self-consistent account for both the attraction and the repulsion bias.

In a study in which subjects recall the placement of three dots, we find both attractive and repulsive biases in our data. Unlike previous work that looked at repulsion from category boundaries, our experiment finds that attractive and repulsive biases appear for non-explicit categories based on a dot's position in a cluster. Our data best supports the hierarchical and self-consistent model. However, we also find a limitation with this model: it does not account for the global context across all trials when predicting the way information is recalled. These findings promote future work to incorporate global priors into the working memory model and to examine the repulsive bias in other kinds of stimuli.

Introduction

The human memory system includes the visual working memory system that temporarily stores a small amount of visual memories that can be easily accessed. Let's say you are making a purchase online and need to insert your credit card information into a payment form. After looking at your credit card, you keep the numbers fresh in your mind in order to type them into the form. The numbers stay in your visual working memory: easily accessible but soon forgotten without extra effort moving the information to long-term memory.

Working memory must efficiently and effectively encode large amounts of incoming information. However, it degrades over time and has limited resources and space. Multiple presented models of working memory (see Figure 1 and Table 1) attempt to deal with this juxtaposition of remembering large amounts of information with limited resources (Cowan, 2001; Orhan and Jacobs, 2013, Stocker and Simoncelli, 2007). We analyze how people encode stimuli with different spatial relationships and find that people build a mental model of the overall structure of a stimulus in order to remember each individual point. The biases that result form this encoding of the overall structure support the hierarchical and self-consistent working memory model. However, we also find limitations of the this model in its current form, as it does not include the effect of priors from the experiment's global context across trials.

Much of the research to understand visual working memory focuses on measuring the capacity of information it can hold (Nassar et al., 2018; Wilken and Ma, 2004; Zhang and Luck, 2008). Most likely, you are unable to hold all digits of the credit card in your working memory at once, so this branch of research looks at the number of items working memory is able to hold at one time. Many models attempt to recreate and understand working memory capacity. In some, memory holds a set number items discreetly, while in others memory has a flexible capacity that adjusts to the stimuli. Regardless, most models have assumed different visual stimuli are stored independently in working memory (Bays and Husain, 2008; Cowan, 2001; Fougnie et al., 2012; van den Berg et al., 2012; Wilken and Ma, 2004; Zhang and Luck, 2008). In the slot based model, for example, working memory contains a set number of slots that each hold individual items independently (Cowan, 2001). Working memory models with this independence assumption can account for some aspects of visual memory. However, they do not account for several key biases and interactions found in psychophysical experiments (Bae and Luck, 2017; Brady, 2015; Brady and Alvarez, 2011; Brady and Tenenbaum, 2013; Ding et al., 2017; Golomb, 2015).

In response to these observed biases, a small but growing cohort of researchers have begun testing a class of models that question the independence assumption. In the chunking model, groups of individual items are encoded jointly into memory, and these chunks are stored as individual units (Cowan, 2001; Nassar et al., 2018.). Alternatively, the hierarchical model holds information about each individual item as well as cumulative information about the higher-level structures of the overall picture (Brady, 2015; Brady and Alvarez, 2011; Brady and Tenenbaum, 2013; Ding et al., 2017; Orhan and Jacobs, 2013). These higher-level structures may include summary statistics of cluster and category (Brady, 2015; Orhan and Jacobs, 2013) or formation of a scene (Brady and Tenenbaum, 2013). Although both chunking and hierarchical models improve on models with an independence assumption, hierarchical models have represented human responses better than chunking models when compared on the same dataset (Brady, 2015; Brady and Tenenbaum, 2013).

It is difficult to measure the structure of memory directly. Although people can experience stimuli and experience the feeling of recalling a memory, they cannot consciously access the method or representations by which the brain holds this remembered information. Instead, researchers can analyze the errors of recall experiments to infer the structure in which items are encoded in memory. The act of recalling a previously seen stimulus uses the decoding process that converts memory storage into remembered visual characteristics about the world. Subjects rarely remember stimuli exactly, so consistent biases on recall hint at how the stimuli are processed and stored in memory (Bae and Luck, 2017; Orhan and Jacobs, 2013). Historically the decoding process had been assumed to start with processing low-level characteristics stored in memory that build up to reconstruct high-level characteristics. However, a growing body of research found psychophysical data that is better explained by a model that processes high-level characteristics before recreating low-level ones (Ding et al., 2017; Luu and Stocker, 2018; Stocker and Simoncelli, 2007).

The influence of ensemble statistics on recall can be better explained if subjects remember stimuli as dependent on one another than if subjects assume independence. Brady (2015) briefly showed subjects a group of 2-6 colors, then asked them to recall those colors. If each item had been encoded individually, each individual color recollection would have some degree of noise, yet it would not be affected by the other colors in the original stimulus. Instead, they found that recall biases were statistically skewed towards ensemble characteristics of the group. For example, if the stimulus showed three similar shades of pink, the recalled colors would also be of three similar shades. Even if the exact colors were remembered incorrectly as shades of red, the distribution among the colors remained consistent. After comparing the data to generated results by item-based, chunk-based, and hierarchical models, Brady postulates that the influence of ensemble statistics suggests a hierarchical model of visual working memory with multiple, interacting levels of representation. In another experiment, Brady and Tenenbaum (2013) found that people remember certain formations better than others, even if the formations have the same number of items. They also found that models with both individual and summary-level statistics better account for this formation preference than independent or chunking models.

Hierarchical models may also account for attraction and repulsion biases (Bae and Luck, 2017; Brady, 2015; Brady and Alvarez, 2011; Brady and Tenenbaum, 2013; Ding et al., 2017; Golomb, 2015; Jazayeri and Movshon, 2007; Luu and Stocker, 2018; Orhan and Jacobs, 2013; Stocker and

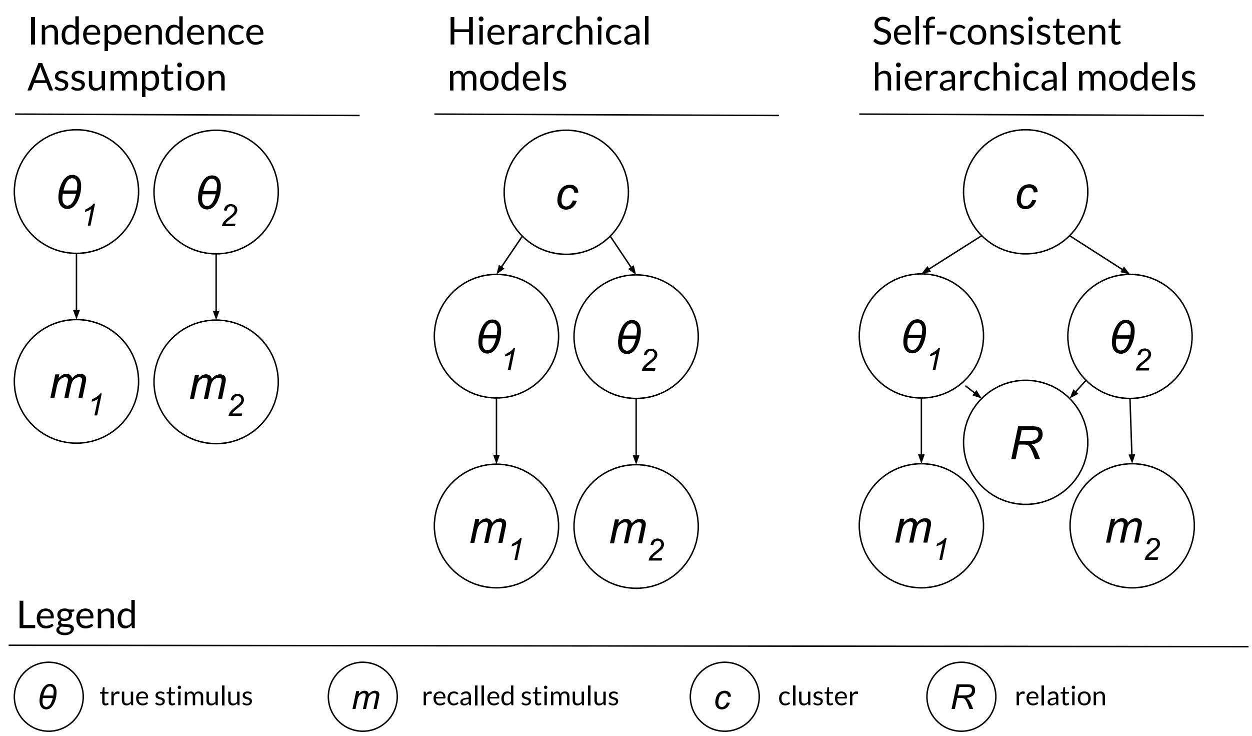

Figure 1: Three categories of proposed models of visual working memory. Models with an independence assumption encode stimuli independently (Bays and Husain (2008), Cowan (2001), Wilken and Ma (2004)). Hierarchical models additionally holds cumulative information about higher level structures (Brady and Alvarez (2011), Brady (2015), Orhan and Jacobs (2013)). Self-consistency adds a relation module (Bae and Luck (2017), Luu and Stocker (2018), Stocker and Simoncelli (2007).

