From Short-Term Tolerance to Long-Term Recognition in Human Visual Memory

By

Mark Schurgin

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for a degree of Doctor of Philosophy

Baltimore, Maryland

March, 2017

Abstract

Humans have a remarkable ability to recognize visual objects following limited exposure and despite changes at the image-level. How humans acquire this ability remains a mystery, and it remains one area in which artificial intelligence has yet to match human performance. I sought to understand this fundamental cognitive ability by leveraging theories and methods from multiple fields. In particular, I examined how rules guiding the perception of objects in visual working memory assist in the construction of visual long-term memories. In four experiments, I reveal that our expectations for how objects move in the world are used to learn and integrate object information into visual long-term memory. Next, I further examined how aspects of memory over the short-term may actually be features used to construct appropriately constrained representations in the long-term. I demonstrate in three experiments that visual working memory is highly tolerant to variability at test, in order to act as a venue to integrate information into long-term memory. Finally, I moved past investigating memory following singular experiences to understand how our memories change over repeated exposures. I discovered across three experiments that the initial quality of an

experience, as well the amount of time between repeated encounters, affects our ability to integrate and remember objects we encounter multiple times. This work contributes to our understanding of the growth process of visual memory, and attempts to form bridges between traditionally disparate fields of vision scientist studying object perception, neuroscientists studying long-term memory, and engineers designing artificial intelligence recognition systems.

Committee Members:

Jonathan Flombaum (Primary Advisor)

Jason Fischer

Howard Egeth

Steven Gross

Soojin Park

Alternate Members:

Marina Bedny

Colin Wilson

Acknowledgments

This dissertation is less a product of my own actions and much more a reflection of the many people who have touched my life.

Mom, this dissertation and my career are as much your accomplishments as they are my own. Thank you for always being there when I was a child, and raising me to be the man I am today. Your deep compassion is a trait I hope to carry with me throughout the rest of your life. Dad, I could have never asked for a better father. Whether you were coaching my little league team as a kid or kicking my butt when I needed it in college, you've always been there for me. If there is one person in this world I aspire to be like, it's you. If I am capable of half of your personal and professional success, I will live a very blessed life. Helen, it's crazy thinking of how far we've both come, and how many accomplishments we've shared and celebrated together. I am so fortunate to have you both as a second-mother and friend.

Liz, I know growing up we had our bickering, but I'm incredibly glad we have been able to become so close as adults. Of anyone in my life you had the most direct influence on my career. If it wasn't for your prodding and support, I would have never

gotten into research. Sam, I'm very fortunate to have you as a member of my family. I'm even more fortunate to be able to call you a close friend. Thanks for being such a supportive brother-in-law.

Alyssa and Becky, it's meant so much to have you both in my life. Even when I haven't asked for it, you've both been there for me. Dave and Mary, I can't thank you both enough for all your love and support. In particular, our political conversations are one of the few places in my life where despite our disagreements, we both listen to and respect one another. You have no idea how much that means to me. Tom, you're the best cousin I could have asked for. You are much more like a brother than a cousin to me.

Grandpa Art, I know you never lived to see me graduate high school or college. It remains one of my greatest regrets in life – I remember vividly how proud you were of Liz when she got into Yale, and I only wish I could have had the opportunity to do the same. I miss you so much, and I will always be grateful for the gift of education you left me. Grandma Evette and Grandpa Bob, your love has been a constant source of support in my life. Grandma, thanks for taking the time to read to me and teach me how to garden when I was little. Grandpa, you've always been full of advice and essential stories to live my life by.

Ms. Gibson, you were the first teacher to see something more in me. I have been underestimated many times in my life, but what you instilled in me as a child has

given me the strength to overcome those obstacles. Put simply, I am a better person for having known you. Rabbi Doug, my passion for science, philosophy and knowledge was first fostered in your office through our many discussions. I don't know if I ever told you this, but when I graduated high school I seriously considered attending rabbinical school. One of the few things in my life I knew was meaningful and wanted to pursue were discussions like ours. While I ultimately didn't go that route, I think obtaining my doctorate remains much in that spirit.

Jack, you mean more to me than you'll ever know. I think we both found each other at a point in life where we needed each other the most. I couldn't have asked for (and will never have) a better dog and companion. I'm so incredibly grateful to have you in my life.

To Eric Sorkin, you are one of my oldest and truest friends – we'll always be each other's "twins". Josh Abrams, you're one of the smartest and most creative people I'll ever know. Our ridiculous childhood projects aside, I'm thankful we're still in each other lives. Wes, I know ever since I left Chicago grad school has absorbed my life and our creative projects have fallen by the wayside. That's completely on me, and something I intend to address post-graduation. To be completely honest, one of the hardest parts of grad-school has been being away from you (oh god, writing this was worse than our in-betweeners moment).

Chris Long, you were a confidant and friend when I needed one most and I'll never forget that. I appreciate, like an initially startled groundhog, you've calmed down in the past few years and come out of your Chris-hole. Dodo, I've had many friends come and go throughout the years, and yet we're still as close as ever. This is despite some incredible distance, which never seems like much when it comes to us. Eventually I'm sure we'll both retire in Switzerland and start our corgi farm, but until then I couldn't be happier to have you in my life. Charlie, we both know we were secretly brothers separated at birth, reunited through destiny and serendipity. Seriously, I don't know what else to say. I love you man.

