UNIVERSITY OF CALIFORNIA
RIVERSIDE

Reevaluation of Formal Model Comparison Between Slot and Resource Models of Visual Working Memory

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Psychology

by

Marcus John Cappiello

December 2019

Dissertation Committee:
Dr. Weiwei Zhang, Chairperson
Dr. Aaron Seitz
Dr. Weixin Yao

The Dissertation of Marcus John Cappiello name is approved:

Committee Chairperson

University of California, Riverside

ABSTRACT OF THE DISSERTATION

Reevaluation of Formal Model Comparison Between Slot and Resource Models of Visual Working Memory

by

Marcus John Cappiello

Doctor of Philosophy, Graduate Program in Psychology
University of California, Riverside, December 2019
Dr. Weiwei Zhang, Chairperson

Visual working memory actively maintains information over brief periods in service of other mental activities. Unlike other memory systems, visual working memory is highly limited in the amount of information that can be retained, however the nature of this limitation is still widely debated. Historically, research on working memory limitation was focused on the number of items that can be simultaneously maintained, but recently limitations in the precision of working memory representations has also been explored. Two theoretical models of working memory limitations that stem from the interest in precision limitations are investigated in depth. The slot model theorizes a limited number of slots for working memory representations. When there are more items than there are slots, then some items will be forgotten entirely resulting in a capacity limit. Conversely, the resource model theorizes all items are retained in visual working memory, and any behavioral limitations stem from low precision of memory representations. That is, high error responses are accounted for by a capacity limit in the slot model, and accounted for by low precision representations in the resource model. Previous formal model comparison between the measurement models of these two theoretical models has found support for the resource model, however important aspects have been overlooked. Support for the slot model is found in the current dissertation by expanding on the previous formal model comparison in four ways. First, a new model comparison analysis is created that includes the flexibility of model parameters; the inclusion of which shows support for the slot model at high set sizes when the capacity limit is exceeded. Second, using the measurement model for the resource model, a capacity limit is still observed. Third, using a non-parametric model fitting approach that does not assume any model, a capacity limitation is again found. Finally, a new measurement model is created that better matches the slot theoretical model. This new model outperforms the variable precision model showing further support for the existence of a capacity limit in visual working memory.

Table of Contents

List of Figures

List of Tables

Chapter 1

1.0. Introduction

If you were asked about your childhood, you undoubtedly currently store thousands of life experiences that you could report. If instead you were asked to remember eight colors for a few seconds, you would only perform well on three or four. This brief, limited memory system, often called working memory (WM; Baddeley & Hitch, 1974), is involved in nearly every cognitive task we perform, from forming full sentences when we speak to deciding what to have for lunch today. The extent of the limitation of WM is a strong predictor of individual differences such as spatial reasoning and fluid intelligence (Kane et al., 2004; Oberauer, Süß, Wilhelm, & Sander, 2008). In addition, deficits in the system relate to neuropathologies such as depression (Joormann & Gotlib, 2008) and schizophrenia (Park & Holzman, 1992). As such, the mechanism behind the limitation of working memory is of vital importance and is currently one of the primary cognitive psychology research topics (Nee & Jonides, 2008).

The finding of a limited capacity of memory items over short delays (Miller, 1956) coupled with the large capacity found for long-term memory (LTM; Standing, 1973) show support for the classic modal model of memory (Shiffrin & Atkinson, 1967). In this model, memory is conceptualized as three separate systems, often called the *system* view of memory. Sensory memory is a high capacity fragile state that resembles residual perception after the offset of presented stimulus. Unless attended to, these items are erased when new information is presented such as what happens after eye movements (Cappiello & Zhang, 2016; Sperling, 1960). The attended items are then put into the limited capacity working memory store. These representations are robust to distractors and eye movements and are stored actively allowing for responses to be made. According to the modal model, representation in WM may be transferred into a large capacity passive long-term memory system following repeated rehearsal. Items in LTM can be retrieved back into working memory and used for performance. The classic neurological evidence for the separation between the WM and LTM systems is patient H.M., where damage to the medial temporal lobe (MTL) lead to a deficit in LTM creation but not WM (Scoville & Milner, 1957). Since then, the MTL deficit finding has been replicated (Simons & Spiers, 2003), and a deficit to both WM and LTM has been found due to lesions in the prefrontal cortex (Voytek & Knight, 2010).

However, recent literature has found MTL involvement in WM tasks in amnesia case studies (Hannula, Tranel, & Cohen, 2006) and neuroimaging (Libby, Hannula, & Ranganath, 2014), challenging the system view of memory. Instead of the *system* view of WM, *state* models have gained popularity, which classify the different types of memory by their level of activation rather than separate systems. Two such competing models are the embedded process model (Cowan, 2001) and the three layer model (Oberauer, 2002). In these models, memory is all one storehouse and the differences seen between the systems predicted by the modal model are due to levels of activation. In the embedded process model, memory representations are stored as passive long-term memories, active long-term memories, or in the focus of attention. The focus of attention is capacity limited (~four items) and relate to the working memory store from the modal model. Memories that are related to the current goal, have been retrieved recently, or have been created recently, are stored as active long-term memory. This store is largely capacity unlimited, and these representations can be quickly activated through shifting the focus of attention. All other memories are stored in the passive long-term memory store that resembles the long-term memory system in the modal model. The three-layer model is similar to the embedded process model, with the major difference being a capacity of one the focus of attention rather than four. With the advent of neuroimaging, neurological support for the state view of memory has been found where MTL activation is seen in WM experiments.

The current dissertation will not focus on the system versus state debate, but rather focus on the WM mechanism. Both system and state models agree on three major aspects that define the WM system. First, due to the active nature of WM processes and the inclusion of both bottom-up and top-down information, WM is unique in its support of a wide range of other cognitive functions (Baddeley, 2012) such as attention (Kane, Poole, Tuholski, & Engle, 2006), fluid intelligence (Conway, Kane, & Engle, 2003), and emotion processing (Xie et al., 2017). Second, while LTM can store memories for up to a lifetime (Standing, 1973), WM representations decays only after a few seconds if no rehearsal tactic is implemented (Zhang & Luck, 2009). Third, and most importantly for the current dissertation, WM is limited in the amount of information it is able to store (Cowan, 2001; Miller, 1956), although the nature of this limitation is widely debated (Bays & Husain, 2008; van den Berg, Shin, Chou, George, & Ma, 2012; Zhang & Luck, 2008).

1.1. Slot and resource model

Much of the past work investigating working memory limitations have focused on a limited quantity of items stored (Cowan, 2001; Koriat & Goldsmith, 1996; Miller, 1956; 1965; Shiffrin & Atkinson, 1967; Standing, 1973). This ‘storehouse’ view of working memory conceptualizes working memory as discrete units of information. It is now common to also investigate the variability of the working memory representations, or the correspondence between the internal representation and the corresponding external stimulus (Koriat & Goldsmith, 1996). Although memory variability is closely related with successful retrieval, recent advances in methodology have allowed both the number and precision of WM representations to be independently measured. These advances have sparked several new categories of models that are a topic of active debate, two of which will be investigated here.

Two prominent models of the working memory mechanism are the slot model (Zhang & Luck, 2008) and the resource model (particularly the variable precision model variant of the resource model; van den Berg et al., 2012). Both models focus on the nature of working memory limitations, such as whether there is a stark capacity limit or just a degradation in precision, but make no assumptions on whether working memory is attention-based (embedded process model) or not (modal model).

One of the first models to include the variability of VWM representations is the slot model, which predicts that there are a limited number of ‘slots’ that can be filled in working memory (Cowan, 2001; Luck & Vogel, 1997; Oberauer, 2002; Sperling, 1960; Zhang & Luck, 2008). Once the slots are filled, any additional memory items will be discarded manifesting as a capacity limitation. When the number of items is lower than the number of slots, the slots can be averaged to increase quality of the memory items (slots plus averaging, Zhang & Luck, 2008). The slot model has been very successful at describing visual working memory performance, such as visual working memory retention over time (Zhang & Luck, 2009), incentivized changes in visual working memory (Zhang & Luck, 2011), dual visual working memory and attention paradigms (Zhang & Luck, 2015), and the interaction between emotion and visual working memory (Xie & Zhang, 2016).

An alternative model, the resource model, instead predicts a resource pool that can be flexibly allocated to all memory items (Bays & Husain, 2008; Wilken & Ma, 2004). For all resource models, there is no capacity limit. Rather, the observed working memory limitation at high set size originates from a decreased precision of the memory representations. That is, all presented items are remembered, but some of them will be very low quality. A recent updated version of the resource model, called the variable precision model, proposes that each memory item receives a variable amount of resource resulting in variable precision of memory representations (van den Berg et al., 2012; van den Berg, Awh, & Ma, 2014). Like the slot model, VP does not specify whether this resource limitation stems from attention, or some other system.

The recent debate between SLOT and VP has primarily focused on formal model comparison e.g. (van den Berg et al., 2012; 2014), which will be addressed in detail here. Using model comparison, we will focus on the core difference between the models – is there a true capacity limit in working memory? The slot model predicts a capacity limit originating from the limited number of slots. The resource models predict no true capacity limit, rather all limitations stem from variability of memory representation precision. Investigation into these differences provides an opportunity to deepen our understanding of the working memory system.

1.2. Theoretical and measurement models

The focus on formal model comparison for model investigation requires us to be specific on the type of models used. All psychological models can be broken down into two types: theoretical models and measurement models (Roberts & Pashler, 2000). Theoretical models (also called abstract or explanatory models) are general principal models that seek to explain theories at a higher level and can be used for all experimental conditions. All models discussed above, including the slot and resource models, are therefore considered theoretical models. Measurement models are mathematical models that can be used to fit empirical data of interest to measure latent variables (such as how many items can be held in memory). They can vary according to the data structure and experimental paradigms (e.g., recognition versus recall,
Xie & Zhang, 2017; Zhang, 2007). For each theoretical model, a measurement model can be created tailored to each experiment allowing for a formal mathematical model comparison approach to support or reject each theory. Doing so forces the researcher to be very specific about each aspect of the theoretical model. For example, if there is variability in internal WM representations, what does that variability look like? Is it normally distributed or take on another form? As will be explained in further detail

below, slight mistakes in the creation of a measurement models has led to gross misinterpretations of model comparison results.

It is important to note that all models of the brain are wrong (Box, 1976). To have a ‘right’ model, we would eventually need to describe every neuron interaction in each individual brain, which would not lead to any insights into the nature of cognition and is currently computationally unfeasible. Instead, these theoretical and measurement models seek to describe higher-level rules that the brain operates on.