Simoncelli, 2007). As shown in Figure 2, the attraction bias describes the phenomena in which subjects recall the characteristics of a group of stimuli as closer to each other than they were in reality. The repulsion bias describes the phenomena in which subjects recall characteristics of a group of stimuli as further to each other than they were in reality (Golomb, 2015).

Recalled items attract towards ensemble means, especially in characteristics relevant to a task (Brady and Alvarez, 2011; Brady, 2015). Orhan and Jacobs (2013) postulate that items in working memory are remembered in clusters. Elements belonging to the same cluster influence and attract towards each other, while elements belonging to different clusters do not. Brady and Alvarez (2011) found that subjects biased the recalled size of a circle towards the mean size of all circles of the same color. Importantly, they found that this effect occurs primarily when color is relevant to the recall task (i.e. remember all the red circles but not the green circles), and that the effect is not visible when color is not relevant to the recall task. Bae and Luck's (2017) work also indicates that attention affects clustering in memory. They gave subjects two stimuli and indicated which was high priority. Characteristics of the high priority stimulus heavily affected recall of the lower

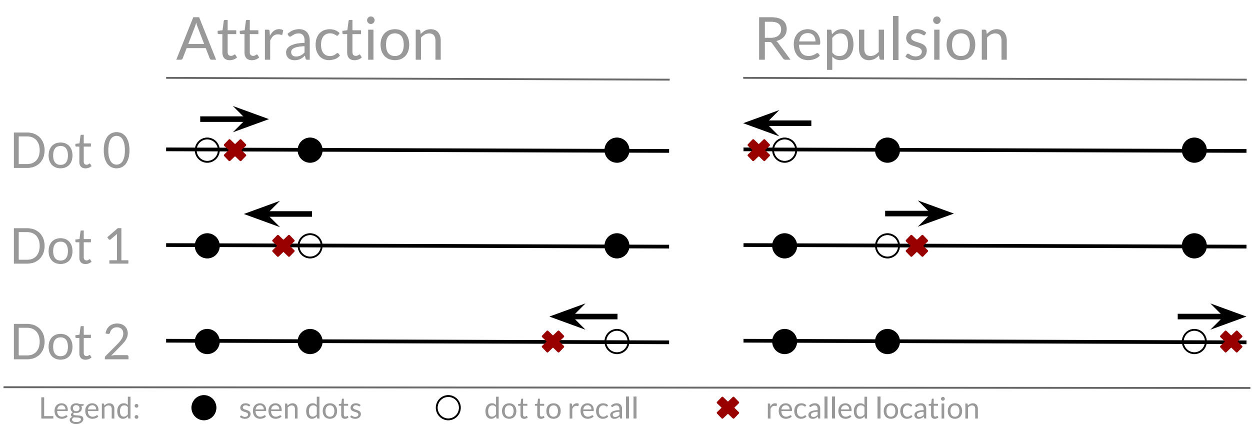

Figure 2: The attraction and repulsion biases appear when people attempt to recall visual memories from working memory. With the attraction bias, people remember stimuli as closer to the others than in reality. With the repulsion bias, people remember stimuli as further from the others than in reality.

priority stimulus, but characteristics of the low priority stimulus had much less of an effect on the higher priority stimulus. The data from these studies suggest that when subjects know of relevant groupings, like color, memory encodes the ensemble statistics about groups considered to be relevant at a higher level of the hierarchy. When remembering items, the hierarchy integrates between these items at each level of abstraction and causes recalled properties to attract towards ensemble means.

A hierarchical structure of working memory may also account for the repulsion bias when subjects are self-consistent within the hierarchy. When recalling data, self consistency means that subjects make decisions at each level of the hierarchy, then treat these decisions as truths in the remainder of their recall process (Stocker and Simoncelli, 2007). Say a subject was tasked with recalling the motion direction of a collection of dots. The motion of this cloud has a high level category of if it moved left or right, and a lower level detail of the precise direction it moved. An optimal Bayesian observer judging the direction the cloud moved would consider all possible categories of movement (left and right) in their prior. In a self-consistent observer, the observer makes a higher level decision of if the cloud moved left or right, then adjusts their prior to only include the possibilities in that category. When a stimulus is close to the category boundary, limiting the prior to focus on one category draws the memory of the stimulus towards category center and repulsed away from the category boundary (Stocker and Simoncelli, 2007).

A self-consistent Bayesian model accounts for human responses like repulsion bias that an optimal Bayesian model does not (Stocker and Simoncelli, 2007). Luu and Stocker (2018) used a reference line at random locations as a category boundary for a stimulus cloud of smaller lines. Participants indicated if they thought the smaller lines pointed on average clockwise or counterclockwise of the reference line. When the average orientation of the stimulus cloud was close to the reference line, the remembered orientation repulsed away from the reference line. The effect occurred at the same magnitude regardless of if the participants made the choice themselves or were told by the experiment which category was correct. This consistency implies that once people make a category choice, they treat it as truth. The repulsion bias is weaker if the line is not there for reference (Zamboni et al., 2016). This dependence of the remembered stimuli on the category boundary implies that decoding processes the categorization before remembering the absolute location. Although self-consistency is not an inherent aspect of the structure of working memory, by showing that self-consistency within hierarchical structures predicts the repulsion bias, these studies suggest that memory is structured in a hierarchical way (Bae and Luck, 2017; Stocker and Simoncelli, 2007; Luu and Stocker, 2018).

There are multiple suggestions as to why memory would be stored hierarchically. Categorical information is more discrete and compact than information along a continuous spectrum. Therefore, storing categorical data is less costly and more computationally efficient than storing detailed individual data about each item. Categorical items may also degrade less over time and help counteract working memory's rapid degradation and noise (Luu and Stocker, 2018). Although it may seem counter-intuitive to ignore certain categories in the prior, remaining self-consistent with higher categories helps the viewer keep a stable view of the world (Stocker and Simoncelli, 2007). Ding et al. (2017) suggest that starting processing with high-level categories first improves perceptual performance. Although this idea was met with some skepticism (Luu et al., 2017), Qiu et al. (2020) report that self-consistent models can have better accuracy in certain constrained situations. Specifically, they report that accuracy can be better if the stimulus is far enough from the discrimination threshold for the chance of an incorrect categorization choice to be low.

This project extends from the current working memory literature in two ways. First, existing literature on self-consistency looks at self-consistency based on some categorical boundary. For example, the stimuli is encoded to be on the right of some boundary, so it is encoded categorically as being on the right and it's recall is affected by that categorization. Rather than stimuli relative to some category boundary, this project finds repulsion and attraction biases based on the relative placement of different stimuli with one another from the size of clusters. We find that the hierarchical and self-consistent model best represents our results. Second, we extend previous literature by also looking at the effect of the average distance between stimuli across all trials. The current hierarchical and self-consistent models does not account for this inter-trial effect.

Table 1: Multiple categories of models have been proposed by a variety of authors to represent working memory and the effects it has on recall.

Model	**Prominent Papers**	**Effects it can account for**
Independent encodings	Bays and Husain (2008) Cowan (2001) Fougnie et al. (2012) Wilken and Ma (2004) van den Berg et al. (2012) Zhang & Luck (2008)	Individual noise dependent on set size
Chunking	Cowan (2001) Nassar et al. (2018)	Attraction
Hierarchical with individual and ensemble level statistics	Bae and Luck (2017) Brady and Alvarez (2011) Brady and Tenenbaum (2013) Brady (2015) Orhan and Jacobs (2013)	Attraction
Hierarchical with Relational representation	Bae and Luck (2017) Ding et al. (2017) Luu and Stocker (2018) Stocker and Simoncelli (2007)	Repulsion

Hypotheses

Hypothesis 1: The relative cluster size of stimuli affects how each individual stimulus is represented and recalled from working memory. This internal model of the cluster structure will manifest in attractive and repulsive biases.

Hypothesis 2: We define the global prior as a prior of the overall wedge size developed from the number of degrees between the two outer dots across all conditions. Our second hypothesis states that the recalled space between the two outer dots will be attracted to the global prior.

Methods

Using AMT

We used Amazon Mechanical Turk (AMT), a crowdsourcing platform, to gather data. Through this format, we could gather psychophysical data from a wide range and larger number of subjects. Furthermore, gathering data through AMT provided fewer health risks than bringing subjects into a lab. We gathered data from 69 US workers, and paid them the equivalent of 15 USD per hour to comply with fair work standards on the Amazon Mechanical Turk platform (Whiting, 2019). We develop our interfaces using JSPsych and EasyTurk(de Leeuw, 2015; Krishna, 2019). We gave bonuses to the 10 workers with the lowest error rates.

Our Interface

We present the AMT workers with a series of 180 trials (see Figure 3). For each trial, they first see a circle and a fixation point. Three dots appear on the circle for 200 milliseconds, followed by a cloud of dots to add noise and force the workers to rely on working memory. Then, two of the three dots re-appear, and workers must click where they believe the third dot had been. They are able to re-adjust the dot if necessary. When they are happy with the dot placement, they press a submit button on the bottom right of the screen. The placement of the button ensures that they take the cursor far enough away from the circle that it does not automatically place a dot at the start of the next trial and ensures that the worker must re-fixate on the circle center each round.