Vani and Matt, we've seen each other through a lot in life and I wouldn't be where I am without your support. Although our random luck of living in the same city seems to have finally run out, I'm so glad that you both are creating a life together in Texas and can't wait to visit you soon. Hrag, while at Hopkins you were not only a great friend, but taught me about the importance of modesty (and dancing at Grand Central every now and then). Darko, you are one of the most amazing, genuine people I know. Despite living quite far apart since you graduated you've always put an effort in to be my close friend, and that's meant a lot to me. Michelle, it's rare to find someone I can have so much fun with, but have a deeply meaningful friendship with as well.

Marie, I'm glad over the years we've gone from "spousal" friends to just friends. Your kindness has meant a lot to me over the years, and will remain one of the

highlights of my time in Baltimore. Corbin, you're the worst. Like, you are quite literally the worst kind of person. You're the type of person who is so fun, goofy, helpful, kind, and loyal that someone in my position would never want to leave Baltimore. There's going to be a major gap in my life without our daily interactions, although I know we'll continue to stay close regardless of where life takes us. Jasmin, I know you don't think you've contributed much to this, but meeting you has changed my life. To everyone else at Hopkins, I can't thank you enough for making this such a memorable experience – Ela, Giulia, GiYeul, and Zheng. I am especially appreciative of all the hard and thoughtful work Hopkins administrative staff puts on behind the scenes – thank you Laura, Julie, Lisa, Rebecca and Jennifer.

Steve Franconeri, thanks for taking a chance on a young Vassar student with no idea what the hell he was doing. It didn't seem like a lot at the time, but that small act of kindness forever changed my life. Dr. Egeth, thank you for always making time for me when I needed it and giving me thoughtful feedback. Justin, what at times you may have thought were small suggestions greatly impacted the overall quality of my work and research. I can't thank you enough for your insight.

Finally, I would like to thank my advisor, Jon. You challenged me to be the quality of researcher I am today. Without your mentorship and guidance, I would simply not think the way I do. I am forever grateful.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This is a review designed with two purposes in mind. The majority of research on visual memory tends to focus exclusively on shorter or longer durations, i.e. on working or long-term memory, respectively. So the first purpose is simply to supply a primer that spans the two areas, with readers in mind who may only be familiar with one or the other. The second purpose is to identify points of overlap and distinction in the two literatures, with the hope that synthesis will help to clarify some of the major questions remaining in these areas.

1.1 Defining Memory and Memory "Systems"

When trying to understand the nature of visual memory, many researchers start by establishing a specific memory system to investigate. There are many approaches to how one might define a memory system. One of the earliest tactics used to define

memory systems was through dissociation. For instance, doctors removed portions of the medial temporal lobe (MTL) in patient H.M. As a result, H.M. suffered from an inability to form new episodic memories after his surgery. When H.M. was given a simple hand-eye coordination task to learn, despite having no memories of the task, he was able to demonstrate improvement in his skills over the course of a few days (Squire, 2004). The brain damage produced a functional dissociation between two different systems of memory: explicit and implicit memory. H.M. could not form new memories of events he experienced (explicit memory), but was still able to learn procedural skills (implicit memory).

The key distinction between these systems pertains to their internal representations. Explicit memory consists of facts and events, whereas implicit memory is an umbrella term for any other type of memory – including procedural skills and habits, priming, classical conditioning, and non-associative learning. Since the content of these memory systems are different, they are assumed to depend on different processes, brain areas, and computations. Explicit memory, for instance, is shown to rely on activity in the medial temporal lobe (through dissociation work established by patients such as H.M.), whereas an implicit system such as procedural skills and habits is more related to neural activity in the striatum (Squire, 2004).

Indeed, additional data confirms that a double dissociation exists between explicit and implicit memory. While I discussed a single example earlier (patient H.M.),

a double dissociation requires a second patient with impaired implicit memory but unimpaired explicit memory. Patient M.S. was an individual thought might meet such criteria, whose right occipital lobe was removed in order to treat symptoms related to epilepsy. Alongside healthy and amnesic controls, in an initial experiment M.S. saw visual stimuli of words either four or five letters long during a study phase, and then at subsequent test was asked to recall whether he had seen those words among foils. M.S. demonstrated no impaired ability for recognizing words, similar to healthy controls, although amnesic patients demonstrated worse performance, consistent with a deficit in explicit memory. In a subsequent experiment, all participants underwent a visual word completion priming task, which consisted of a similar study phase, but at test participants saw three-letter stems and were asked to complete each stem with the first word that came to mind. When comparing the proportion of stems completed to studied words relative to baseline (i.e. words not seen during the study phase) amnesic and healthy controls demonstrated similar performance, whereas M.S. demonstrated significantly less priming. This difference in priming was interpreted as a deficit in implicit memory, suggesting M.S.'s brain damage (or removal) had resulted in impaired implicit memory but unaffected explicit memory (Gabrieli et al., 1995). This provided support for the existence of two orthogonal processes that rely on distinct brain areas and internal content.

However, it may not be the case that distinguishing memory systems based on dissociation is the best approach. Does this mean we can only define or classify a memory system if it is supported by dissociations in the brain? Adding to this confusion, brain areas commonly associated with memory (that many memory dissociations rely on) have been shown to also contribute to a variety of other behaviors. Specifically, research has implicated the role of the hippocampus in both visual statistical learning and decision-making processes (Shohamy & Turk-Browne, 2013).