1.3. Dissertation overview

Support for resource models over slot models has been found previously in formal model comparison, including goodness of fit and residual patterns (van den Berg et al., 2012; van den Berg & Ma, 2014). However, upon closer inspection the support found does not stem from theoretical differences between the models, but rather issues in formal model comparison methods and measurement model creation. The current dissertation first develops a new formal model comparison method that more accurately addresses the differences between the theoretical models (Chapter 2). Next, a comparison between the slot and resource measurement models is done using a variety of methods including the new formal model comparison method (Chapter 3). Finally, an entirely new measurement model for slot is created that better represents the theoretical model and a final formal model comparison is performed using the new measurement model (Chapter 4). All together the dissertation will largely be focused on primary difference between the slot and resource models: is there a capacity limit in WM.

Chapter 2

2.0. Delayed-estimation continuous recall task

Measurement models for the SLOT and VP model have been created previously to describe delayed-estimation continuous report working memory data for simple circular features (Wilken & Ma, 2004; Zhang & Luck, 2008). In these tasks, participants see an array of to-be-remembered items, such as colors, lines of different orientations, or shapes, following a probe that indicates which memory item to be reported (Figure 1). The participants select the color, orientation, or shape that best matches their memory of the probed item from a continuous spectrum (e.g. from a color wheel).

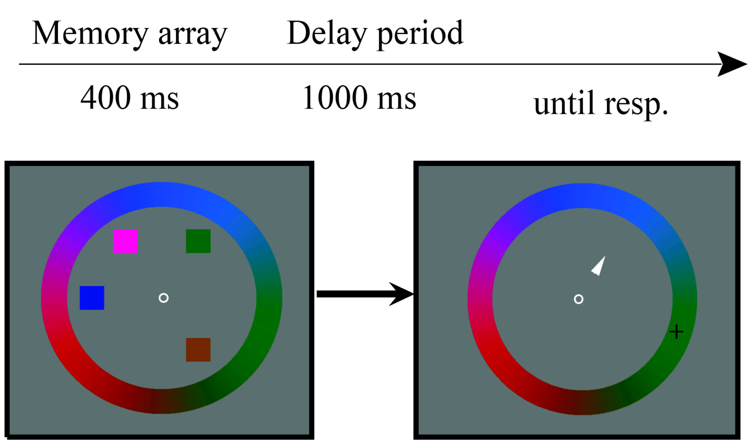

Figure 1. Continuous-recall VWM experiment using color.

Error is calculated as the distance in degrees between the correct answer and the estimated response. Over many trials, a bell shaped curve is observed centered around zero degrees error. Much of the debate between slot and resource models has centered around describing these error distributions (e.g. van den Berg et al., 2012; 2014; van den

Berg & Ma, 2014; Zhang & Luck, 2008). For the remainder of this discussion, color will be discussed as the tested feature for clarity. However all principles described below can be extended to other continuous feature dimensions such as orientation and shape.

2.1. Measurement models of VWM

The measurement model for the slot model stems from understanding what participants will do given different circumstances. If the probed color at test is not contained in WM, then participants will respond by clicking randomly on the color wheel (forced choice). Because the wheel is rotated randomly on each trial, this will manifest as a random distribution of responses, or a uniform distribution, across all errors. On the other hand, if the probed memory item is contained in WM, then their responses should cluster around zero degrees error. The spread of this bell-shaped distribution will rely on participants' internal WM representation quality. While the representation quality will likely vary trial by trial, the *mean* internal representation quality can be found by the width of this bell-shaped distribution. For ease of computation, a normal distribution can therefore be used to account for all in-memory responses. Together, the measurement model for the SLOT consists of a mixture model of a normal distribution centered at zero and a uniform distribution (Zhang & Luck, 2008).

On the other hand, there is no capacity limit in the resource model, and therefore no uniform distribution is included in the measurement model (variable precision model, VP). Instead, high error responses are accounted for by very low-quality representations. Here, all items are remembered, but with varying levels of quality that all come from the same resource pool. To account for this, the VP is a mixture model of many normal distributions whose precision is sampled from a continuous gamma distribution over precision (related to SD, see below). VP also has two free parameters, the shape and scale of the gamma distribution. For high error responses, the gamma distribution over precision can (and often does) cover a large range of values for precision, making this model highly flexible. As such, VP has found some success in model comparison (van den Berg et al., 2012; 2014).

The continuous report task used currently results in empirical data in circular space (i.e. 20° error on one trial). To formalize the measurement models for this task, all models must therefore be converted to circular space. A Gaussian analog in circular space is the von Mises distribution, described as:

$$p(x \mid \mu) = \frac{e^{k\cos(x-\mu)}}{2\pi I_0(\kappa)} = VM(x;\mu,\kappa)$$

where μ is the mean, I_0 is the Bessel function of the first kind, and κ is the concentration parameter that controls the width of the distribution. Using the von Mises distribution, SLOT is formalized as:

$$p(x \mid \mu) = (1-g) * VM(x;\mu,\kappa) + g\frac{1}{2\pi}$$

Unless experimental manipulation is introduced, the normal distribution is always centered on zero degrees error, removing the need for μ as a free parameter. SLOT model therefore only has two free parameters: *SD* (a linear analog of *SD* can be calculated from κ and will be used here for clarity) and the height of the uniform distribution (*g*). Given the set size (SS) of the to-be-remembered stimuli, the number of items each participant is able to hold in WM, or their capacity at high memory set sizes, can be calculated as (1-*g*)*SS. The VP model is formalized as:

$$p(x \mid \mu) = \sum VM(x; \mu, k)dk \text{ where } k \sim \Gamma(modek, sdk)$$

where k is precision ($k \sim 1/SD^2$), which can be used to calculate κ for the von Mises distribution (often using a look-up table; Suchow, Brady, Fougnie, & Alvarez, 2013; van den Berg et al., 2014), and Γ is a gamma function characterized by mode and standard deviation of the precision (*modek* and *sdk*, respectively). Mode and SD of the precision is used instead of the more traditional shape and scale parameters due to their ease of estimation and interpretability. VP therefore also has two free parameters – *modek* and *sdk*. It is important to note that these parameters are in a higher order space than *g* and *SD* from SLOT, which leads to model comparison issues discussed below.

2.2. Simulations

To determine which measurement model best fits empirical data, a plethora of different fitting techniques have been used previously, such as least-squares, maximum-likelihood estimation (MLE), Bayesian hierarchical modeling, conversion to Fischers J with simplex (van den Berg & Ma, 2014). By testing the validity of each model comparison technique, MLE is found to be ideal for our current model comparison, where least-squares and MLE give nearly identical results, yet MLE is used more commonly in the literature and will therefore be used here. Bayesian hierarchical model comparison is computationally intensive and does not produce better results than MLE except when the trial number per condition is very low (which will not be true here). Finally, MLE does a far better job at parameter estimation than conversion to Fischers J with simplex, which is important when understanding the psychological meaning behind each parameter.

Before any formal model comparison can begin, it is crucial to test the reliability of the fitting technique. To do so, data will be simulated for both the SLOT and VP model using a wide variety of parameters. Then, a formal model comparison will be performed by comparing the log-likelihood estimation of the fits of each model to the simulated data. If the data is simulated using the SLOT model, the formal model comparison should choose SLOT over VP, and vise versa. If the formal model comparison is unable to make that distinction reliably, then the model comparison tactic must be re-evaluated.

2.2.1. Simulations with MLE

The first set of simulations used the following procedure:

1. Fit empirical data for all set sizes using MLE with both SLOT and VP to get parameters (see *Chapter 3* for data summary).
2. Simulate data with the number of trials and parameters from step 1 for both SLOT and VP
   - Parameters are idealized versions and representative of the parameters from data fits (Table 1)
3. Fit all datasets with SLOT and VP using MLE and compare log-likelihoods.
4. Repeat steps 2 and 3 (100 runs used here)
5. Determine how frequently the model comparison method chose the correct model across all repetitions.

Table 1

Parameters used for MLE simulations by set size mimicking empirical fits.

Set Size	SLOT Pars	VP Pars
1	*g*=0, *SD*=10	*modek*=0.007 , *sdk*=0.005
2	*g*=0, *SD*=20	*modek* =0.005 , *sdk* =0.005
3	*g*=0, *SD*=25	*modek* =0.003 , *sdk* =0.006
4	*g*=0.25, *SD*=25	*modek* =0.002 , *sdk* =0.006
5	*g*=0.4, *SD*=25	*modek* =0.001 , *sdk* =0.007
6	*g*=0.5, *SD*=30	*modek* =0.0008 , *sdk* =0.007
7	*g*=0.57, *SD*=30	*modek* =0.0006 , *sdk* =0.008
8	*g*=63, *SD*=30	*modek* =0.0004 , *sdk* =0.009

The simulation results can be seen in Table 2. Overall, MLE formal model comparison is able to detect the correct mode 78.5% of the time. This varies by model used, where MLE chose SLOT model correctly less often than VP model ($Mean_{SLOT}$=73.3%, SD_{SLOT}=12.7%, $Range_{SLOT}$=[54, 100]; $Mean_{VP}$=83.7%, SD_{VP}=11.0%, $Range_{VP}$=[64, 100]). This stark difference suggests a bias toward VP in model fits, which will be discussed in detail below.

Table 2

Model recovery results using MLE

	Set Size	1	2	3	4	5	6	7	8
Number Trials	Model								
110	SLOT	60	60	69	77	83	82	72	84
	VP	90	86	86	75	78	67	69	70
120	SLOT	58	63	61	78	79	74	72	74
	VP	87	86	84	75	76	75	64	72
130	SLOT	63	67	71	66	88	73	68	77
	VP	93	87	84	72	72	76	69	70
150	SLOT	66	55	55	76	86	60	79	75
	VP	92	86	88	85	67	74	78	68
320	SLOT	61	54	62	88	88	86	78	86
	VP	98	97	99	95	87	82	90	83
800	SLOT	55	58	55	99	100	91	95	90
	VP	100	100	99	99	97	96	98	97

Note. Numbers indicate percent correct model choice.

As seen in Table 2, these values also change depending on number of trials and parameters used as expected. The change in results over parameters used is different for each model, with SLOT doing well at high set size parameters and VP doing well at low set size parameters. This is to be expected. The uniform distribution in SLOT is only useful when participants cannot remember all colors, or at large set sizes. At low set sizes, the free parameter *g* accounts for a very small proportion of trials (likely lapses in attention). Interestingly, the advantage of SLOT peaks for parameters around set size 5, then decrease after. This will be discussed in further detail below. With an overall bias toward VP in model fits, the model fit procedure must be updated before these measurement models can be used to support slot or resource theoretical models.