As AMT workers' workplaces are more varied than a lab's, we standardize the set-up in two ways. First, we ask workers to stay one arm's length away from the screen. Second, we provide an interface in which they adjust the size of the circle such that it is approximately the size of their fist. The adjustment parameter is used to scale the size of the stimuli to achieve comparable display size across workers. To encourage workers to contribute the best data possible, we provide

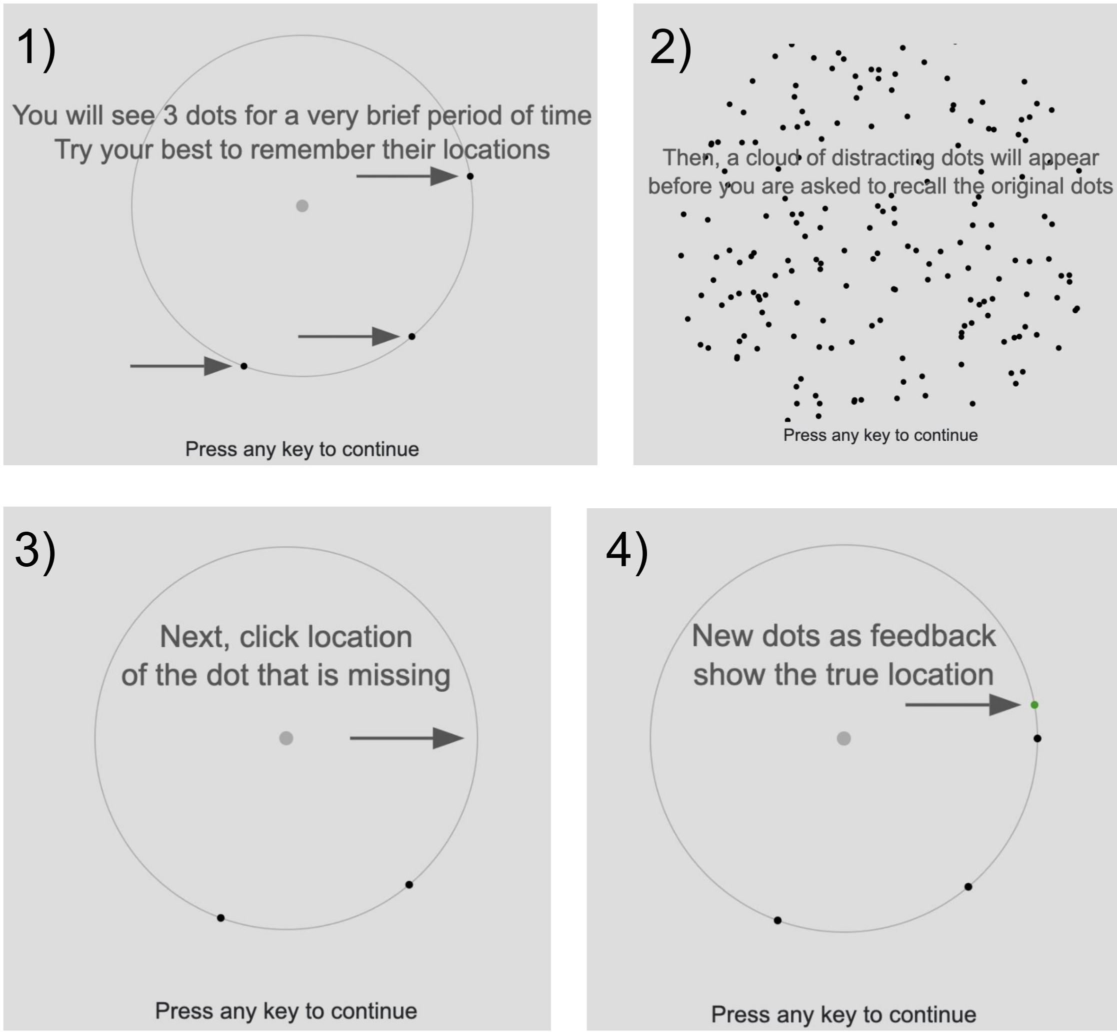

Figure 3: In our experiments, workers see three dots. They are then presented with a mask of dots. Finally, they see two of the dots and guess the location of the third dot. We provide feedback of the ground truth location of the dots to the participants after each trial.

feedback each round in the form of showing the placement of the original dot. If the amount of distance between the guessed dot and the real dot is less than 15 degrees, the placement of the original dot appears in green. Otherwise, the original dot appears in red. We advertised and gave bonuses to 10 workers for high performance, and told workers we would reject work that had a high overall error rate. After reviewing responses, we did not reject any workers.

Parameters

Our stimulus shows three dots on a circle. The absolute positions are randomly selected each trial. To measure the effects of clusters, each trial either has the dots equidistant from one another, or

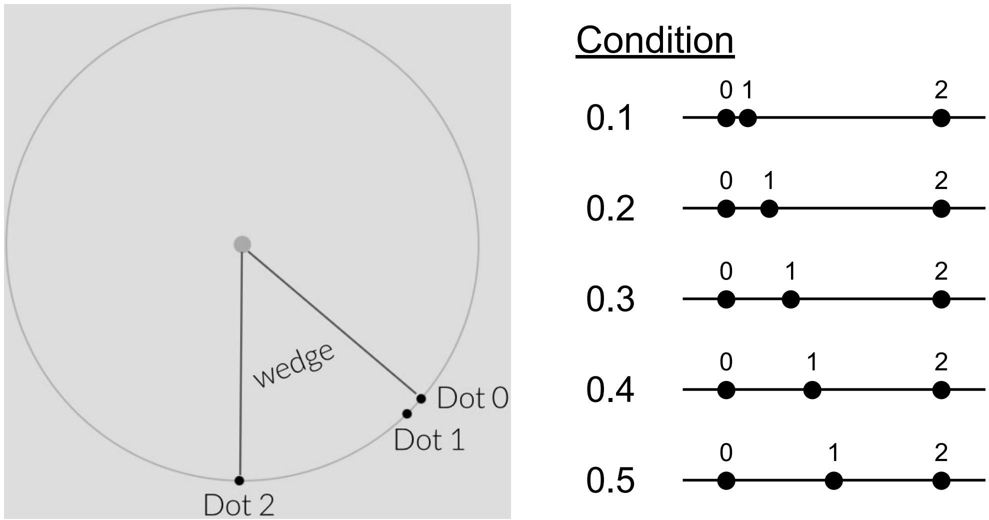

Figure 4: The wedge refers to the number of degrees between the two outer dots. We refer to the middle dot as Dot 1, the closest outer dot as Dot 0, and the furthest outer dot as Dot 2. The middle dot can be at 10, 20, 30, 40, or 50 percent distance between the two outer dots.

has one cluster of two dots and one dot relatively farther away. The middle dot placement varies among 10, 20, 30, 40, and 50 percent of the way between the two outer dots (conditions 0.1, 0.2, 0.3, 0.4, and 0.5). We doubled the conditions such that the side of the smaller cluster alternates to maintain symmetry among structures (i.e. the furthest outer dot would not always be on the clockwise side) and reduce the effect of priors.

We refer to the overall cluster of three dots as a wedge. We refer to the dot in the middle of the wedge as the middle dot or Dot 1. The dot it is closest to we refer to the near outer dot or Dot 0. The dot it is furthest from we refer to as the far outer dot or Dot 2. For a visual representation of these terms, please refer to Figure 4. To reduce priors on the overall shape of the three dots, we vary the distance between the outer two dots, or the wedge, among 60, 80, and 100 degrees. Some workers saw wedges that were 60 and 80 degrees, while another group of workers saw wedges that were 80 and 100 degrees. We split the trials into these two groups in order to place the same absolute wedge size (80 degrees) in two different global contexts, with either a larger wedge or a smaller wedge. In total, workers saw 180 total trials that varied across the wedge size, cluster size, and which dot of the three would be the target for recall.

Results

Figure 5: Here we present the raw data from our experiments. Each column shows a condition in which the middle dot is a certain percentage between the two outer dots. Each row section shows responses when Dot 0, 1, or 2 is missing, and each row section is broken down by wedge size. Please refer to Figure 4 for a visualization of these terms. The raw data shows the repulsion bias, the attraction bias, and bimodality. These phenomena best support a self-consistent hierarchical model of working memory.

We share combined participants' recall distributions for each testing condition from our experiments in Figure 5. Each column represents a different cluster size. Each row section contains the data in which the worker guessed a particular dot. Within each row section, each individual row represents a different wedge size condition. 80-a contains the data from wedge size 80 in the condition when the worker also saw 60 degree wedges. 80-b contains the data from wedge size 80 in the condition when the worker also saw 100 degree wedges. We overlay the results from all wedge sizes, normalized to the same scale, in supplementary materials.

Based on the recall distributions we can see the attraction and repulsion biases. Attraction biases occur when the distribution of responses is biased to be closer to the closest dot (e.g. Dot 0 in the 0.3 condition). The repulsion bias occurs when the distribution of responses is biased to be further away from the closest dot (e.g. Dot 1 in the 0.1 condition).