In one study on decision-making, participants were first given an association task where pairs of pictures were shown together (forming an association) while they performed a cover task (detecting an upside-down image). Then participants were given a reward phase, where half of the previously paired items were followed by either no reward or a monetary reward. In the final task, participants were shown the other half of paired items not present during the reward phase, and were asked to decide between two stimuli to receive possible monetary reward. It was found that participants' decisions were biased towards photos previously paired with items that received a reward, despite these photos never receiving a reward themselves. Additionally, this decision bias was predicted by activity in the hippocampus (Wimmer & Shohamy, 2012). This suggests that the role of the hippocampus extends past its

relationship with explicit memory, and is also implicated in reward and decision-making processes.

Given the ever expanding role of brain areas used to support dissociations in memory, it may be better to distinguish between memory systems using different approaches. There are other ways for systems to be established that don't have clear dissociations, but are supported by a variety of other differences. Dissociation implies independence, when it may be possible that different memory systems may at times support or relate to one another.

1.2 Distinguishing VWM & VLTM

Memories for visual information are typically distinguished as belonging to either visual working memory (VWM) or visual long-term memory (VLTM). VWM is considered an online system that retains and manipulates information over the short-term (Baddeley & Hitch, 1974; Vogel, Woodman & Luck, 2006; Cowan, 2008, Ma, Husain & Bays, 2014), whereas VLTM is typically defined as an automatic storage of visual information over longer periods of time (Squire, 2004; Cowan, 2008; Brady, Konkle & Alvarez, 2011).

It is intuitive to distinguish different memory systems, especially VWM and VLTM, by the time scale over which the memory takes place. However, while time may constrain systems in different ways it may not be sufficient to understand all possible differences and similarities between them. A more holistic approach may be to

distinguish these systems in terms of their functions. If we know the function of a system, we can clearly identify the specific challenges facing such functions, even if we do not know the specific solutions to these challenges a system may choose to employ.

The Function of VWM

When establishing the function of VWM, researchers early on defined it as a space divided between storage and other processing demands (Baddeley & Hitch, 1974). It is thought to be the combination of multiple processes, providing an interface between perception, short-term memory, and other mechanisms such as attention (Cowan, 2008). VWM's most notable characteristic is perhaps that it's capacity limited (Vogel, Woodman & Luck, 2001; Awh, Barton & Vogel, 2007), and it is thought to be a core cognitive process underlying a wide range of behaviors (Baddeley, 2003; Ma, Husain & Bays 2014). Given these descriptions, the function of VWM may be best described as supporting complex cognitive behaviors that require temporarily storing and manipulating information in order to produce actions.

VWM representations must therefore be in a format that is easily amendable and malleable, but also in a general format able to support a wide range of behaviors: it needs to output representations that can be taken as input by relevant processes. For example, let's say you need to solve the math problem: $\frac{2+6}{4}$. Your visual system needs

to be able to parse and store each number and symbol. Then, you must apply knowledge about the symbols to manipulate numbers according to specific rules (i.e. 2 + 6 = 8). While the manipulated information doesn't necessarily need to be visual, it should be in a format relevant to complete the appropriate process (i.e., 2 + 6 doesn't output a number 8 in your visual system, but an abstract representation of 8 that can then be divided by 4).

The Function of VLTM

VLTM is typically described in terms of episodic recollection, which is commonly defined as the conscious recollection of visual events (Squire, 2004; Brady, Konkle & Alvarez, 2011). This is a generally vague description that doesn't elucidate the function of VLTM. However, we can find insight into the function of VLTM from the study of object recognition. The goal of any object recognition system is to identify an object based on some previous input. Clearly, when VLTM studies utilize images of real-world objects in their tasks the situation is analogous to an object recognition task.

The fundamental challenge facing object recognition is the need for the system to demonstrate both tolerance and discrimination.[1] When an observer recognizes an object, the 2-D image hitting their retina will be vastly different due to changes in size,

[1] It is important to note that discrimination is sometimes referred to as "explicitness," and tolerance is sometimes referred to as "invariance" in the object recognition literature. For the purposes of this review, the terms are interchangeable.

lighting, orientation, or viewing angle across encounters (Cox, Meier, Oertelt & DiCarlo, 2005; Cox & Dicarlo, 2008; Dicarlo & Cox, 2007; Rust & Stocker, 2010; Wallis & Bulthoff, 1999). The representations supporting our object recognition ability need to be 'tolerant' in order to address this issue. Tolerance refers to our ability to recognize something despite considerable changes in inputs across encounters. For example, if you toss a ball and watch it move through the air, the image hitting your retina transforms considerably over the ball's journey. However, despite drastic changes in the appearance of the ball from moment to moment, you are still able to recognize that the ball is the same at each point in time. This is because your visual system is relatively tolerant to these changes.

While tolerance allows us to identify previous experiences despite considerable changes across encounters, there is a potential danger of over-tolerance. One could imagine that if a representation is too tolerant, it risks an observer mistaking every object she encounters as one that she's seen before. To address this problem, our representations must also demonstrate a certain level of discrimination. In contrast to tolerance, discrimination refers to our ability to distinguish similar but distinct inputs from one another.