2.2.2. Model Flexibility

One possibility for the observed bias for VP is a difference in model flexibility between SLOT and VP, or the amount of the data space each model can account for. For example, assume that we find a better fit for Model 1 over Model 2. It is temping to assume this means that Model 1 comes from a theoretical perspective that aligns better with the psychological system than Model 2, but there are two alternative possibilities. First, Model 1 may have more free parameters than Model 2, which is commonly accounted for by adding a penalty parameter to the log likelihood (e.g., AIC; Akaike, 1973). This is not currently an issue because both models contain two free parameters, but will be important for *Chapter 4*. Second, the free parameters for Model 1 may be more flexible, or may be able to account for more of the data space, than Model 2. Model flexibility is particularly important currently because the free parameters of the SLOT and VP models reside in different order spaces.

Although model flexibility has not been implemented in the slot versus resource model debate, it has gained popularity in psychological research largely due to Roberts and Pashler's article on the persuasiveness of a good fit (Roberts & Pashler, 2000). They report thousands of articles that have used a 'good fit' to support or reject theoretical models in psychological science, none of which considered the flexibility of model parameters. A key factor of their suggested fix for the issue is to keep models as parsimonious as possible and to account for model flexibility in all fits. The typical definition of a parsimonious model is one that gives a desired level of explanation with as few predictor variables as possible. It is now necessary to extend this idea to finding a model that gives a desired level of explanation with as little model flexibility as possible. After Roberts and Pashler's article, model flexibility has been successfully used in many areas of research (e.g. perception, LTM; Jang, Wixted, & Huber, 2007) but has yet to make its way to the slot versus resource debate.

There are several model flexibility analyses available, and it is important to choose one that works well for the empirical question (Veksler, Myers, & Gluck, 2015). For example, parameter space partitioning separates the data space into regions, and determines how many regions each model can account for. While this results in a qualitative understanding of the flexibility of the models, it does not propose a way to penalize the more flexible model. An alternative approach, minimum description length, finds an appropriate quantitative penalty term by simulating all possible datasets and fitting all models to be compared. While this approach would be useful for our current comparison, it is not feasible within a reasonable time frame to simulate all possible datasets (estimated to take ~1 year without a super computer).

Here, a model flexibility analysis is developed that will produce a penalty term to more flexible models and be computationally feasible allowing other researchers to use the technique. The analysis developed here would be classified as a model mimicry analysis, which investigates how each model being compared can account for each other's predictions (Veksler et al., 2015). That is, if Model 1 can fit more of Model 2's predictions than vise versa, then it is a more flexible model. Of the various model mimicry approaches, one parametric bootstrap approach proposed by Wagenmakers *et al*. allows for a rank order flexibility analysis that gives a penalty term for the more flexible model (Wagenmakers, Ratcliff, Gomez, & Iverson, 2004). The current approach aligns closely with this proposal, with some slight modifications to match the current datasets.

In the following explanation, SLOT and VP will be used as examples, but this method will work for any measurement models. The overall approach is to simulate data using the models in question from a parameter range observed in empirical datasets (rather than the entire dataspace as in minimum description length), and then find how well each model fits all simulated datasets. The procedure is as follows:

1. Fit participant data with SLOT and VP and extract parameters.
2. Using one parameter set, simulate data for both SLOT (D_{SLOT}) and VP (D_{VP}).
3. Fit D_{SLOT} and D_{VP} with both SLOT and VP. Ideally, SLOT will fit D_{SLOT} better than VP and vice versa.

4. Determine which model fits better for the fits to both (D_{SLOT}) and (D_{VP}) by finding the difference in AIC. Here, it is important to keep the order of the model fits the same (e.g. always SLOT-VP rather than VP-SLOT). To simplify the notation, the following will be used for the rest of the analysis (GOF will be AIC here):
   - $\Delta GOF_S = GOF_{SLOT|DSLOT} - GOF_{VP|DSLOT}$
   - $\Delta GOF_V = GOF_{SLOT|DVP} - GOF_{VPDVP}$
   - Note: A lower AIC indicates a better fit. Therefore, if the models are equally flexible, then ΔGOF_S should always be negative and ΔGOF_V should always be positive.
5. Repeat this process M for the same parameter set (M=500 currently).
   - Due to the random sampling of simulated datasets, ΔGOF_S and ΔGOF_V may be different for each run using the same parameter set.
6. Find the distribution of values for ΔGOF_s and ΔGOF_V.
7. Find the optimal criterion, or the ΔGOF value that optimizes the choice between the two models given ΔGOF_S and ΔGOF_V.
   - The optimal criterion is the point at which you choose the correct model the maximum number of times across all fits. If the models are equally flexible, the criterion will be zero.
8. Calculate bias as the difference between the criterion and zero (β).
9. Include β in the model comparison results.
   - $\Delta AIC_F = \Delta AIC + \beta$

Again, model mimicry is attempting to find how well a model can account for another models predictions. If data is simulated with SLOT, that data is what SLOT predicts. When SLOT is fit to that simulated data, it will be a perfect fit (with some residual due to noise). If instead VP is fit to the same simulated data, that fit is indicative of how much of SLOT model predictions can be accounted for by VP. As an extreme example, if VP fits just as well as SLOT to data simulated by SLOT, then the measurement model VP is so flexible that it accounts for both models at once and is not useful to support or reject either model. Here the difference between model fits will be zero. As this difference becomes larger, the better the measurement models are able to determine which model is closer to the underlying truth. Due to the number of data points simulated, parameter sets used for simulation, and random variability from the random sampling, this difference term, ΔGOF_S, will change for each new simulation. Now, still using a simulated dataset from SLOT as an example, when we plot the difference between the fits of SLOT and VP over M simulations (steps 5 and 6), the resulting distribution shows how much of SLOT model predictions can be accounted for by VP. If the peak of the distributions lies at zero, this suggests that VP can account for SLOT predictions as well as SLOT can. As the peak shifts more negative (assuming $\Delta GOF_S = GOF_{SLOT|DSLOT} - GOF_{VP|DSLOT}$), the worse VP is at accounting for SLOT predictions. Then the same procedure is done with data simulated with VP to find how much of VP predictions SLOT can account for. As the peak of this distribution becomes more positive (assuming $\Delta GOF_V = GOF_{SLOT|DVP} - GOF_{VPDVP}$), the worse SLOT is at accounting for VP predictions.

With these two distributions, it is now possible to determine if there is a bias toward one model or the other. Using the same extreme example as above, say VP is able to account for SLOT predictions as well as SLOT, however SLOT is not able to account for VP predictions. Here, the peak for ΔGOF_S will lie at zero, and the peak for ΔGOF_V will be positive. If these distributions are used to determine when SLOT or VP is supported, VP will be supported much more often than SLOT, which in classic model fitting would be the end of the SLOT model. However, in this example, VP is so flexible that it cannot tell between data simulated from SLOT or VP, giving us no indication of which model is closer to the truth. If instead, the models were equally flexible and only accounted for predictions from their own model, then the peaks would be equally spaced across zero. That is, when data simulated by SLOT is fit, SLOT fits well and VP does not, and vise versa. To quantify any bias (β) between the model flexibility, the point at which the correct model is chosen the maximum amount of the time is calculated (often called the optimal criterion is signal detection theory). If β is zero, the models are equally flexible. If β is non-zero, then one of the two models is accounting for more of the data space than the other. If positive, VP is accounting for D_{SLOT} better than SLOT is accounting for D_{VP} and vice versa. Since β is in the same units as GOF (AIC in this case), it can simply be added to the resulting fit (Step 9) to obtain ΔAIC_F – the model fit statistic that includes model flexibility. Using ΔAIC_F, it is now possible to find which measurement model best fits the data and, by extension, which theoretical perspective is closer to the underlying truth.

2.2.3. Simulations with model flexibility

The validity of the new model comparison method designed in 2.2.1 is explored using the same simulation results from 2.2.1. A bias term is calculated separately for each trial number and each parameter set, the results of which can be seen in Table 3 (Mean=0.45, SD=0.57, Range=[-0.77, 2.01]).

Table 3

Model flexibility bias.

Set Size	1	2	3	4	5	6	7	8
Number of Trials								
110	0.67	0.90	0.67	0.09	-0.47	0.14	0.15	0.15
120	0.84	0.75	0.93	0.15	-0.36	0.19	0.15	0.13
130	1.00	0.68	0.63	0.24	-0.40	0.26	0.34	0.11
150	1.02	1.11	0.78	0.00	-0.44	0.13	0.02	0.16
320	1.14	1.40	0.78	0.34	-0.59	0.39	0.39	0.24
800	2.02	1.65	1.13	0.04	-0.77	0.96	0.90	0.75

Note. Values are in log likelihood units, and positive indicates a bias toward VP over SLOT.

For the majority of trial numbers and parameter sets, β is positive as is expected due to the original discrepancy found between SLOT and VP fits on simulated data in 2.2.1. The change in β across set sizes also matches the original simulation results (Table 2) where the largest β value is found when MLE could recover VP better than SLOT (set size 1) and the lowest β value is found when MLE could recover SLOT better than VP (set size 5). Here, the bias term is added to the results of the simulations from 2.2.1., the results of which can be seen in Table 4 ($Mean_{SLOT}$=85.3%, SD_{SLOT}=8.1%, $Range_{SLOT}$=[65, 99]; $Mean_{VP}$=80.5%, SD_{VP}=11.3%, $Range_{VP}$=[63, 100]).

Table 4

Model recovery results using MLE with model flexibility.

	Set Size	1	2	3	4	5	6	7	8
Number Trials	Model								
110	SLOT	83	94	85	77	75	84	80	85
	VP	71	73	80	73	86	64	63	70
120	SLOT	89	91	91	79	76	75	76	78
	VP	72	83	74	71	79	68	64	71
130	SLOT	93	90	91	69	86	78	73	79
	VP	80	84	81	70	81	68	65	65
150	SLOT	93	87	88	76	81	65	79	78
	VP	76	82	86	85	72	74	78	66
320	SLOT	90	93	88	91	87	87	85	87
	VP	90	95	96	95	90	79	83	82
800	SLOT	97	94	88	99	99	95	98	91
	VP	100	100	99	99	97	95	95	95

Note: Numbers indicate percent correct model choice.

Overall, the inclusion of β in simulation results improved all fit recovery by 4.5% from 78.5% to 83%. The model recovery for SLOT improved by 12% and the model recovery for VP decreased only 3.2%, supporting the use of model flexibility in the formal model fitting moving forward.