For some conditions, such as guessing Dot 0 in the 0.1 condition, the distribution has two peaks. This bimodality is an effect of the overall encoding of the cluster structure. People will remember approximately the distance between the two clustered dots, but they may think they need to guess a different dot and instead guess on the other side. Consider an example in which, Dot 0 disappears and the participant sees Dot 1 and Dot 2. They may think that the Dot 1 they are shown is actually Dot 0, and then guess a point in the middle of the two dots they see. This "flipping" shows that people remember the overall structure of the stimuli, and they do not remember each dot independently.

The attraction biases, repulsive biases, and bimodality we see reflect a hierarchical self-consistent model (see Figure 7). In the self-consistent hierarchical model, people encode the true stimuli in a hierarchy based on the clusters. When recalling the dot, this hierarchy attracts the prior of the missing stimulus towards the sub cluster (orange). However, when the missing stimulus is close enough, the relation between the two dots in the cluster (teal) pushes the prior of the missing similar (grey) away from the cluster center. This encoding accounts for both the attraction and repulsion biases. When the participant mistakes which dot in the cluster is missing, we see a flipping effect that leads to bimodal distributions.

Aggregation Methods

We then further investigate the biases that emerge in our experiments. We look at three measurements: bias, flipping percentage, and variance.

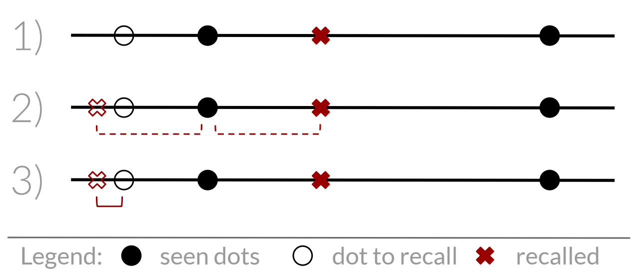

Figure 6: Participants sometimes make flipping errors. A flipping error occurs when the participant mistakes which dot is missing. In this case, the outer dot is missing but the participant believes the middle dot is missing. To better measure repulsion and attraction biases, we aggregate the bias by first flipping the guess over the dot it is closest to before measuring the bias.

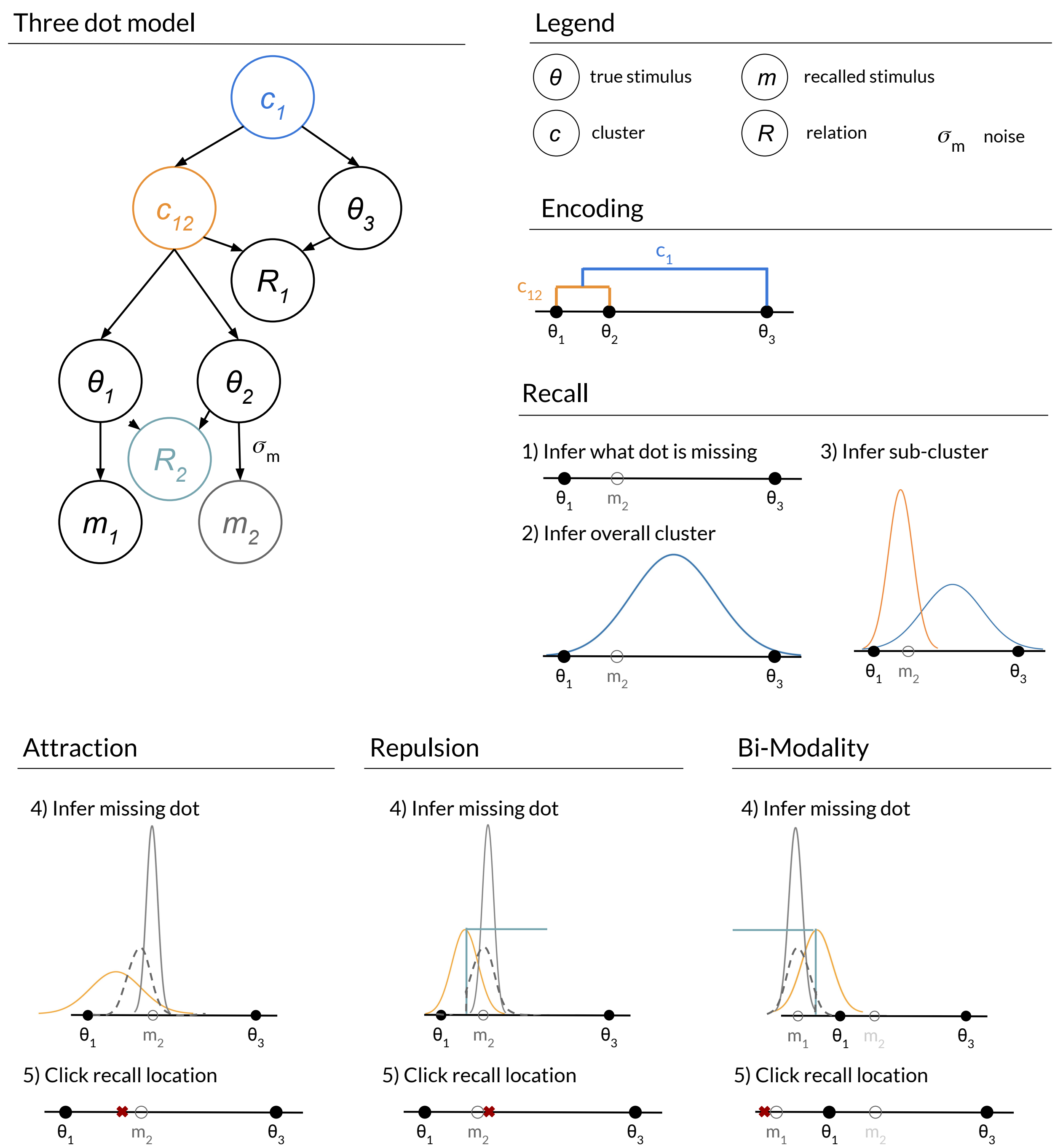

Figure 7: The self-consistent hierarchical model accounts for the attraction, repulsion, and biomodality we see in our results. We visualize the methods by which these phenomena appear in steps 4 and 5 of recall. Attraction: When the dots are further away, the dot to guess is attracted towards the sub-cluster. Repulsion: When the dots are close, the sub cluster instead repulses the missing dot's prior distribution away from the sub-cluster. Bimodality: When the participant mistakes which dot they believe is missing, we see a flipping effect.

Bias:

To aggregate the bias, we take the average distance between the worker's guess and the ground truth. However, this formula is less representative of the bias in the worker's mind when the percent of responses that are flipped is high. To account for this, we aggregate the bias of flipped responses using a different strategy as outlined in Figure 6. We count a response as flipped if it is closer to a different dot than to the dot that was missing. We first flip the clicked response over the closest dot before measuring the bias. The bias when considering the distance between the clicked dot and the stimulus dot can be found in supplementary materials.

Flipping percentage: This metric measures the percent of all "flipped" responses. In other words, it measures the percent of responses that are closer to a dot other than the missing dot.

Variance:

We measure the variance of the distribution of responses across all participants.

Intra-condition

Within the 60a-80a and 80b-100b conditions there are some common threads (see Figure 8). We refer to 60a and 80b as the smaller wedges and 80a and 100b as the larger wedges. As expected, the repulsive bias appears for Dot 0 and Dot 1 when they are in a close cluster. We expect the dots to show attractive bias towards the other clusters when they are far enough away to not have the repulsive bias. All wedge sizes show more attractive bias for Dot 0 than Dot 1. This effect may occur because the middle dot can use both outer dots as reference in the 0.4 and 0.5 conditions, while the outer dots are pulled towards the other two shown dots.

For both conditions, the outer dots of the larger wedge sizes show higher levels of attractive bias. This is likely because they are attracting towards the global wedge size prior which is smaller than their ground truth wedge size. For example, in the 60a-80a condition the average wedge size is 70, so the stimuli with a wedge size of 80 recall the outer dots as pulled in towards the average. For the smaller wedge size conditions, the outer dots still show an attractive bias, but it is much smaller. Since we expected to see an attraction bias when outer dots are not close to the cluster, the recall may still be attracting towards the global prior and showing a less strong attractive bias than it would have otherwise.