For example, imagine one day my co-worker and I are having lunch in a break room and we each have an apple to eat. We place our lunches beside each other, but

we both exit the room to grab some napkins. When we return, there are two apples on the table that are quite similar in appearance, but one apple is mine and the other is my co-worker's. In order to know which is mine, I must have a representation that is discriminating. Despite the similarities between both apples, I should be able to identify which apple is mine.

Similar to over-tolerance, if a memory system is too discriminating it would operate poorly since an observer would likely fail to recognize any previously seen input unless it was exactly the same as when it was first encountered. Given the variability in our visual inputs described earlier (i.e. differences in size, lighting, orientation, etc.), as well as the general assumption that our visual experience is noisy, this is simply an impossible threshold to meet. As a result, there's a natural tension VLTM must manage between tolerance and discrimination. This is the function of VLTM – to manage this tension in order to recognize previous visual experience given new input.

1.3 Core Concepts of VLTM

As discussed previously, VLTM is typically defined as an automatic storage system of visual episodic memory. The function of this system is to identify a past visual experience given some previous input, while balancing the need for both tolerance and discrimination. Having defined VLTM and identified its functions, we can expand

our understanding of VLTM by establishing what core concepts researchers have identified in the field.

Familiarity and Recollection

When trying to understand the nature of VLTM, researchers have focused on investigating the kinds of information VLTM may utilize. One particularly influential distinction in VLTM (and long-term memory research generally) has been recollection and familiarity. Familiarity refers to an observer knowing an item is old or new, without specific details associated with that memory - "familiarity is the process of recognizing an item on the basis of its perceived memory strength but without retrieval of any specific details about the study episode" (Diana, Yonelinas & Ranganath, 2007, p. 379). This notion of familiarity is quite common, and perhaps something we have all experienced. Say you happen to have gone to a friend's house for a party the other night, and interacted and met with a variety of new people. In particular, you happened to meet a man named Bob, whom you talked at length about the Chicago Cubs. The next day, you enter a coffee shop and as you get in line you happen to see that Bob is in front of you. You instantly recognize Bob's face as someone you know, but you can't remember any of the details related to this knowledge. This is familiarity.

In contrast to familiarity, recollection refers to an observer accessing specific details about a previously experienced item. Given the above example, when you see

Bob in the coffee shop you remember all the specific details associated with him – his funky hairstyle, the color of his eyes, along with all the details of when you last met (your conversation about the Cubs, etc.).

DPSD Model. One model used to explain recollection and familiarity processes is the dual-process signal-detection (DPSD) model (Yonelinas et al., 2010). The DPSD model specifies that familiarity is primarily a signal detection process of discriminating between two Gaussian distributions between old and new items. Therefore, your ability to be familiar with something you've seen before depends on the overlap between the signal of the original memory and the new input. Experiences that are very different should produce very little to no overlap, so familiarity should not occur. But given visual inputs that are highly similar (even if you have not necessarily encountered them before), this will likely result in a familiarity signal (Figure 1).

The DPSD model defines recollection as a threshold-based process. In order to recollect an item, an observer must collect enough information. Once enough information is collected it has passed the "threshold," and specific details will then be recollected. These classifications are exclusive and are an "all-or-nothing" process. It does not matter if an observer collects more information at this point, as long as the threshold has been passed. However, if an observer does not accumulate enough information the threshold is not passed and recollection will not occur. Figure 1

schematizes this process. For example, let's imagine you have a memory associated with your favorite coffee mug, which has a certain recollection threshold. If you see your coffee mug clearly on your desk you receive a strong memory signal (Stimulus A) so the threshold is passed and your memory is recalled. The next day you see the coffee mug on your desk, but it's in a different place and is partially occluded (Stimulus B). While the signal is not as strong, you still gain enough information that it passes the threshold and your memory is recalled with the same amount of detail as the day before. However, later in the week you see your coffee mug from very far away (Stimulus C). While you do receive some memory signal, it's not enough to pass your threshold and so you don't recollect the memory.