2.3. Discussion

It is common to assume the literature standards for model fitting are giving results that can be used for theoretical support without checking. Here, it is now obvious that simulations are necessary before any formal model comparisons can be performed. Any differences in data type and model parameters may alter the ability of standard model comparison methods such as MLE to detect differences in theoretical models. Here, the difference in the order of space between SLOT and VP (with VP having free parameters in higher-order space) lead to a large difference in flexibility between the models. Simulations show this leads to an overall bias toward VP that is not indicative of any differences between slot and resource theoretical perspectives, but rather a difference in choice in measurement models. Once the penalty term for flexibility is added, not only is the overall performance of the model comparison method improved, but the bias toward VP is erased allowing for a true model comparison method. It is likely that the strong support found for the VP model over SLOT model found in the past e.g., (van den Berg et al., 2014) may no longer hold once flexibility is included.

Chapter 3

3.0. Empirical Data

A dataset consisting of empirical data from continuous recall VWM experiments has been compiled previously (van den Berg et al., 2014) which will be useful currently to avoid replication issues. Ten datasets in total were used in the original dataset. Two of the papers (Anderson & Awh, 2012; Anderson, Vogel, & Awh, 2011) from which the data was gathered have since been retracted, and will therefore not be included here. One additional dataset was collected personally at UC Riverside, and the remaining 8 datasets were then scrutinized under the following two exclusion criteria. First, for each experiment, model comparison recovery and model parameter recovery using the number of trials included in the study must fall above 70%. Second, for each participant's data, the data will not be included if, when fit with SLOT, g is above 0.8. Such a large guessing rate often shows clear misfits for both SLOT and VP. For the first criterion, two additional experiments were removed due to a low trial number (108 trials). For the second criterion, 10 participants were removed for low performance. All together, the data used currently can be viewed in Table 5.

Table 5

Data summary

Exp	Article	Subjects	Feature	Set Size	Trials/SS	Eccentricity (Degrees)	Stimulus Duration (ms)	Delay
1	(Wilken & Ma, 2004)	15	Color	1,2,4,8	128	7.2	100	1500
2	(Zhang & Luck, 2008)	8	Color	1,2,3,6	125	4.5	100	900
3	Bayes, 2009	12	Color	1,2,4,6	150	4.5	100, 500, 2000	900
9	(van den Berg et al., 2012)	6	Orientation	1-8	320	8.2	110	1000
10	(Rademaker, Tredway, & Tong, 2012)	6	Orientation	3,6	800	4.0	200	3000
11	N/A	11	Color	1,2,4,8	150	5.3	400	1000

The Exp. number is kept the same as the original article for reference, where experiment 11 is the new dataset. While all are experiments designed to test VWM, the eccentricity, stimulus duration, and delay all influence the results. Therefore, a multitude of values for these design aspects are used so any differences found between model fits will be due to theoretical differences, rather than experimental design.

3.1. Replication

Before using the model comparison procedure created in Chapter 2, a replication of previous research is important to 1) make sure our MLE procedure works properly and 2) give us a reference point for which our model flexibility analysis can be compared.

It is common to fit all datasets at once rather than separately for each set size. There are two major ways of fitting data for all set sizes: 1) simply fit data for each set size and pool the results into one log-likelihood (Suchow et al., 2013) and 2) include free parameters that match the explanatory model across set sizes (e.g. step function versus power function for representation variability (see van den Berg & Ma, 2014; Zhang & Luck, 2008). The largest issue with the hierarchical model fitting procedure in method 2 is the combination of highly variable data and the comparison of two very similar distributions – a step function and a power function. Figure 2 shows the subject-by-subject SLOT (step function) and VP (power function) fits to SD calculated from empirical data from (van den Berg et al., 2012). Although VP outperforms SLOT in this example, simulations show this difference is not reliable (56% correct distribution chosen across 100 runs). Therefore the current replication will focus on method 1.

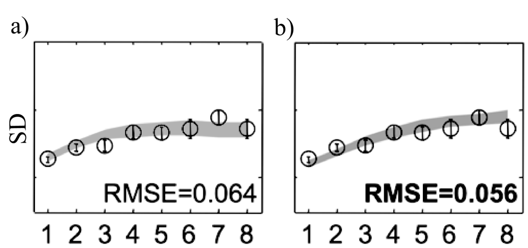

Figure 2. Graphs of SD across set size from (van den Berg et al., 2012). Black circles represent empirical data, and colored bands indicated fits for a) SLOT and b) VP.

A formal model comparison is done between SLOT and VP for the data discussed in *3.0.* Overall, VP wins over SLOT 61% of all fits (198 total subjects) and the mean delta AIC value is -1.32 in favor of VP. The magnitude of the difference is smaller than in previous literature (mean delta AIC ~ -10; van den Berg et al., 2012; 2014) which is likely due to their inclusion of a lapse term, however VP clearly outperforms SLOT using MLE as seen previously. Figure 3 shows the mean delta AIC values by set size. As expected, VP outperforms SLOT at low set sizes. The model comparison is inconclusive at high set sizes, suggesting both models handle the data equally well. A large variability is seen at set size 7 due to the low participant number (only 5 datasets) and will therefore not be included in the current investigation.

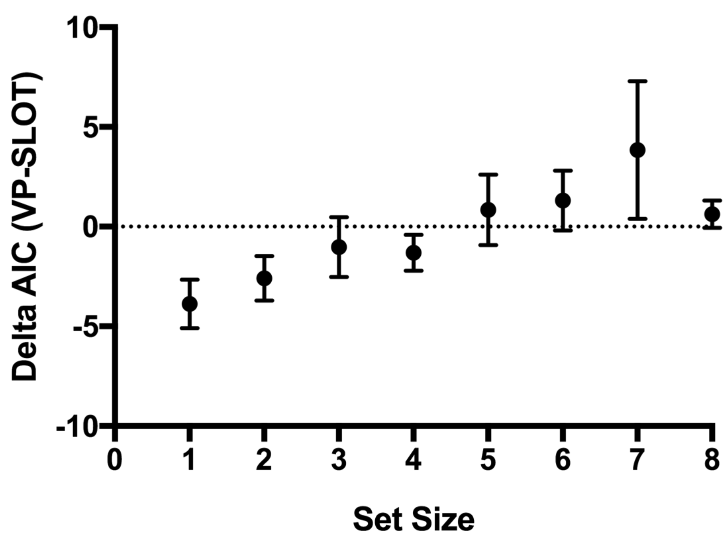

Figure 3. Delta AIC for VP minus SLOT for each set size. A negative value indicates a better fit for VP.

3.1.1. Residuals

If a misfit is found, it is important to understand what aspect of the model is leading to the misfit so we may better understand if it is a failure of the explanatory model or a misinterpretation of the measurement model. To achieve this, we observe residuals from fits to empirical data as well as residuals from fits to simulated data in a process used previously (van den Berg & Ma, 2014). If the explanatory model mimics underlying brain processes well, then empirical data should resemble data simulated from the corresponding measurement model. As no model is perfect as discussed in *Chapter 1*, the residuals found from measurement model fits to empirical data often form a pattern across error degrees (e.g. the common ‘mexican hat’ residual found when fitting SLOT to empirical data). If you see the same residual pattern when fitting that measurement model to simulated data, then that is support for the model used for the simulated data. This process with SLOT and VP includes:

1. Fit empirical data with both SLOT and VP, extract parameters
2. Use parameters to simulate data with both SLOT and VP
3. Find residual pattern when fitting SLOT and VP to empirical and simulated data

   Note: Fitting SLOT to data simulated with SLOT will result in no residual, and likewise for VP.
4. Compare residual patterns

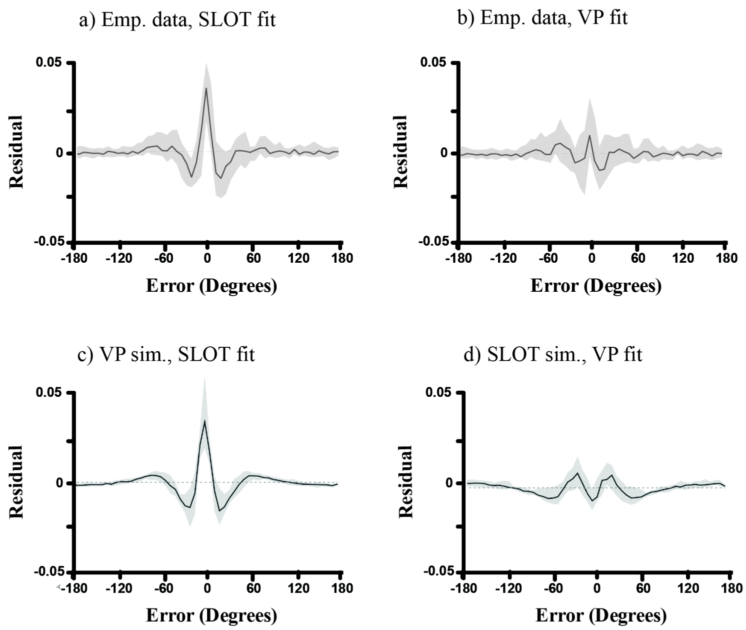

Figure 4. Residual plots for empirical data fit with SLOT (a) and VP (b), along with simulated data from VP fit with SLOT (c) and simulated data from SLOT fit with VP (d).

Replicating previous findings (van den Berg & Ma, 2014), the pattern of residuals when fitting SLOT to empirical data resembles the pattern of residuals when fitting SLOT to data simulated with VP (Figure 4a and 4b), yet this is not true vise versa (Figure 4b and 4d) suggesting data simulated with VP resembles empirical data the best. It is possible that the observed residual pattern difference does not relate to any theoretical difference between slot and resource models, but rather by a choice in measurement model parameters. This is explored in *Chapter 4*.

3.2. Formal model comparison

Now that the results from previous research have been replicated, the new model comparison method including model flexibility is used to investigate support for SLOT or VP. Again, the major difference between the two models is the existence of a capacity limit. At low set sizes, the existence of a capacity limit is not seen; therefore the key comparison is SLOT versus VP at high set sizes.

The parameters found in *3.1* are used to calculate the bias term for each subject fit using the procedure found in *2.2.2*. That is, a new bias term will be calculated for each set size for each participant's data separately. It is important to note that the inclusion of model flexibility does not change the fit, but rather simply penalizes models if too flexible. Therefore the residual results found in *3.1.1* still apply and must be addressed (see Chapter 4).

Bias calculated for each subject separately by set size can be seen in Figure 5 (mean = 0.22, SD = 0.84, range = [-1.72, 4.39]). Again, a positive value indicates a bias toward VP, so VP will be penalized when the bias term is added to the GOF. The large range of bias values indicates the necessity of calculating bias for each dataset and condition separately rather than using an idealized version.