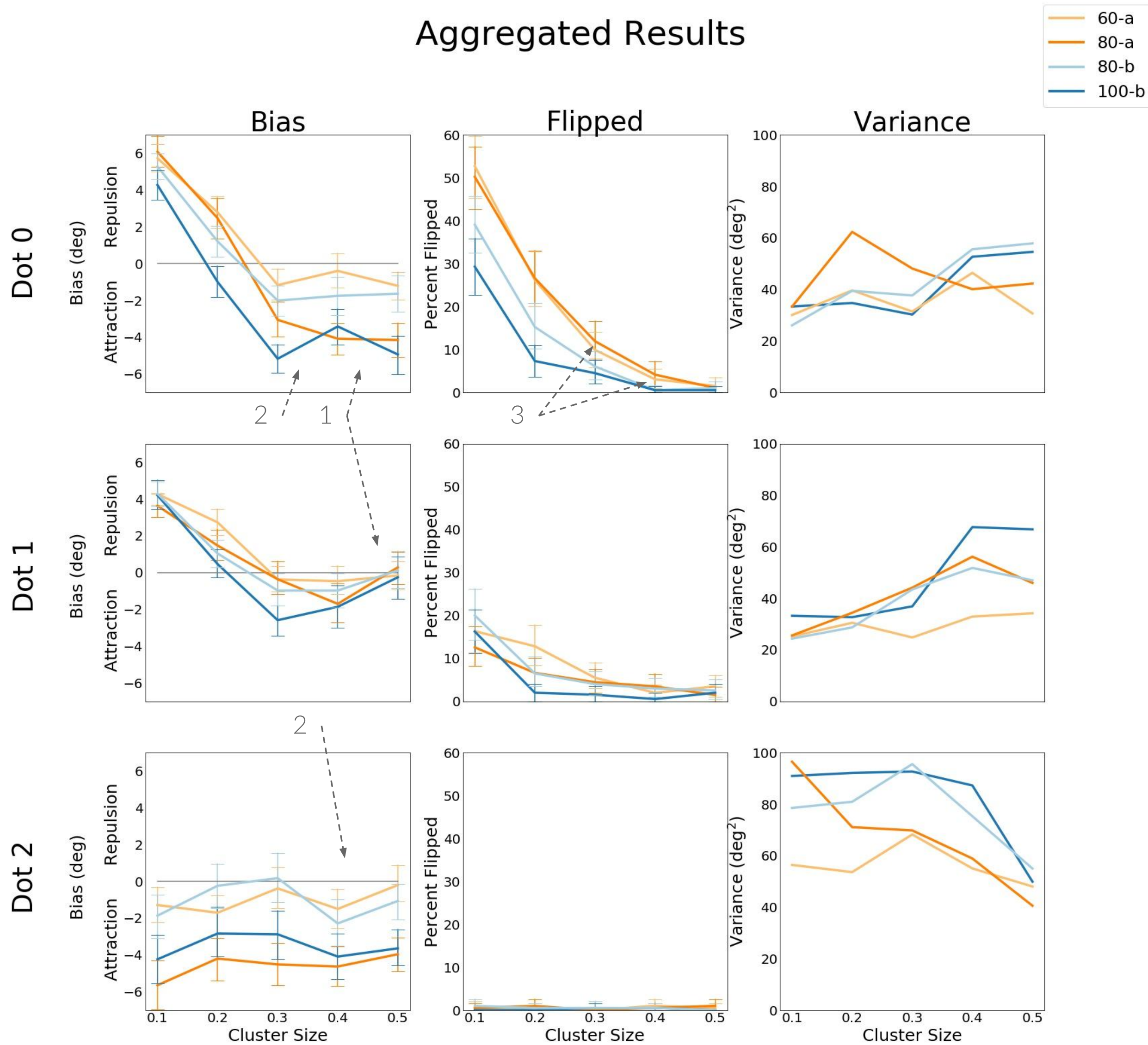

Figure 8: The results when guessing Dot 0, 1, or 2 aggregated for the average bias, the percent of results that were flipped, and the variance. We see both the repulsive bias in Dot 0 and 1 when the cluster is small, and the attractive bias as the cluster becomes larger. (1) The attraction bias is stronger for Dot 0 than Dot 1 because Dot 1 has both outer dots for reference, while Dot 0 attracts towards the other two. (2) The larger wedges (80-a and 100-b) have greater attractive biases than the smaller wedges. (3) The 60a-0.4 condition and 80a-0.3 both have clusters of 24 degrees, but there are more flipping errors in the 0.3 condition than the 0.4 condition. This implies that the relative size of the cluster is more influential than the absolute difference between two dots.

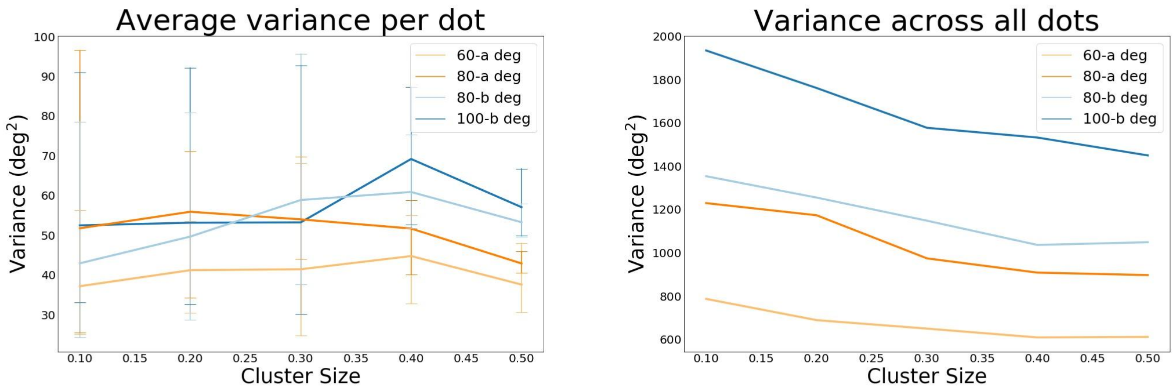

Figure 9: Here we show the variance when split by dot and on the overall responses across all dots. When accounting for the variance across all dots, 80-b has a higher variance than 80-a, even though both have wedge sizes of 80 degrees.

There is a higher chance of flipping when the dots are in a relatively small cluster (i.e. when the middle dot is 10% or 20% of the distance between the two outer dots). Even though all possibilities are equally likely, people also appear to have a prior that the dot they are guessing will be in the middle. This prior manifests as people make the flipping error more often on Dot 0 than Dot 1. The chance of flipping depends more on the relative size of the cluster than the absolute distance in degrees between the two dots in the cluster. For example, the .4 condition of 60a degrees (60a-0.4) and the .3 condition of 80a degrees (80a-0.3) each have 24 degrees difference between the two dots in the cluster. However, when the participant is tasked with guessing Dot 0, they make the flipping error on 3.03% of the 60a-0.4 condition and 11.91% of the 80a-0.3 condition. Although the absolute distance between the two dots was 24 degrees in both cases, participants made flipping errors nearly 4 times as often on the condition when the cluster made up a relatively smaller proportion between the two outer dots.

Finally, variance increases slightly as the dot to be guessed is further from the other dots. This may be because the recall of each dot is aided by remembering it's position relative to other dots, so that noise increases when it does not have other nearby dots to ground it. The 100 degree wedge has a higher overall variance than the 60 degree wedge only when the dot to be guessed is further from other dots (i.e. the 0.4 and 0.5 condition for Dot 0 and Dot 1, and the 0.1, 0.2, and 0.3 condition for Dot 2).