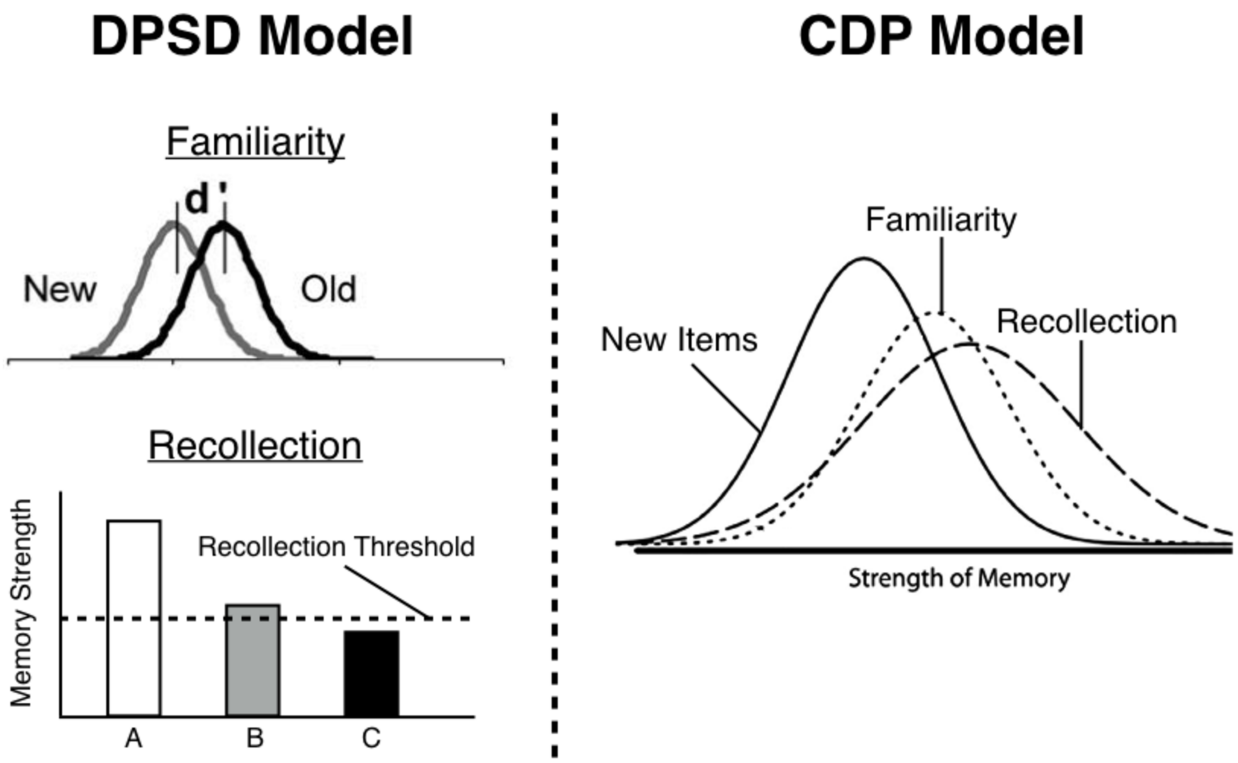

Figure 1. Illustration of the DPSD and CDP models. In the DPSD Model, familiarity and recollection are two distinct but parallel processes. Familiarity is a signal detection process of discriminating between two Gaussian distributions between old and new items. When there is more overlap between the distributions, familiarity is more likely to occur. Recollection is a threshold-based process where signal strength passes a certain threshold and is recollected or not. In the diagram, both stimuli A & B pass the threshold and would thus be recollected with the same amount of details, regardless that each stimulus may illicit different amounts of memory strength. Stimuli C does not pass the threshold, and would thus not be recollected. In the CDP Model, both familiarity and recollection vary continuously and operate using signal detection based processes. These processes are interactive and are combined during decision making.

Given the DPSD model, familiarity and recollection are two distinct but parallel processes. Despite their distinction from one another, humans likely utilize both processes in a variety of long-term memory tests and behaviors. Familiarity or recollection would be sufficient to recognize something you've seen before. For

example, in order to recognize someone you've previously met you could rely on familiarity (i.e. an associative feeling that you've seen that person before) or recollection (i.e. specific details related to that person). Along similar lines, many common long-term memory tests ask observers to classify images as ones they saw during encoding ("old") or as completely new images. An observer could use either familiarity or recollection to correctly classify an image as "old".

CDP Model. There exist numerous studies in support of the DPSD, extending past humans (Bowles et al., 2010; Yonelinas et al., 2010) to research involving rodents (Fortin, Wright & Eichenbaum, 2004) and monkeys (Miyamoto et al., 2014). However, this is not the only computational approach to understanding the possible distinction between familiarity and recollection. For example, Dede and colleagues (2014) suggested these processes are better described computationally by the continuous dual process (CDP) model. In this model, familiarity operates exactly as defined in the DPSD model. However, unlike the DPSD, the CDP model proposes recollection can vary continuously using the same signal detection process as familiarity. This means recollection isn't an all-or-nothing process, and that different memories that illicit different strengths will vary in detail. Furthermore, during memory decision-making, familiarity and recollection are combined and are thus interactive (in contrast to the DPSD, which as stated previously assumes they are independent but operate in parallel).

Both the CDP and DPSD models have similar predictions in terms of performance in recognition tasks, but vary according to the types of systematic errors they would produce. By modeling systematic errors, Dede and colleagues (2014) found support that human data was better fit by the CDP model than a DPSD equivalent. This demonstrates the CDP model is consistent with previous research in support of the DPSD model, but may more accurately capture the specific computations used in recognition decision-making. It also alters the previous distinctions between familiarity and recollection processes in positing they are interactive, rather than orthogonal.

Potential Limitations. It is unclear if researchers intend for the familiarity/recollection dichotomy to designate different types of representations, or different affordances from a single VLTM representation. In general, a feature of these models that leaves them difficult to interpret is that they are content-neutral; the same models are applied in the same way regardless what is to be remembered, and without any claims about how those things are described in symbolic or activation terms.

For example, it is possible that familiarity and recollection reflects memories with different contents, something like a gist (familiarity) and a more detailed representation (recollection). Yonelinas and colleagues (2010) appear to endorse this kind of view, as they argue that recollection and familiarity are separate process that make independent contributions to recognition memory. This is an intuitive conclusion under

the DPSD model, as it specifies familiarity is a signal-detection based process, whereas recollection is a threshold-based process, and thus likely differ in their content and format. This would suggest that individuals are able to make accurate familiarity-based judgments even when detailed representations fail to consolidate.