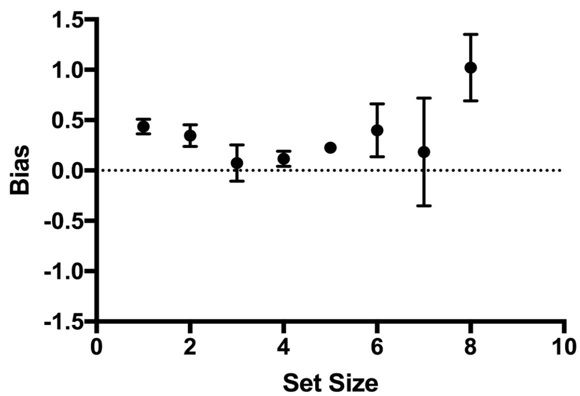

Figure 5. β for each set size in log likelihood units. A positive value indicates bias toward VP. Error bars in SE.

Overall the model comparison method is biased toward VP, especially at set size 8. Although the magnitude of the bias is not large, it makes a large difference in the results of the empirical model comparison as seen in Figure 6.

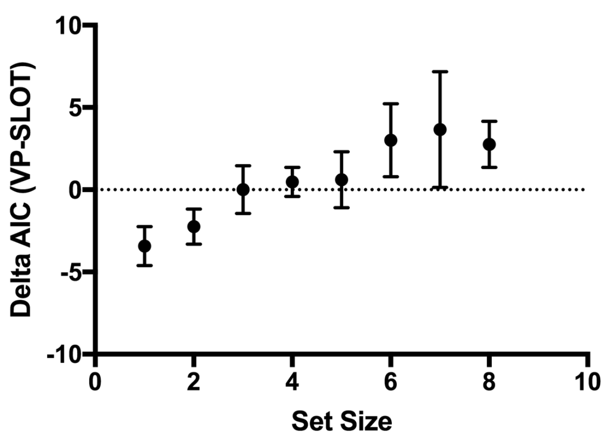

Figure 6. ΔAIC for model comparison between VP and SLOT for each set size. Error bars in SE.

Overall, VP wins over SLOT 58% of all fits (198 total subjects) and the mean delta AIC value is 0.6 in favor of SLOT. As mentioned previously, the key is to compare SLOT and VP fits at high set size where a capacity limit may be seen. SLOT outperforms VP at these high set sizes supporting the existence of a capacity limit as proposed by the slot theoretical model (set size 6: percent win for SLOT = 70%, mean ΔAIC = 3.00, SD ΔAIC = 5.00, range = [-0.96, 11.4]; set size 8: percent win for SLOT = 80%, mean ΔAIC = 2.76, SD ΔAIC = 3.3, range = [-0.06, 7.7]).

3.3. Proportion of guessing in VP

Since the model flexibility analysis is model mimicry, saying VP is more flexible than SLOT suggests that VP is able to account for more of SLOT predictions than vise versa. It is important to understand exactly what aspects of SLOT can be accounted for by VP. In particular, if VP can account for guessing due to its large flexibility, then it is possible to find a capacity limit within the VP model itself. It is important to note that VP is continuous, whereas SLOT is discrete which will make the guessing estimation differ between the two models. That is, for VP, as set size increases, the whole gamma distribution will be shifted toward a lower precision, which will affect the fits on the low-error responses. For SLOT, the guessing parameter varies independent of the normal distribution avoiding the potential misfit for low-error responses.

To find the proportion of guessing in VP, it must be understood what guessing looks like using normal distributions. As the standard deviation of a normal distribution is increased, it will get more difficult for our model comparison method to tell the difference between a normal distribution and a uniform distribution. The more trials we have, the easier it will be to tell the difference. To determine what proportion of the gamma distribution is capturing a uniform distribution, we need to determine the point at which the model fitting procedure can no longer distinguish between a uniform and normal distribution. To find this point for each of the 11 datasets, the following procedure is used.

1. Find the number of trials for the dataset.
2. Simulate data for a von mises while varying the concentration parameter (in degrees, from 100°-180°, increment of 1°).
3. Do a model comparison between a uniform and normal distribution for the simulated data.
4. Find the threshold past which the model comparison can no longer tell the difference (~115°)
5. Fit the data using VP to find the parameters for the gamma distribution.
6. Find the proportion of trials that are captured by the gamma distribution below the threshold found in (4).

The results are seen in Figure 7. As expected, the proportion of trials that fall below the threshold increases once set size is above 4. For low set sizes, very little of the gamma distribution falls below the threshold, which indicates participants are remembering the majority of the colors. At set size 5 and higher, we find capacity limitations. Due to the continuous nature of the VP model, the estimated capacity is lower than what is found using SLOT model, though not unreasonable compared to previous literature (Zhang & Luck, 2008).

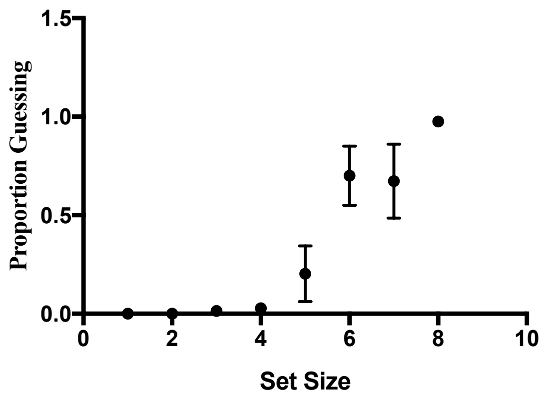

Figure 7. Proportion guessing estimate from VP fits for each set size.

3.4. Non-parametric MLE (NPMLE)

Measurement models of explanatory models are useful for formal model comparison but may lead to conclusions based on aspects of the model that are not part of the explanatory models (see Chapter 4 for an example of this issue). It is therefore useful to ask the primary question of interest, the existence of a capacity limit in WM, without assuming a measurement model. To do so, an estimation of the proportion of trials accounted for by a uniform distribution is found non-parametrically. The NPMLE method used is similar to the VP measurement model, in that the resulting model is a mixture model consisting of many von Mises distributions. Unlike the VP model, the von Mises distributions are added to the mixture distribution one at a time with variable precision (as a free parameter). In this way, the precision estimates are discrete and may very independently of each other.

A constrained Newton method with multiple supports has been shown to allow fast computation of the non-parametric MLE in circular space (Wang, 2007), and is used here. The means of the von Mises distributions are constrained to zero, such that the only free parameter is the precision of the distributions. For each dataset, the concentration parameters (*kappa*) and the proportion of data captured by the corresponding von Mises distributions (*pi*) are calculated. Note that a lower pi indicates a wider von Mises distribution, with *pi*=0 indicating infinite width (a uniform distribution). Across all experiments and subjects, the fitting procedure converged after adding 2-6 von Mises distributions. That is, after adding 2-6 von Mises distributions, adding more did not significantly improve fits and the fit converged. With so few distributions, the NPMLE method cannot be used to determine whether the empirical data stems from a continuous or discrete mechanism, and therefore cannot test between SLOT and VP directly. However, it can be used to find the proportion of guessing trials without assuming a measurement model.

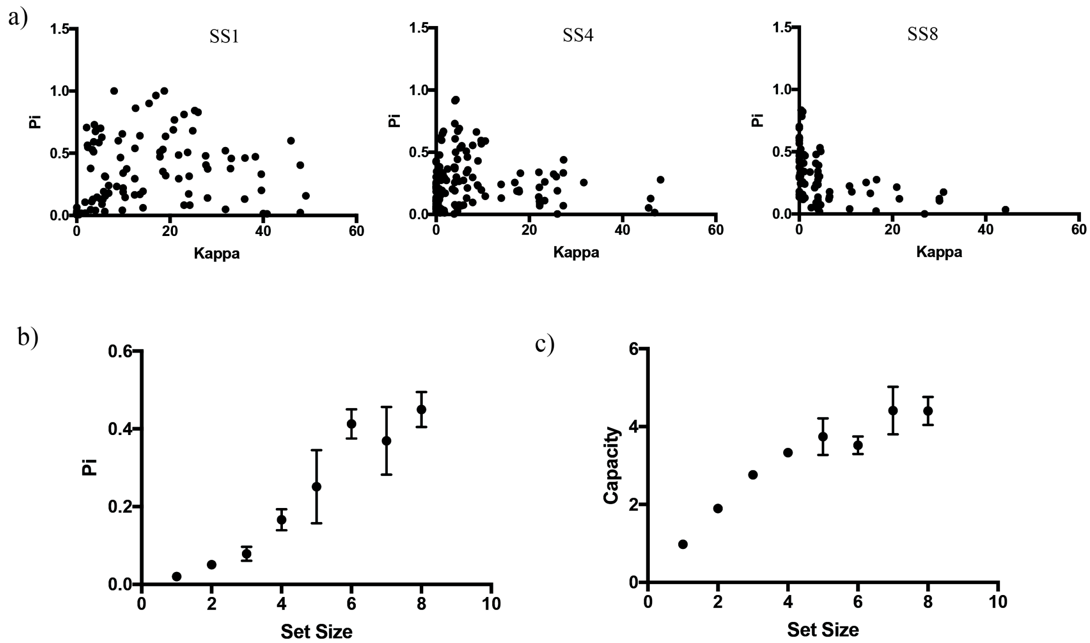

Figure 8. a) NPMLE results for empirical data from set size 1, 4 and 8. b) Estimated mean *pi* for *kappa*=0 by set size. c) Estimated capacity by set size. Note: All error bars in standard error.

As seen in Figure 8a and 8b, the proportion of data captured by a *kappa*=0 increases as set size increases. Non-zero values are highly variable as expected with individual differences between participants. Capacity is estimated using (1-(*pi*|*kappa=0)*)*SS resulting in a capacity of ~4 from set sizes 5 – 8 (Figure 8c). If there was no capacity limit and all items were contained in memory, the low-resolution trials could be captured by non-zero *kappa*. Using NPMLE, it is found that the high error responses do require a uniform distribution to be captured, supporting the existence of a capacity limit as theorized by the slot model.

3.5. Discussion

Chapter 3 compared the slot and resource theoretical models using several different methods. The key difference between the theoretical models is a capacity limit, where slot predicts a limit on capacity, and the resource model does not. Using continuous-recall VWM experiments, any capacity limit of the VWM system would manifest as a uniform distribution across circular space. Therefore the current formal model comparison focused on the need for a uniform distribution to capture empirical data, particularly at high set sizes above participant capacities.