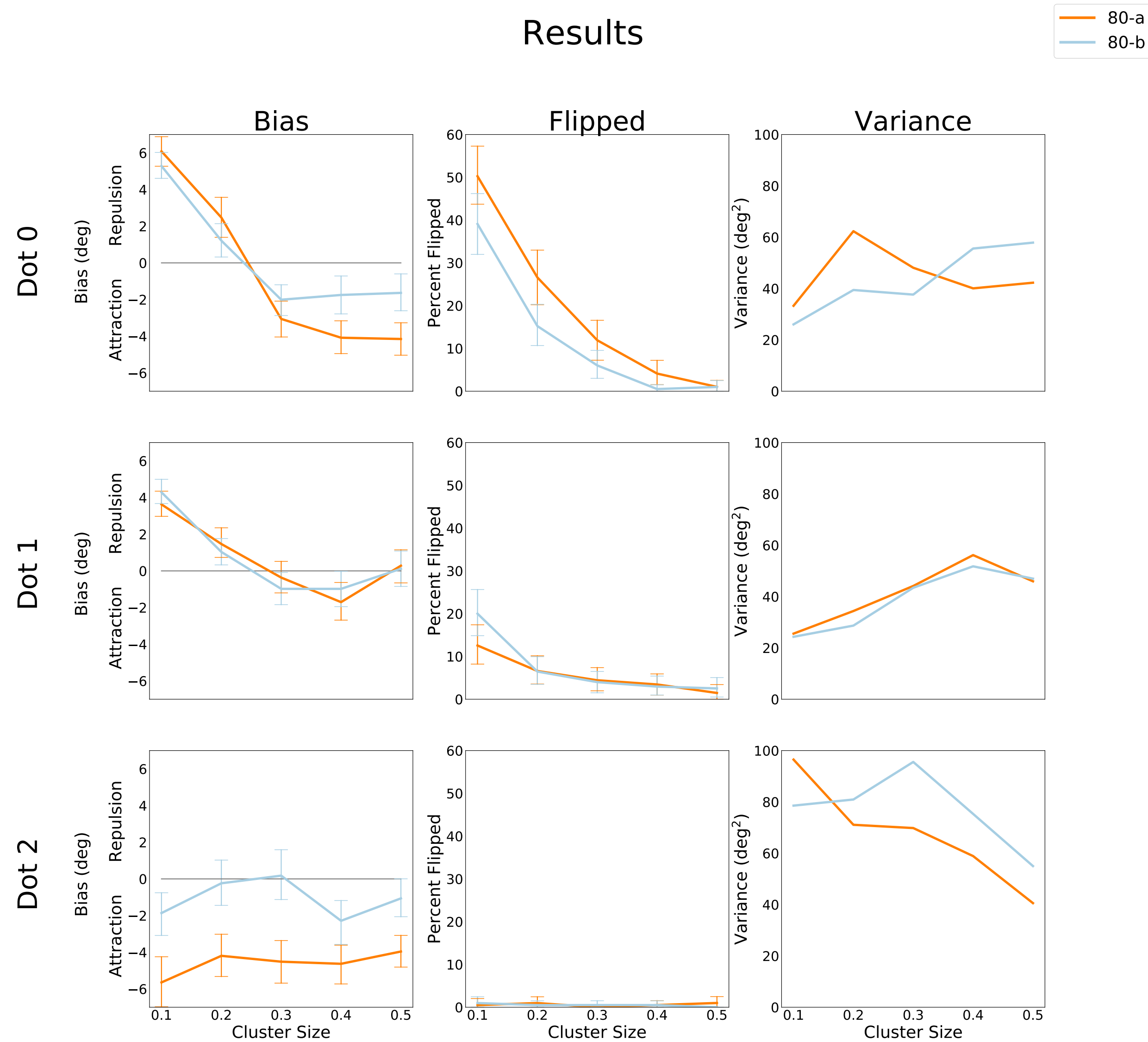

Figure 10: The same results as shown in Figure 8, but only the wedges with 80 degrees. Although the wedges have the same wedge size, they show different trends. The 80-a condition shows more attraction on the outer dots because they are attracting towards the global prior. For the same reason, the 80-a condition finds more flipping mistakes on Dot 0. Both conditions show similar variance on Dot 1, when the global wedge size is less relevant, but vary on the variance when guessing Dot 0 and Dot 2.

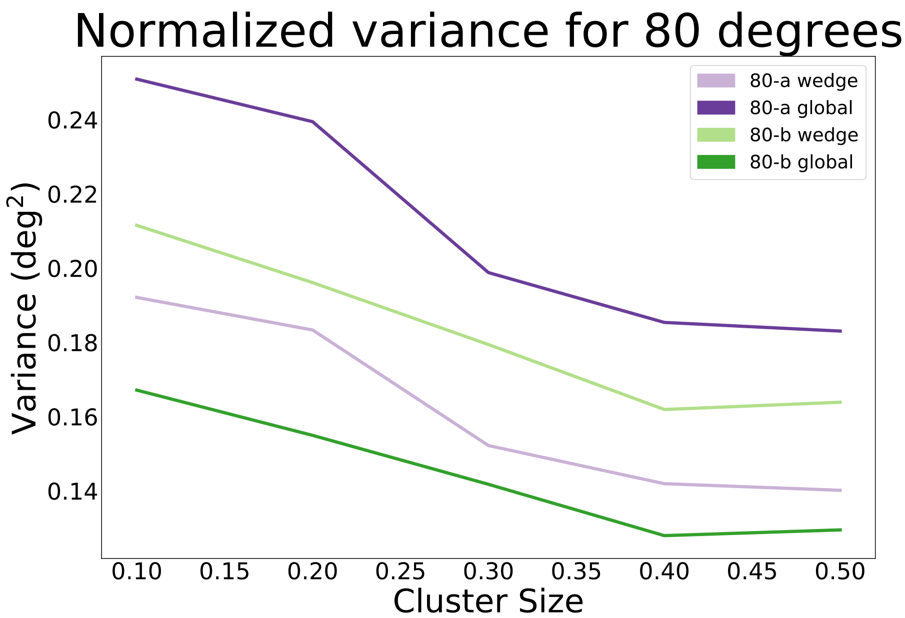

Figure 11: Here we investigate the effect of the global prior on variance of responses in the 80 degree conditions. We normalize each of the responses by either the wedge size (80 degrees), or the global mean (70 degrees for 80-a and 90 degrees for 80-b) before calculating the variance.

Inter-condition

Although both the 80-a and 80-b conditions contain identically stimuli, their results are affected by the other condition. We highlight just these two conditions in Figure 10. The 80-a stimuli are paired with 60 degree wedges, while the 80-b stimuli are paired with 100 degree wedges. All wedge conditions reveal similar levels of repulsive biases for Dot 0 and Dot 1. In these conditions, the global prior has minimal impact because the participant uses the closest dot for point of reference. However, the 80-a condition exhibits much higher levels of attractive bias on the outer dots than the 80-b condition. This is because 80-a is the larger wedge compared to the 60 degree wedge it is paired with, so the outer dots experience more attraction towards the global prior of wedge size. This difference does not manifest for the middle dot (Dot 1) because the overall wedge size is known through the two visible outer dots.

This attraction towards the global prior also manifests in the percentage of flipped guesses. 80-a is more likely to be flipped when the worker is guessing Dot 0, while 80-b is more or equally likely to be flipped when the worker is guessing Dot 1. 80-a's global prior is smaller than it's ground truth, so attraction toward the global prior makes people more likely to make a flipping mistake and think that the middle dot they can see is actually representing the outer dot. 80-b's global prior is larger than it's ground truth, so attraction towards the global prior makes people more likely to make a flipping mistake and think that the outer dot they can see is actually representing the middle dot.

In both conditions, 80 degree wedges experience similar levels of variance when guessing the middle dot. This may occur because the global wedge size is less relevant when guessing the middle dot, so participants show similar behaviors for both. However, the variance the exhibit differs when guessing the outer dots. In most conditions where the outer dots are far away from the middle dot, 80-b exhibits more variance than 80-a. As seen in Figure 11, the variance of the two 80 degree conditions follows similar trends, but different magnitudes, when normalized by the wedge size (80 degrees) and by the global mean (70 degrees for the 80-a condition and 90 degrees for the 80-b condition). The higher the value by which the distribution is normalized (i.e. 70 degrees vs 80 degrees vs 90 degrees), the lower the variance of the normalized distribution.

Discussion

Our experiment investigates the impact of structure in visual working memory by analyzing biases in recalling three dots with varying clustering schemes. The structure of working memory affects the way that people encode large amounts of information with limited resources that degrade over time. Encoding structure in working memory would facilitate people to memorize large amounts of information with limited resources and make the information less likely to decode over time. We find attractive biases and repulsive biases and bimodal distributions that support a hierarchical and self-consistent model of visual working memory. We also find that identical stimuli present different recall biases based on a global prior of the average wedge size across. As existing working memory models do not account for this global prior, our results suggest avenues for future work.

Model fitting:

Quantitatively fitting the model was outside the realm of this experiment. Future work could find the best way to quantitatively represent the effect of global priors and find the best parameters to fit this model structure to this data.

Other types of memory: Our study examines visual working memory for three dots in different clustering structures. However, further study of the effect of global priors on remembering other types of stimuli would be a valuable contribution to understanding the other areas to which this effect generalizes. Within spatial relationships, further work could explore more complex structures, including those that involve higher level semantic meaning (i.e. the dots form a shape that is a semantically defined concept).

Orientation: We randomize the cluster orientation along the circle in order to negate the orientation to the recall. However, it has been well established that the orientation of a stimulus affects the sensitivity of recall (De Gardelle et al, 2010; Girshick et al, 2011; Tomassini et al, 2010). Furthermore, the workers may have used some form of the orientation to remember. For example, one worker said in our feedback box that *"I tried to see what quarter of the circles were falling in. That helped me with placement."*

Global priors: We looked at a global prior across all trials of the space between the outer two dots. However, we made all other characteristics across trials identical. Future work can adjust other parameters of our task between two different experimental conditions and see if the attractive bias we saw extends to other task parameters as well.

Remembering multiple dots: We only ask workers to remember one dot. Future work could investigate the effects of a person's mental model of the structure of the stimulus by asking them to guess all three dots.

In sum, we present an experiment that supports the hierarchical self-consistent model of visual working memory. We uncover attractive and repulsive biases in the relative cluster shape of three dots and find attractive biases towards the global prior of presented information. There are many avenues for future work exploring these phenomena in other settings.

Acknowledgements

Thank you to the National Science Foundation of the United States of America (Awards BCS-1350786 and IIS-1912232) for their support of this project.

References

1. Gi-Yeul Bae and Steven J Luck. Interactions between visual working memory representations. *Attention, Perception, & Psychophysics*, 79(8):2376–2395, 2017.

2. Christopher Bates and Robert Jacobs. Efficient data compression leads to categorical bias in perception and perceptual memory. In *CogSci*, pages 1369–1375, 2019.

3. Timothy F Brady and George A Alvarez. Hierarchical encoding in visual working memory: Ensemble statistics bias memory for individual items. *Psychological science*, 22(3):384–392, 2011.

4. Timothy F Brady and George A Alvarez. Contextual effects in visual working memory reveal hierarchically structured memory representations. *Journal of vision*, 15(15):6–6, 2015.

5. Timothy F Brady and Joshua B Tenenbaum. A probabilistic model of visual working memory: Incorporating higher order regularities into working memory capacity estimates. *Psychological review*, 120(1):85, 2013.