But under this view, we should then want specific accounts of the contents and formats in each representation. What goes into a gist? One possibility is that it's simply some sort of categorical or semantic knowledge for something you've seen before. Such knowledge would be able to interact with more detailed representations, allowing for observers to utilize both types of information when making memory judgments. However, defining gist in this way seems problematic, since almost any information related to a category should therefore create some familiarity signal.

By defining recollection and familiarity as exclusive processes, this also creates potential limitations in the kinds of information gist and more detailed representations can provide. Can an observer have familiarity with some aspects of an object and recall others? Given a DPSD framework, this is not possible. Familiarity is recognition without any specific details, and recollection is an all-or-none process. What happens then when an observer confuses a previously seen object with a similar-looking item? Does this mean that they incorrectly recollected that object? Or does it mean that they relied

only on familiarity to make their judgment? It is not clear in the DPSD model what information an observer may be relying on when making such errors.

In contrast, the CDP model would be able to account for such differences, as recollection can vary continuously (and is thus not all-or-none), and is also integrated into familiarity-based information when making a decision. Given that both familiarity and recollection can vary continuously, this creates more flexibility for different kinds of memory performance. For example, an observer could recollect varying amounts of specific details related to an object.

However, the CDP model has its own limitations. It defines familiarity and recollection as varying along the same memory continuum. They are separate processes operating in the same terms (i.e. signal detection), but are integrated when making a decision. Why then are we defining these as separate processes? Under this framework, it becomes less clear how familiarity and recollection are distinct. It may be that the CDP is simply describing a single memory process operating along a continuum of memory strength. Again, it would be useful to specific the contents and formats of the representation in some detail to understand how it produces the response patterns taken to indicate familiarity and recollection.

Pattern Separation and Pattern Completion

Another way researchers have tried to understand the nature of VLTM is to take an approach informed by neuroanatomy. The idea is that certain brain areas related to memory have specific properties that may make them amenable to completing different types of processes. Pattern completion and pattern separation is one such model seeking to explain distinctions in memory, but also supported by brain localization properties.

Pattern completion is the process by which incomplete or degraded signals are filled-in based on previously stored representations (Yassa & Stark, 2011). After you see an object, the next time you encounter that object the image hitting your retina is not going to be exactly the same (due to differences in viewpoint, orientation, lighting, etc). Pattern completion is the process that would assist in VLTM being tolerant to variability in inputs related to the same object across encounters.

Pattern separation is the process that reduces the overlap between similar inputs in order to reduce interference at later recall (Yassa & Stark, 2011; see Figure 2). This process supports the discriminatory ability of VLTM. It seeks to parse overlapping signals into distinct representations in order to assist in the individual recall of both items.

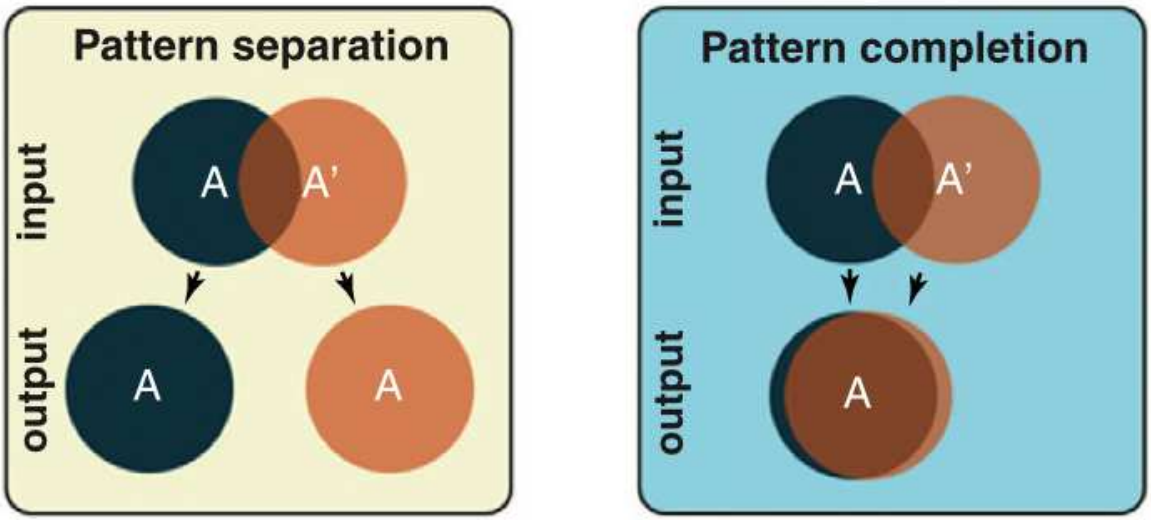

Figure 2. Illustration of pattern separation and pattern completion processes. Pattern separation is the process by which two overlapping signals are recognized as distinct from one another. Pattern completion is the process by which two overlapping signals are recognized as being from the same source and are combined into a single output or representation.