Support for the VP model over SLOT is first found using a simple MLE formal model comparison, replicating past results (van den Berg et al., 2012; van den Berg & Ma, 2014). Observing fits for each set size, the major issue is the SLOT model's inability to handle data from low set sizes, where there is no clear support for either model at higher set sizes. This is also observed in the pattern of residuals from SLOT and VP fits to empirical data. The SLOT residuals show a clear patter at low set sizes, with the largest residual (error in fit) at zero degrees error, while VP has lower residuals overall. Together, these results suggest both SLOT and VP are able to fit high error responses, but VP is better equipped to capture variability in low error responses. In addition, when data is simulated with VP, the residual pattern from SLOT fits is once again observed suggesting VP is capturing the underlying VWM mechanism.

Simulations from Chapter 2 suggest that VP is more flexible than SLOT, giving it an unfair advantage in model comparison. A bias term is therefore added to the model fits that allows for a fair comparison between SLOT and VP. With this bias term included, SLOT now outperforms VP at high set sizes supporting the need for a uniform distribution when the set size is above participants' capacity limit. The need for a uniform distribution can also be found using VP, where some of the normal distributions included in the VP mixture model have widths so wide that they are indistinguishable from a uniform distribution. Looking at the proportion of trials that are captured by these wide normal distributions, a capacity is observed that resembles what is seen in the literature (Zhang & Luck, 2008).

Finally, the need for a uniform distribution to capture continuous-report VWM empirical data is explored non-parametrically, without assuming any theoretical model. This is important because a measurement model may lead to unforeseen issues due to choices in free parameters that do not have anything to do with the theoretical models. Using the NPMLE approach, support is again found for the need of a uniform distribution.

All together, strong support is found for the need of a uniform distribution, and by extension the need for a capacity limit, using three separate methods. However, the VP model still outperforms the SLOT model overall, but this is not due to the inclusion of a uniform distribution. Therefore the normal distribution of the SLOT model must be updated to capture empirical data, which is explored in Chapter 4.

Chapter 4

4.0. Slot with variability (SLOTv)

The original measurement model for the slot model (SLOT) has lead to misinterpretations of the explanatory model due to the lack of the term for variability of memory representations. VP assumes this variability stems from noisy memory representations, and accounts for the variability using a mixture of many normal distributions with varied precision. VP has been compared to SLOT, which accounts for memory representation noise using one normal distribution. The decision to use one normal distribution in SLOT was not reflecting a theoretical viewpoint that internal representations may not vary in precision, but was rather a simple statistic that reflected average memory precision across memory items and experimental trials. Therefore any of the improvements in the performance of VP compared to SLOT may be due to the lack of variability, rather than investigating the true difference between the two explanatory models. Instead, a measurement model needs to be created that investigates the major theoretical difference between the two models: is there a capacity limit in VWM? The slot-model assumes there is, where any memory items that are not collected by the limited number of slots are lost. VP assumes no capacity limit to working memory, where all items are retained until retrieval.

To test for a capacity limit, an updated measurement model for the slot-model is needed that captures both the capacity limit as a uniform distribution and the variability of memory representations (SLOTv). By comparing SLOTv to SLOT, we can directly investigate whether or not the addition of variability to the measurement model is necessary. Next, by comparing SLOTv to VP, we investigate whether or not a capacity limit is beneficial to describe working memory performance. As discussed above, the VP measurement model may also include a uniform distribution by allowing precision values of zero. The key difference between SLOTv and VP, then, is the independence between the variability of memory representations and the uniform distribution. If this independence more accurately represents the underlying WM mechanism, we expect SLOTv to outperform VP at high set size. As seen with SLOT in *Chapter 3*, at low set sizes the uniform distribution will not benefit model fits.

4.1. SLOTv Measurement model

The new measurement model for the slot model (SLOTv) contains a normal distribution in precision space that is sampled from to allow for variability of in memory responses. A normal distribution is used because the free parameters of the normal distribution, mean and standard deviation, have psychological meaning: the mean indicating the overall VWM precision of the participant and the standard deviation indicating the overall variability of the VWM representations. Three parameters are used, one for the uniform distribution and a mean and standard deviation of the normal distribution. SLOTv is formally defined as:

$$p(x \mid \mu) = (1-g) * \sum VM(x;\mu,k)dk + g\frac{1}{2\pi} \text{ where } k \sim N(m,\sigma)$$

where m and σ are the mean and standard deviation of the normal distribution from which precision is sampled from. SLOTv therefore has three free parameters: *g*, *m*, and σ. The addition of an extra free parameter may lead to overfitting, which will be accounted for by the penalty term in AIC (Akaike, 1973).

4.2. Residuals

In section *3.1.1*, a residual pattern was found when fitting SLOT to data simulated with VP that resembled the residual pattern when fitting SLOT to empirical data, yet this was not true when fitting VP to data simulated with SLOT. The observed residual pattern resides within the range -60° to 60° suggesting the issue is with describing low error responses, or the single normal distribution in SLOT. As discussed above, an updated version of SLOT, SLOTv, is better equipped to handle variability in memory representation while retaining the primary theoretical viewpoint – a capacity limit. An identical procedure used in *3.1.1* is used again now including three models: SLOT, VP and SLOTv. The results can be seen in Figure 9.

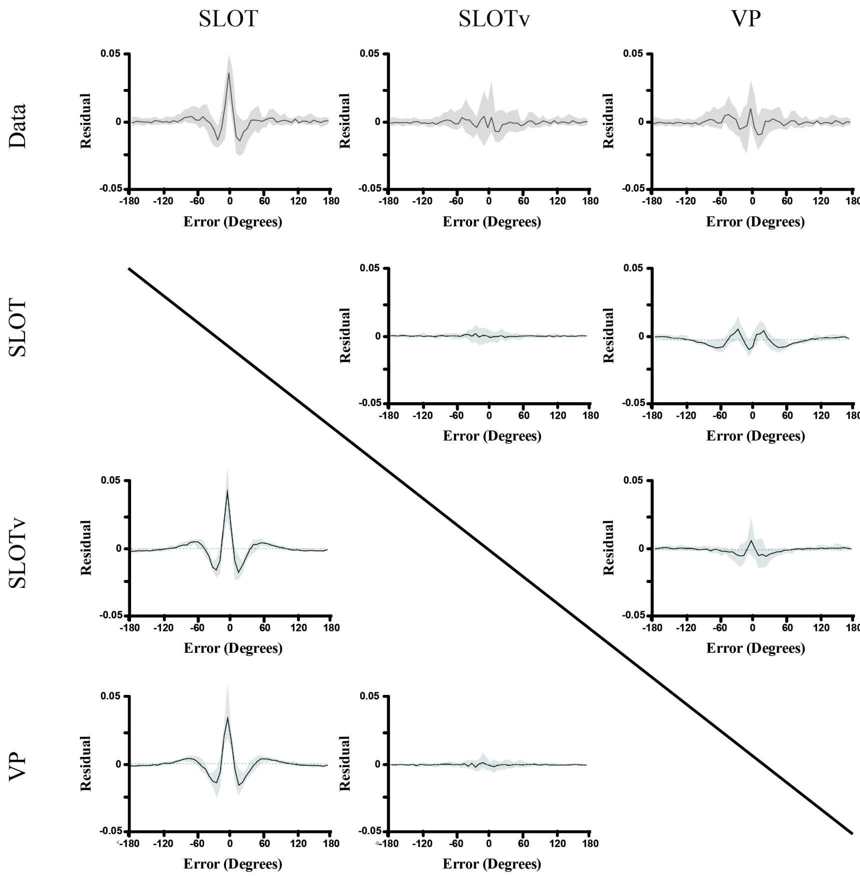

Figure 9. Residual plots of model fits including SLOT (first column) SLOTv (second column) and VP (third column) to empirical data (first row) and data simulated with models including SLOT (second row) SLOTv (third row) and VP (fourth row).

Several key results are found in the residual patterns. First, the 'Mexican hat' residual pattern seen when fitting SLOT to empirical data is not evident when fitting SLOTv to empirical data. Second, SLOT fits to data simulated with both SLOTv and VP show the same Mexican hat pattern, suggesting both models accurately represent empirical data.

4.3. Formal model comparison

A bias term is calculated using the same procedure as used in section *2.2.2* for SLOTv versus SLOT and for SLOTv versus VP. Results are seen in Figure 10 with a larger value indicating a bias toward SLOTv.

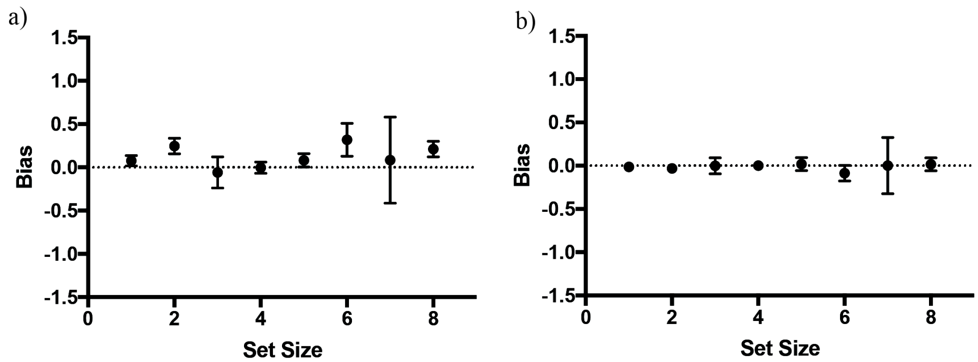

Figure 10. β for a) SLOTv versus SLOT and b) SLOTv versus VP. Note a larger value indicates a bias toward SLOTv.

A bias toward SLOTv is found when compared to SLOT (Figure 10a). This bias is similar to the bias toward VP found in Chapter 3.2, but with a smaller magnitude likely due to the AIC penalty term for the number of free parameters. Interestingly, nearly no bias is found comparing SLOTv and VP, suggesting the AIC penalty term penalized SLOTv for the third parameter to an appropriate extent. Without the addition of a uniform distribution, SLOTv and VP are nearly identical models, which results in no bias observed. Due to the variability of the bias and the value of including model flexibility in formal model comparison seen in Chapter 2, the bias terms will be included in the current formal model comparison.

Using the bias term, a formal model comparison is performed first comparing SLOT and SLOTv to determine if the addition of another free parameter to allow for variability for in-memory representations is necessary. Overall, SLOTv wins over SLOT 63.3% of the time. As seen in Figure 11, the only clear advantage that SLOTv has over SLOT is at low set sizes (set size 1: percent win for SLOTv = 74%, mean ΔAIC = -5.5, SE ΔAIC = 1.24, range = [-27.1, 2.4]; set size 2: percent win for SLOT = 67%, mean ΔAIC = -4.0, SE ΔAIC = 1.0, range = [-19.2, 2.1]). At higher set sizes, the addition of the third free parameter does not improve SLOTv fits enough for it to overcome the penalty added for the free parameter in AIC. The results suggest the addition of the third free parameter in SLOTv improves the model performance at low set sizes, and does not impair model fits at high set sizes, suggesting it is preferred as the new measurement model for the slot theoretical model.