6. Vincent De Gardelle, Sid Kouider, and Jerome Sackur. An oblique illusion modulated by visibility: Non-monotonic sensory integration in orientation processing. *Journal of Vision*, 10 (10):6–6, 2010.

7. J. R. de Leeuw. jspsych: A javascript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, 47(1):1–12, 2015.

8. Stephanie Ding, Christopher J Cueva, Misha Tsodyks, and Ning Qian. Visual perception as retrospective bayesian decoding from high-to low-level features. *Proceedings of the National Academy of Sciences*, 114(43):E9115–E9124, 2017.

9. Samuel J Gershman, Joshua B Tenenbaum, and Frank Jäkel. Discovering hierarchical motion structure. *Vision research*, 126:232–241, 2016.

10. Ahna R Girshick, Michael S Landy, and Eero P Simoncelli. Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nature neuroscience*, 14(7):926–932, 2011.

11. Julie D Golomb. Divided spatial attention and feature-mixing errors. *Attention, Perception, & Psychophysics*, 77(8):2562–2569, 2015.

12. Mehrdad Jazayeri and J Anthony Movshon. A new perceptual illusion reveals mechanisms of sensory decoding. *Nature*, 446(7138):912–915, 2007.

13. Ranjay Krishna. Easyturk: A wrapper for custom amt tasks. `https://github.com/ranjaykrishna/easyturk`, 2019.

14. David A Lagnado and David R Shanks. The influence of hierarchy on probability judgment. *Cognition*, 89(2):157–178, 2003.

15. Thomas Langlois, Nori Jacoby, Jordan W Suchow, and Thomas L Griffiths. Uncovering visual priors in spatial memory using serial reproduction. In *CogSci*, 2017.

16. Long Luu and Alan A Stocker. Choice-induced biases in perception. *BioRxiv*, page 043224, 2016.

17. Long Luu and Alan A Stocker. Post-decision biases reveal a self-consistency principle in perceptual inference. *Elife*, 7:e33334, 2018.

18. Long Luu, Cheng Qiu, and Alan A Stocker. High-to low-level decoding does not generally improve perceptual performance. *BioRxiv*, page 240390, 2017.

19. Matthew R Nassar, Julie C Helmers, and Michael J Frank. Chunking as a rational strategy for lossy data compression in visual working memory. *Psychological review*, 125(4):486, 2018.

20. A Emin Orhan and Robert A Jacobs. A probabilistic clustering theory of the organization of visual short-term memory. *Psychological review*, 120(2):297, 2013.

21. Cheng Qiu and Paige Orner. Structural influence in multiple dots recall. 2019.

22. Cheng Qiu and Alan A Stocker. Grouping and segregation in visual working memory. *Journal of Vision*, 20(11):932–932, 2020.

23. Cheng Qiu, Long Luu, and Alan A Stocker. Benefits of commitment in hierarchical inference. *Psychological review*, 127(4):622, 2020.

24. Alan A Stocker and Eero P Simoncelli. A bayesian model of conditioned perception. *Advances in neural information processing systems*, 2007:1409, 2007.

25. Alessandro Tomassini, Michael J Morgan, and Joshua A Solomon. Orientation uncertainty reduces perceived obliquity. *Vision research*, 50(5):541–547, 2010.

26. Xue-Xin Wei and Alan A Stocker. A bayesian observer model constrained by efficient coding can explain'anti-bayesian'percepts. *Nature neuroscience*, 18(10):1509, 2015.

27. Xue-Xin Wei and Alan A Stocker. Lawful relation between perceptual bias and discriminability. *Proceedings of the National Academy of Sciences*, 114(38):10244–10249, 2017.

28. Yair Weiss, Eero P Simoncelli, and Edward H Adelson. Motion illusions as optimal percepts. *Nature neuroscience*, 5(6):598–604, 2002.

29. Mark E. Whiting, Grant Hugh, and Michael S. Bernstein. Fair work: Crowd work minimum wage with one line of code. *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, 7(1):197–206, Oct. 2019. URL `https://ojs.aaai.org/index.php/HCOMP/article/view/5283`.

30. Elisa Zamboni, Timothy Ledgeway, Paul V McGraw, and Denis Schluppeck. Do perceptual biases emerge early or late in visual processing? decision-biases in motion perception. *Proceedings of the Royal Society B: Biological Sciences*, 283(1833):20160263, 2016.

Supplementary

Feedback from Amazon Mechanical Turk study

At the end of our study we provided a textbox where people could optionally provide feedback and describe any strategies they used. Of the 37 participants who responded by describing a strategy, 31 of them stated that they did not use extra strategies beyond focusing and trying to remember the dot placements, sometimes specifically mentioning the fixation dot in the middle of the circle. However, some people talked about using their *"intuition"*, while others created a *"mental image"* visualization of the stimuli: (*"My strategy was to simply try and keep a picture in my mind and then recreate that picture."*). Multiple people also discussed how as they continued with the study they found that their intuition and first instinct seemed to be more correct than when they over-thought their response:

"I didn't use any specific strategy initially. I did notice that as it progressed, I did better making "quicker" decisions. If I thought about it a few extra seconds I would tend to lose my orientation a bit and would be further off the mark. So, my first instinctive guess after the scattered/distracting dots moved off the screen was my best guess usually."

Of the people who used strategies, some looked at the orientation of the dots on the circle overall (e.g. remembering what quarter of the circle they fell in or remembering by imagining what time they fell on in the face of a clock). Others mentioned using the overall structure of the three dots (e.g. remembering the outer two dots, or focusing on the overall curve among the three dots).

Although most of the participants who responded to our query were not consciously aware of any strategies they used, the repulsion and attraction biases are still apparent in their work as a result of the unconscious way their working memory encoded the dots.

Overlaid raw data

Figure 12: When normalized by wedge size, the raw data of clicked angles follows similar distributions.

Figure 13: The same data as Figure 12, highlighting the conditions with 80 degrees.

Results when normalized

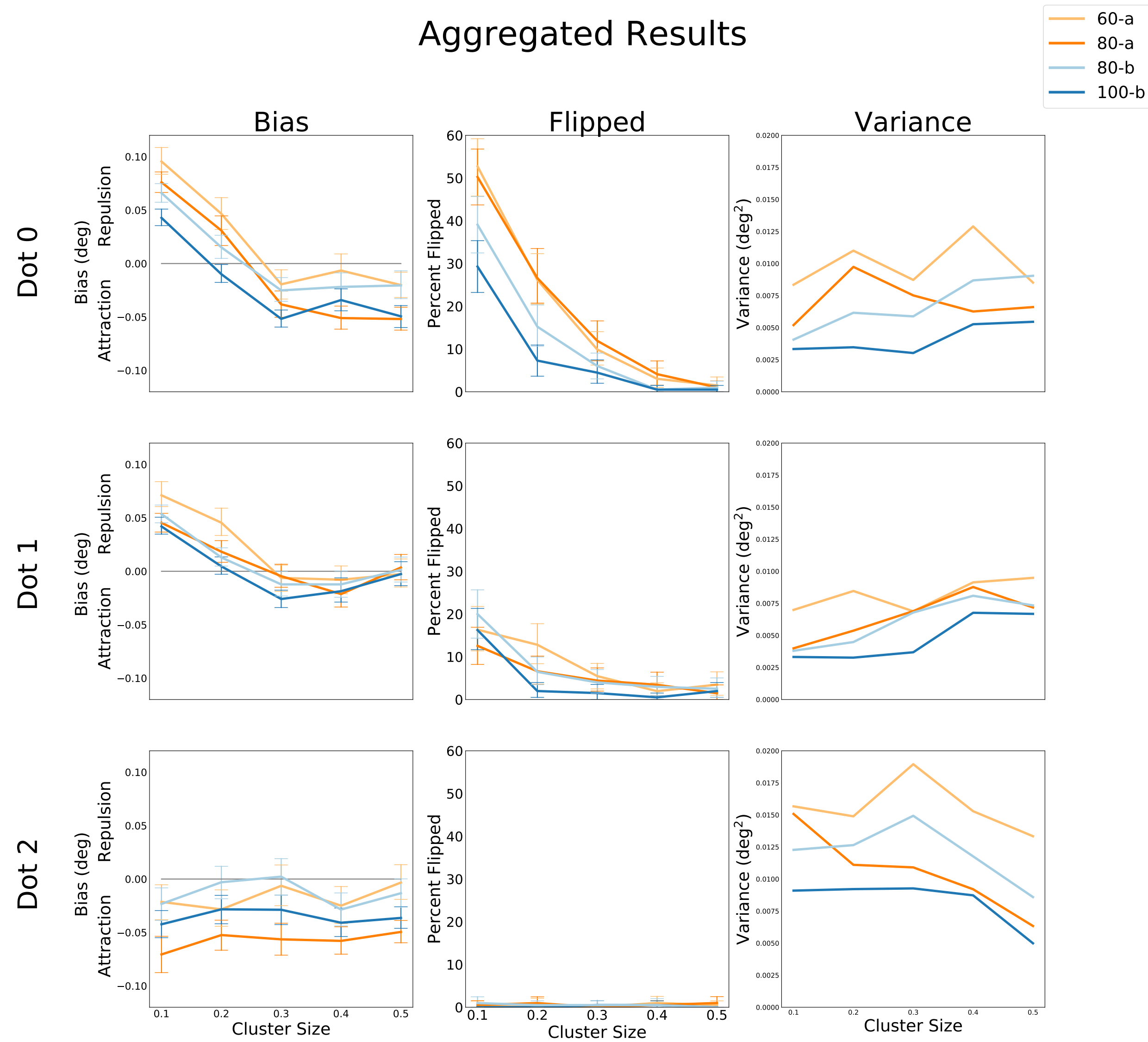

Figure 14: Here we summarize the bias, flipping percentage, and variance. The participant's responses were normalized by wedge size before bias and variance were calculated.