This model was in part designed to address the issue that we want to be able to remember certain events as related but other events as distinct (i.e. tolerance and discrimination). However, given a limited number of neurons in the brain, there may be too many patterns for the representations of events to be distinct – too much overlap. Thus, even when an observer is given different stimuli there might exist considerable overlap between the inputs into the system, causing different items to be incorrectly be identified as related to one another (i.e. the problem of saturation). One possible way to address this issue would be to make the input representation very sparse, with a large and distributed network of neurons creating multiple potential pathways of

activation for inputs. The hippocampus, which previous examples have shown is critical to episodic and recognition memory, has a region with such properties – the dentate gyrus (DG). This is consistent with pattern separation, which seeks to separate overlapping representations as distinct from one another.

Another issue this model sought to address is to explain how we remember certain events as related to one another. Specifically, what is the process that recognizes two partially overlapping signals as arriving from the same source? One possible solution would be to have a set of pathways that feeds back into itself to fill in potential missing information, so that partial or incomplete signals can be recognized. Another area of the hippocampus, CA3, has such properties, through recurrent collaterals that form a feedback loop (i.e. the axons of neurons in the circuit circle back to the inputs [dendrites] of neighboring axons), consistent with pattern completion computations (Yassa & Stark, 2011).

Even though there can be many different computations in support of distinctions in memory, the framework of pattern separation and pattern completion is made more viable by taking into account the specific properties of brain areas that may implement such processes. In fact, recent evidence using rats in a spatial maze has shown that the CA3 demonstrated coherent responses during conflict (when the global cues of the maze were rotated but local information remained the same), consistent with pattern

completion, whereas the DG demonstrated disrupted responses during such conflict, consistent with pattern separation (Neuneubel & Knierim, 2014). While the content of these processes are not specified, it provides a useful example for how a distinction supported by brain localization properties may inform our understanding of memory.

Potential Limitations. Unfortunately, many aspects of pattern separation and pattern completion remain unspecified. This results in a host of potential limitations. Often this framework is referred to as a "neurocomputational" approach, but the computations of these processes are never specified. There are many possible ways that one could design a system to parse or complete the same overlapping inputs that would result in vastly different outputs. In addition, the content and format of these inputs are never specified. This information would greatly affect the kinds of computations one might use to address these types of processes.

Pattern separation and pattern completion are also limited by its focus to explain memory based processes exclusively on properties of the hippocampus. Pattern completion shares similarities with familiarity, as both processes could be used to explain how an observer might recognize a previously seen image with degraded input. But pattern completion is specified to occur in the CA3 region of the hippocampus. Thus, it is unable to explain why amnesic patients with brain damage to

the hippocampus can still demonstrate unimpaired associative (or familiarity-based) memory performance (Gabrieli et al., 1995).

1.4 Common Methods in VLTM

Old/Similar/New Judgment

One way to measure the discriminatory ability of VLTM is to utilize a paradigm often referred to as the Old/Similar/New Judgment. Participants first view and encode images of objects, typically while doing an incidental encoding cover task (i.e. would this object fit in a shoe box?). At test, participants serially view objects that were present in the encoding task (Old images), objects similar but not identical to ones in the encoding task (Similar images), and completely new objects that were not present during encoding (New images). Participants then judge whether the images they see are old, similar, or new.

The primary purpose of this task is to evaluate the discriminatory ability of VLTM, and how memories may acquire different levels of discriminability. For example, for an observer to correctly classify a similar item it is assumed to require more specific details in memory (i.e. recollection-based) than correctly identifying an old item, which could be accomplished using a gist- or familiarity-based process (Kensinger, Garoff-Eaton & Schacter, 2006; Stark, Yassa, Lacy & Stark, 2013; Kim & Yassa, 2013).