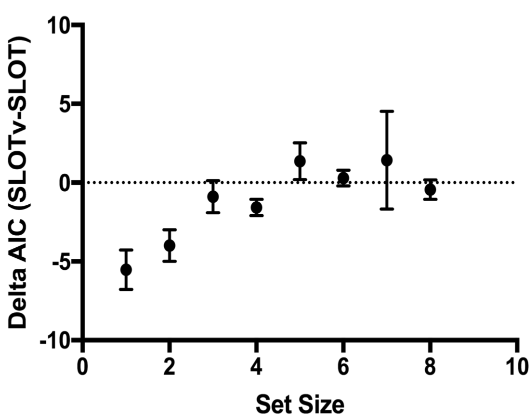

Figure 11. ΔAIC for model comparison between SLOTv and SLOT for each set size. Error bars in SE.

Since the addition of the extra free parameter is found to be necessary in the measurement model for the slot theoretical model, SLOTv is now compared to VP. Overall, SLOTv outperforms VP 93% of fits. No difference in fits is observable at low set sizes, but SLOTv clearly outperforms VP at high set sizes as seen in Figure 12b. A side-by-side comparison between SLOT-VP and SLOTv-VP shows the value of adding variability to the measurement model. Originally, VP outperformed SLOT at low set sizes and SLOT outperformed VP at high set sizes. Now, with the updated measurement model, SLOTv and VP show similar fits at low set size but again SLOTv outperforms VP at high set size. All together these results suggest 1) variability must be included in measurement models to account for in-memory representations and 2) a uniform distribution is needed to account for high set size conditions supporting the existence of a capacity limit.

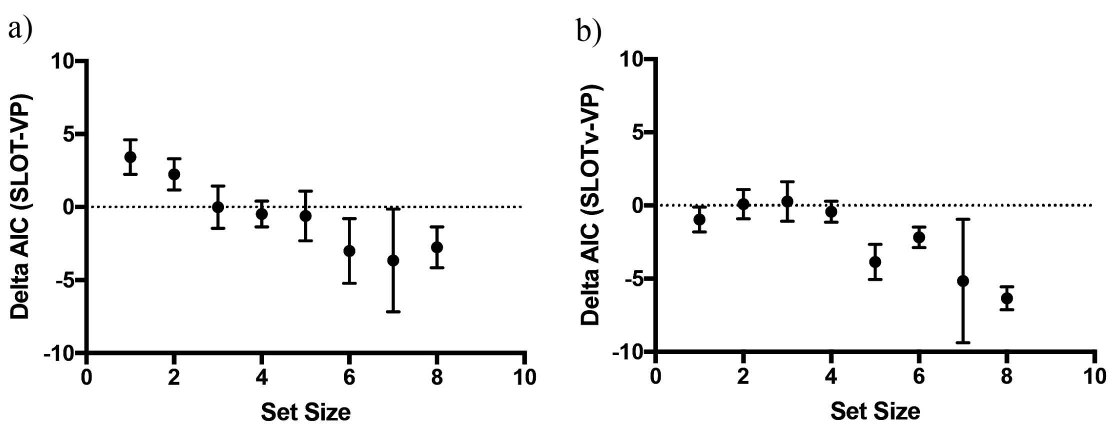

Figure 12. ΔAIC for model comparison between a) SLOT and VP and b) SLOTv and VP for each set size. Error bars in SE.

4.4. Discussion

Creating measurement models forces researchers to be very specific about what each aspect of the model represents using free parameters. The difference between SLOT and SLOTv is a good example of how these choices may lead to incorrect conclusions. Originally, in-memory representations were captured by a single normal distribution that acted as a mean precision statistic. VP capitalized on this use by creating a largely flexible model that allowed for in-memory variability and found it outperformed SLOT (van den Berg et al., 2012; van den Berg & Ma, 2014).

However the slot theoretical model makes no claims about a lack of variability in memory representations, so the original model fits did not support any theoretical differences in the theoretical models. Here, the updated measurement model SLOTv solves this issue by allowing for variability in the free parameters while retaining a capacity limit parameter. It outperforms SLOT suggesting that in-memory representations have variability that does not resemble a single uniform distribution, and also outperforms VP supporting the existence of a capacity limit.

Chapter 5

5.0. Discussion

Several lines of research are used to investigate the primary difference between the slot and resource theoretical models: is there a capacity limit in VWM? The majority of the current dissertation focuses on two primary measurement models for slot and resource models (SLOT and VP, respectively). When model flexibility is appropriately penalized, a large benefit is found for the inclusion of a capacity limit in the measurement models. This finding is supported by determining the proportion of VP that captures a capacity limit, which follows the same pattern as proposed by the slot model. Even when no model is assumed using a non-parametric approach, the data requires a capacity limit to be fit appropriately. Finally, an updated version of SLOT, named SLOTv, is created that more accurately represents the slot theoretical model. In formal model comparison, SLOTv outperforms VP at high set sizes and removes the residual pattern observed using SLOT, again adding support for a capacity limit in VWM.

5.1. Future research

The slot model predicts independence between capacity, which relies on the number of slots, and memory precision for the representations within the slots. Resource models assume there is no set capacity limit, and any guessing we see is due to low precision representations. Therefore it is important to test if we can manipulate the capacity and resolution of memory representations independently. This is not novel, and has been used to support the slot model previously e.g. (Zhang & Luck, 2008). Consolidation masking, where a mask is displayed after the offset of the memory array to interrupt consolidation of working memory representations, decreases the number of items encoded while leaving the precision of the memory representations intact. On the other hand white noise masking, a noise mask presented with the memory array, decreases memory precision but does not affect the number of items encoded. However, resource models may account for these differences using a flexible measurement model such as VP. When VWM is manipulated using these paradigms, the gamma distribution may be able to shift and continue to describe the data. To get around this issue, the neural systems that are involved in capacity and resolution can be investigated. If support is found for independent brain systems for capacity and resolution, it will be strong support for the slot model.

5.1.1. Neural mechanism behind working memory capacity

Historically, working memory representations were thought to be retained as sustained neural activity due to the persistent neural firing observed during the delay interval of a working memory task in the prefrontal cortex (Fuster & Alexander, 1971), the inferotemporal cortex (Fuster & Jervey, 1982), and parietal cortex (Todd & Marois, 2005). The sustained neural activity in the parietal cortex can even be used to predict working memory capacity (Todd & Marois, 2005), however the growing consensus is that the capacity of working memory is dependent on multiple brain regions (Eriksson, Vogel, Lansner, Bergström, & Nyberg, 2015; Postle, 2015).

One prominent theory for the mechanism behind the limited capacity observed in visual working memory focuses on neural oscillations between the parietal lobe and frontal cortex (Raffone & Wolters, 2001). In this model, oscillations in the alpha range between the parietal lobe and frontal cortex carry active working memory representations. If the phase of these oscillations match, then the two representations will combine into one, effectively decreasing the capacity of the system by one representation. Through simulations, Raffone & Wolters show that there can only be 3-4 simultaneous oscillations at a time before they begin to phase lock, which results in the capacity limit of three to four items as observed behaviorally (Luck & Vogel, 1997).

5.1.2. Neural mechanism behind working memory precision

Two primary models have been proposed to account for working memory precision. First, the neural noise hypothesis is a neural instantiation of the resource model, and attributes all limitations found in working memory to randomness in neural spiking (neural noise; Bays, 2015). Each working memory representation is retained as the sustained firing of an ensemble of neurons, where each neuron in the ensemble contributes to one preferred feature (e.g. vertical orientation). Each individual neuron's activity is related to the activity a large population of neurons due to interconnections, called normalization. As more neurons are active (i.e. as the set size increases), each neuron becomes more normalized resulting in less precision as observed behaviorally (Zhang & Luck, 2008). However, this model cannot account for the capacity limit observed in working memory performance.

Second, the sensory recruitment hypothesis assumes working memory representations 'recruit' the same sensory neurons that were used to encode them while they are being retained over a delay. This recruitment results in the precision of the working memory representation. Previous research has found that you can decode the contents of working memory during the delay interval as early as V1 (for orientation bars), suggesting these areas are still active during the delay interval (Ester, Anderson, Serences, & Awh, 2013; Harrison & Tong, 2009). However, the precision of perception or sensory memory does not seem to predict the precision of working memory at the behavioral level (Brady, Konkle, Gill, Oliva, & Alvarez, 2013; Cappiello & Zhang, 2016), suggesting another mechanism is needed.

One candidate for this mechanism suggested by the hippocampus-precision hypothesis, is the pattern separation computation in the hippocampus. A large literature has investigated the ability to discriminate similar long-term memory representations, and strong support has been found for a mechanism that converts population codes coming from sensory areas into sparse codes in the hippocampus e.g. (S. M. Stark, Yassa, Lacy, & Stark, 2013; Yassa & Stark, 2011) . This pattern separation computation could also help keep working memory representations separate, and would give each representation precision.

5.1.3. Research opportunity

The ongoing debate on VWM mechanisms does not shed a clear light on whether there are separate neural mechanisms for capacity and resolution. However, the theoretical viewpoints discussed above suggest an opportunity for neural manipulation that may dissociate these two aspects of VWM. One option for such and investigation is using non-invasive brain stimulation (NIBS) to modulate areas that are thought to be involved in VWM capacity and resolution. NIBS has become a common way to find causal relationships between brain region and function (Nitsche et al., 2008). These techniques are relatively new, and a full understanding of the mechanism behind any observed effects has not been reached. Transcranial direct current stimulation (tDCS) and transcranial alternating current (tACS) use low voltage electric currents to modulate neural activity. tDCS has been found to facilitate transsynaptic activity of the stimulated region in cats even in deep brain tissues (Bolzoni, Pettersson, & Jankowska, 2013). It is therefore a useful tool for manipulating the activity in deep brain structures, unlike other NIBS such as transcranial magnetic stimulation. tACS has been shown to entrain brain oscillations at low and high frequencies in the animal model (Reato, 2013) by affecting excitability, shifts in spike timing, and modulation of firing rate.

If the neural oscillation theory for VWM capacity is true, then tACS in the parietal lobe should modulate VWM capacity but not precision. Likewise, if the hippocampus-precision hypothesis is true, then anterior temporal lobe tDCS should modulate VWM precision but not capacity. Preliminary results show this is the case, with parietal lobe tACS decreasing VWM capacity, and ATL tDCS decreasing precision. However, due to the widespread effect of tACS and tDCS, it is difficult to know exactly what part of the brain being manipulated is resulting in the observed VWM effects.