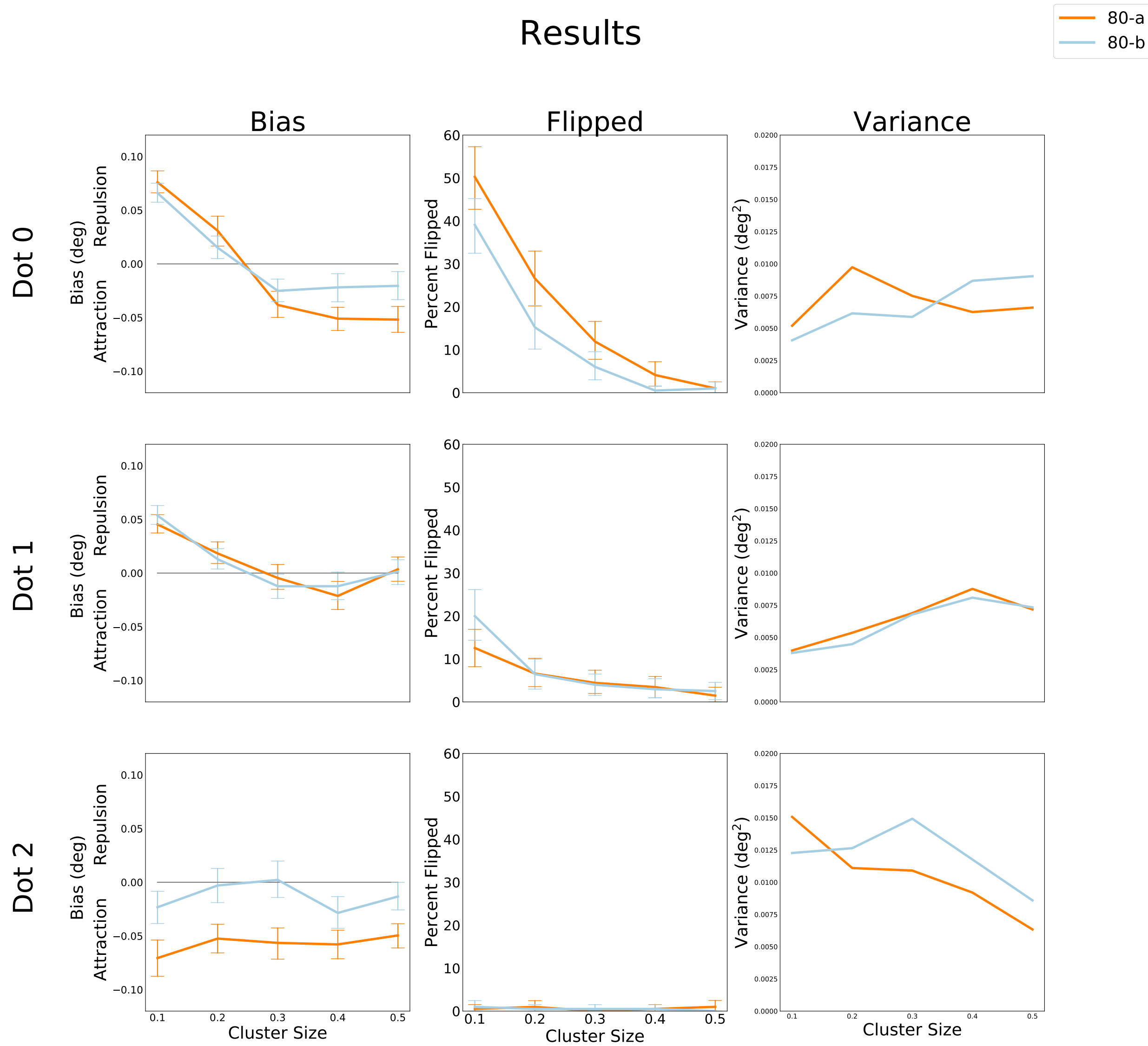

Figure 15: Here we summarize the bias, flipping percentage, and variance for the two 80 degree conditions. The participant's responses were normalized by wedge size before bias and variance were calculated.

Results from bias aggregation without flipping adjustment

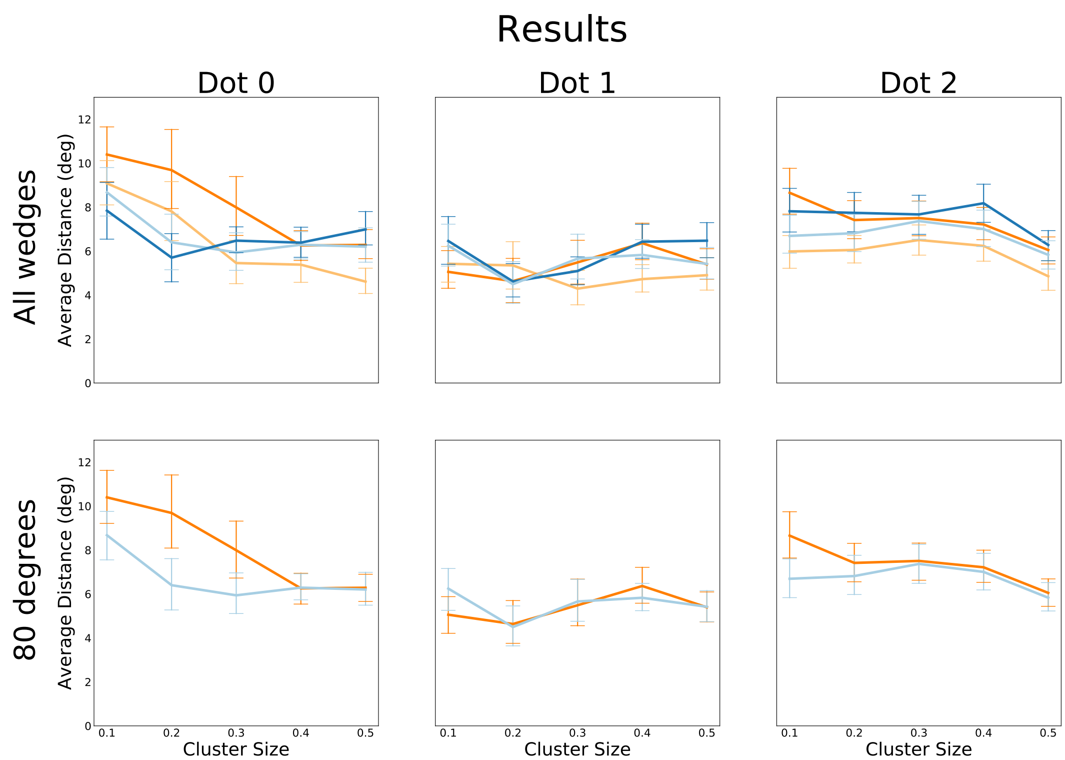

Figure 16: Here we show the average bias of the absolute distance between the guessed location and the true location.

Variance of raw data

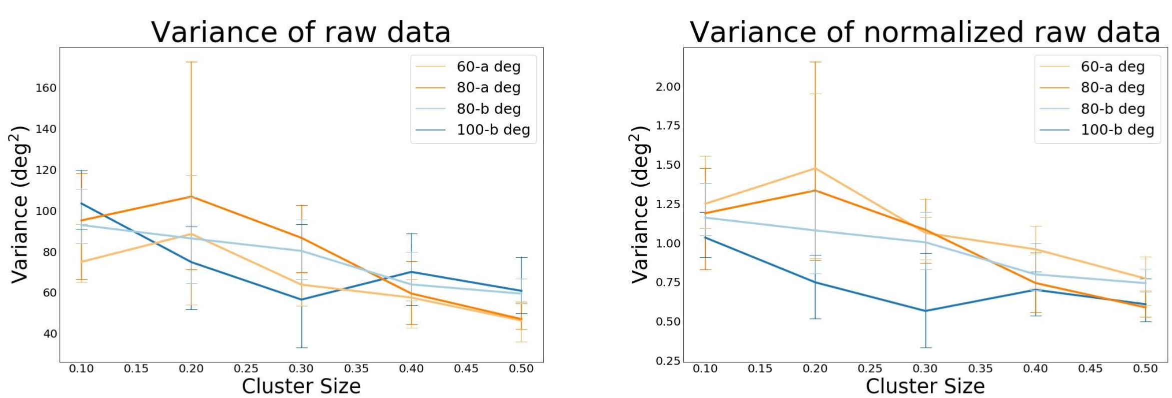

Figure 17: The variance of the raw data (unaggregated for flipping errors). Normalized by wedge size on the right figure.