Therefore more research is needed to become confident in the separate capacity and resolution mechanisms for VWM.

5.1.4. Alternative computational models of VWM

An alternative account for VWM limitation has recently been proposed called the interference model (IM; Oberauer & Lin, 2017). IM proposes all observed working memory limitations are due to variable precision of the in-memory representations and interference between the memory representations at retrieval (Oberauer & Lin, 2017). Similar to resource models, the interference model predicts all memory items are stored simultaneously, however IM predicts one representation is in focus of attention and the rest in active long-term memory. During retrieval, the context of the cue may target multiple stored representations for response, which leads to interference and decreased performance. As set size increases, the likelihood of the interference increases resulting in the observed working memory limitation.

SLOT, IM, and VP have been compared in a formal model comparison (Oberauer & Lin, 2017), however this model comparison used the original version of the slot-model, with one normal and one uniform distribution, and did not address model flexibility. It is therefore important to redo the formal model comparison using the techniques and models created in this dissertation. However IM requires 6 parameters, which requires a larger number of trials for each condition to appropriately compare the models (~1000 trials per condition based on simulations).

It is also possible to investigate IM experimentally. According to IM, all guessing is due to interference at retrieval. If this is true, then VWM of items from difference features (mixed condition; e.g. color and shape) should show a higher capacity than those from the same features (non-mixed condition; e.g. color only). Even if this is the case, however, it may be due to differences between the memory items at encoding rather than interference at retrieval. Instead, CDA could be measured to determine if all items are maintained during the delay interval.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *Proc. 2nd Inter. Symposium on Information Theory*, 267–281.

Anderson, D. E., & Awh, E. (2012). RETRACTED ARTICLE: The plateau in mnemonic resolution across large set sizes indicates discrete resource limits in visual working memory. *Attention, Perception & Psychophysics*, *74*(5), 891–910.

Anderson, D. E., Vogel, E. K., & Awh, E. (2011). RETRACTED ARTICLE: Precision in visual working memory reaches a stable plateau when individual item limits are exceeded. *Journal of Neuroscience*, *31*(3), 1128–1138.

Baddeley, A. (2012). Working Memory: Theories, Models, and Controversies. *Annual Review of Psychology*, *63*(1), 1–29.

Baddeley, A., & Hitch, G. (1974). Working Memory. *Psychology of Learning and Motivation*, 47–89.

Bays, P. M. (2015). Spikes not slots: noise in neuralpopulations limits working memory. *Trends in Cognitive Sciences*, *19*(8), 431–438.

Bays, P. M., & Husain, M. (2008). Dynamic Shifts of Limited Working Memory Resources in Human Vision. *Science*, *321*(5890), 851–854.

Bolzoni, F., Pettersson, L.-G., & Jankowska, E. (2013). Evidence for long-lasting subcortical facilitation by transcranial direct current stimulation in the cat. *The Journal of Physiology*, *591*(13), 3381–3399.

Box, G. E. P. (1976). Science and Statistics (pp. 791–799). Wiley-Interscience.

Brady, T. F., Konkle, T., Gill, J., Oliva, A., & Alvarez, G. A. (2013). Visual Long-Term Memory Has the Same Limit on Fidelity as Visual Working Memory. *Psychological Science*, *24*(6), 981–990.

Cappiello, M., & Zhang, W. (2016). A dual-trace model for visual sensory memory. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(11), 1903–1922.

Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences*, *7*(12), 547–552.

Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114– discussion 114–85.

Eriksson, J., Vogel, E. K., Lansner, A., Bergström, F., & Nyberg, L. (2015). Perspective. *Neuron*, *88*(1), 33–46.

Ester, E. F., Anderson, D. E., Serences, J. T., & Awh, E. (2013). A Neural Measure of Precision in Visual Working Memory. *Journal of Cognitive Neuroscience*, *25*(5), 754–761.

Fuster, J. M., & Alexander, G. E. (1971). Neuron activity related to short-term memory. *Science*. *173*(3997) (1971): 652-654.

Fuster, J. M., & Jervey, J. P. (1982). Neuronal firing in the inferotemporal cortex of the monkey in a visual memory task. *Journal of Neuroscience*. *2*(3), 361-375.

Hannula, D. E., Tranel, D., & Cohen, N. J. (2006). The Long and the Short of It: Relational Memory Impairments in Amnesia, Even at Short Lags. *Journal of Neuroscience*, *26*(32), 8352–8359.

Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, *458*(7238), 632–635.

Jang, Y., Wixted, J. T., & Huber, D. E. (2007). The diagnosticity of individual data for model selection: Comparing signal-detection models of recognition memory. *Psychological Bulletin*, *133*(5), 800–832.

Joormann, J., & Gotlib, I. H. (2008). Updating the contents of working memory in depression: Interference from irrelevant negative material. *Journal of Abnormal Psychology*, *117*(1), 182–192.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The Generality of Working Memory Capacity: A Latent-Variable Approach to Verbal and Visuospatial Memory Span and Reasoning. *Journal of Experimental Psychology: General*, *133*(2), 189–217.

Kane, M. J., Poole, B. J., Tuholski, S. W., & Engle, R. W. (2006). Working memory capacity and the top-down control of visual search: Exploring the boundaries of "executive attention". *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(4), 749–777.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*. *103*(3), 490.

Libby, L. A., Hannula, D. E., & Ranganath, C. (2014). Medial Temporal Lobe Coding of Item and Spatial Information during Relational Binding in Working Memory. *Journal of Neuroscience*, *34*(43), 14233–14242.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*. *63*(2), 81.

Nee, D. E., & Jonides, J. (2008). Neural correlates of access to short-term memory. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(37), 14228–14233.

Nitsche, M. A., Cohen, L. G., Wassermann, E. M., Priori, A., Lang, N., Antal, A., et al. (2008). Transcranial direct current stimulation: State of the art 2008. *Brain Stimulation*, *1*(3), 206–223.

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 411–421.

Oberauer, K., & Lin, H.Y. (2017). An interference model of visual working memory. *Psychological Review*, *124*(1), 21–59.

Oberauer, K., Süß, H.-M., Wilhelm, O., & Sander, N. (2008). Individual Differences in Working Memory Capacity and Reasoning Ability. *Variation in working memory,* 49–75.

Park, S., & Holzman, P. S. (1992). Schizophrenics show spatial working memory deficits. *Archives of General Psychiatry*, *49(12), 975-982.*

Postle, B. R. (2015). The cognitive neuroscience of visual short-term memory. *Current Opinion in Behavioral Sciences*, *1*, 40–46.

Rademaker, R. L., Tredway, C. H., & Tong, F. (2012). Introspective judgments predict the precision and likelihood of successful maintenance of visual working memory. *Journal of Vision*, *12*(13), 21–21.

Raffone, A., & Wolters, G. (2001). A cortical mechanism for binding in visual working memory. *Cognitive Neuroscience, Journal of*, *13*(6), 766–785.

Reato, D. (2013). Effects of weak transcranial alternating current stimulation on brain activity—a review of known mechanisms from animal studies. *Frontiers in human neuroscience, 8,* 687.

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, *107*(2), 358–367.

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *The Journal of neurology, neurosurgery, and psychiatry, 20*(1), 11.

Shiffrin, R. M., & Atkinson, R. C. (1967). Human Memory: a Proposed System and Its Control Processes. In *Psychology of learning and motivation* (Vol. 2, pp. 89-195). Academic Press.

Simons, J. S., & Spiers, H. J. (2003). Prefrontal and medial temporal lobe interactions in long-term memory. *Nature Reviews Neuroscience*, *4*(8), 637–648.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, *74*(11), 1.

Standing, L. (1973). Learning 10000 pictures. *The Quarterly Journal of Experimental Psychology*. *25*(2), 207-222.

Stark, S. M., Yassa, M. A., Lacy, J. W., & Stark, C. E. L. (2013). A task to assess behavioral pattern separation (BPS) in humans: Data from healthy aging and mild cognitive impairment. *Neuropsychologia*, *51*(12), 2442–2449.

Suchow, J. W., Brady, T. F., Fougnie, D., & Alvarez, G. A. (2013). Modeling visual working memory with the MemToolbox. *Journal of Vision*, *13*(10), 9–9.

Todd, J. J., & Marois, R. (2005). Posterior parietal cortex activity predicts individual differences in visual short-term memory capacity. *Cognitive, Affective, & Behavioral Neuroscience, 5*(2), 144-155.

van den Berg, R., & Ma, W. J. (2014). "Plateau-"related summary statistics are uninformative for comparing working memory models. *Attention, Perception & Psychophysics*, *76*(7), 2117–2135.

van den Berg, R., Awh, E., & Ma, W. J. (2014). Factorial comparison of working memory models. *Psychological Review*, *121*(1), 124–149.

van den Berg, R., Shin, H., Chou, W.-C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences*, *109*(22), 8780–8785.

Veksler, V. D., Myers, C. W., & Gluck, K. A. (2015). Model flexibility analysis. *Psychological Review*, *122*(4), 755–769.

Voytek, B., & Knight, R. T. (2010). Prefrontal cortex and basal ganglia contributions to visual working memory. *Proceedings Of the National Academy of Sciences*, *107*(42), 18167-18172.

Wagenmakers, E.-J., Ratcliff, R., Gomez, P., & Iverson, G. J. (2004). Assessing model mimicry using the parametric bootstrap. *Journal of Mathematical Psychology*, *48*(1), 28–50.

Wang, Y. (2007). On Fast Computation of the Non-Parametric Maximum Likelihood Estimate of a Mixing Distribution. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, *69*(2), 185–198.

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12), 11–11.

Xie, W., & Zhang, W. (2017). Dissociations of the number and precision of visual short-term memory representations in change detection. *Memory & cognition, 45*(8), 1423-1437.

Xie, W., Li, H., Ying, X., Zhu, S., Fu, R., Zou, Y., & Cui, Y. (2017). Affective bias in visual working memory is associated with capacity. *Cognition and Emotion*, 1–16.

Yassa, M. A., & Stark, C. E. L. (2011). Pattern separation in the hippocampus. *Trends in Neurosciences*, *34*(10), 515–525.

Zhang, W. (2007). *Resolution and capacity limitations in visual working memory: A new approach*. The University of Iowa.

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

Zhang, W., & Luck, S. J. (2009). Sudden Death and Gradual Decay in Visual Working Memory. *Psychological Science*, *20*(4), 423–428.

Zhang, W., & Luck, S. J. (2011). The Number and Quality of Representations in Working Memory. *Psychological Science*, *22*(11), 1434–1441.

Zhang, W., & Luck, S. J. (2015). Opposite effects of capacity load and resolution load on distractor processing. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(1), 22–27.