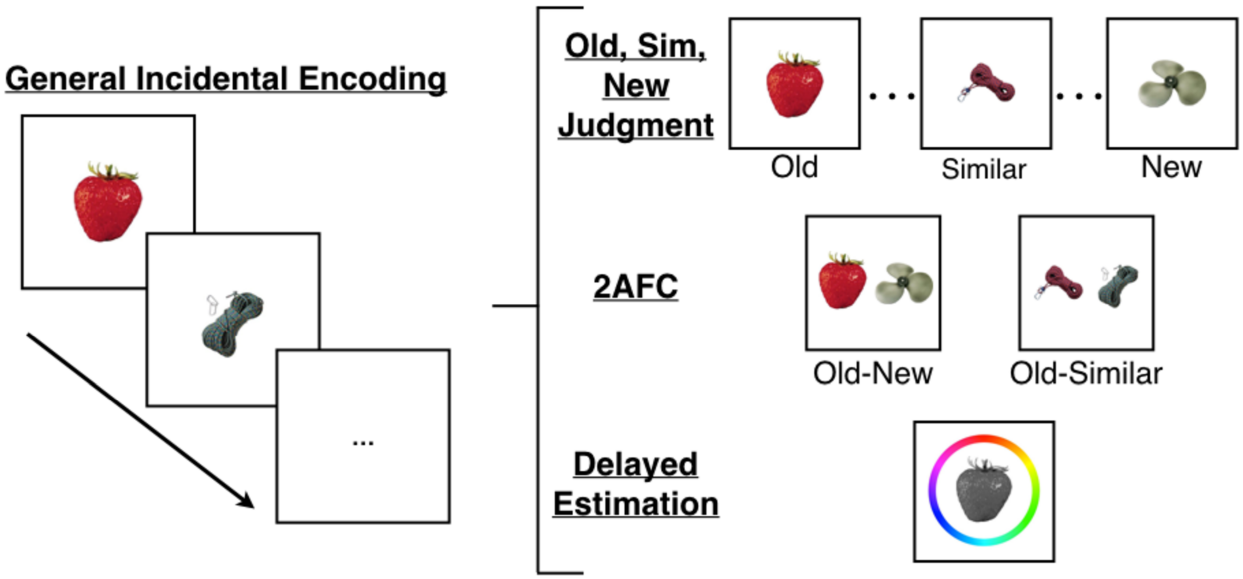

Figure 3. Illustration of typical VLTM methods. In a general incidental encoding task, participants see a stream of images of real-world objects and make some judgment about those objects (i.e. indoor/outdoor, does this fit in a shoe-box?). After they may be tested using several methods. Using Old/Similar/New Judgement, participants are shown objects that were exactly the same as encoding (Old), similar but not identical to images at encoding (Similar) and completely new images (New), and are asked to classify them accordingly. In 2AFC tasks, they are shown two images, one they've seen before and either a completely new (Old-New comparison) or similar looking object (Old-Similar comparison), and are asked to judge which of the two images they've previously seen. In delayed estimation tasks, participants are initially shown a grayscale image of a previously seen object and are asked to report its color using a color wheel.

How Does Emotion Affect Visual Memory? One study that utilized this method sought to investigate how negative emotional context may affect the likelihood of remembering an item's specific visual details. Participants first completed an incidental encoding task, where they were exposed to hundreds of images of real-world objects (for 250 or 500ms) and had to judge whether each object would fit in a shoebox. Half of the images were rated as being negative and arousing, and the other objects were

rated as neutral. Two days after incidental encoding, participants were then given a surprise test. Participants viewed old images (exactly the same as encoding), similar images (similar but not identical), and completely new images relative to encoding. They were told to classify them accordingly (old/similar/new) (Kensinger, Garoff-Eaton & Schacter, 2006).

At test it was observed that for old items, negative emotional context led to an increase in correct classification. This was true for items presented both for 250ms and 500ms, but was stronger as encoding time increased. However, for similar items there was no main effect of emotional context or encoding time. As a result of the main effect of emotional context for old images, the researchers concluded that negatively arousing content increased the likelihood that visual details of an object would be remembered (Kensinger, Garoff-Eaton & Schacter, 2006). However, given the lack of an effect for similar images, it would be more accurate to conclude that emotion may have enhanced certain aspects of visual memory (i.e. for old images).

How do Familiarity and Recollection Contribute to Responses? This method has also been used to investigate the relative contributions familiarity and recollection processes may have in the behavioral responses of old, similar, new judgments. In the study, participants completed a two-stage recognition test. In the first phase, participants saw 128 images of real-world objects on a computer screen for 2 seconds

each, and were asked to report whether an object was an "indoor" or "outdoor" object. In the second phase, participants were given a surprise test, where they viewed old images (exactly the same as encoding), similar images (similar but not identical), and completely new images. They were told to classify the images as either old, similar, or new. Additionally, after indicating what category an image belonged to, participants were then instructed to indicate whether they "remember" seeing the same image in the study session or if they just "know" that they have seen the same image without any conscious recollection of its original presentation (Kim & Yassa, 2013). It is theorized that "remember" judgments reflect recollection-based processes, whereas "know" judgments reflect familiarity-based processes, although there exist several criticisms that these are not true indices of these processes but rather reflect subjective states of awareness or differences in confidence (Yonelias, 2002).

As expected, at test they found very different accuracy for classifying old (70% correct), similar (53% correct) and new (74% correct) images. When analyzing these responses according to whether a participant reported they "remembered" (recollection-based) or "knew" (familiarity-based), a few interesting patterns emerged. When judging old items correctly, observers made primarily "remember" responses, suggesting that correct classifications of old items was primarily driven by recollection. For similar items, there was a slight trend to report "remember" rather than "know" both when the item was judged as old or similar. This suggests that observers can

classify similar items with or without recollection, and that incorrectly identifying similar items (i.e. misclassifying them as old) is not simply driven by familiarity signals (Kim & Yassa, 2013).

Source Localization Judgments

One approach to evaluate the strength of memory and further distinguish potential errors is to use a paradigm combining Old/New judgments with source localization. At encoding, participants view images of objects typically presented in one of four quadrants in the display. Then, at test, participants are shown old, similar and new images in the center of the screen they must classify as "Old" (previously seen images) or "New," (similar or new images) and indicate which of the four quadrants the object originally appeared in (Cansino, Maquet, Dolan & Rugg, 2002; Reagh & Yassa, 2014a).

The underlying assumption of source localization manipulations is that more episodic information is retrieved on trials when the source judgment was successful than on trials when it was not. This is similar to the logic behind the "remember/know" procedure discussed previously. When an observer makes a correct classification and source judgment, the assumption is that this indicates a recollection-like memory, whereas if an observer makes a correct classification of an image but incorrect source judgment, this may indicate a familiarity-based memory. However, unlike the

"remember/know" procedure, source localization judgments do not rely on the observer's own introspection in order to classify memory quality. Thus, the aim of the paradigm is to evaluate what percentage of responses using Old/New judgments may rely on memories that contain more or less information, and what brain areas may be involved in these processes.

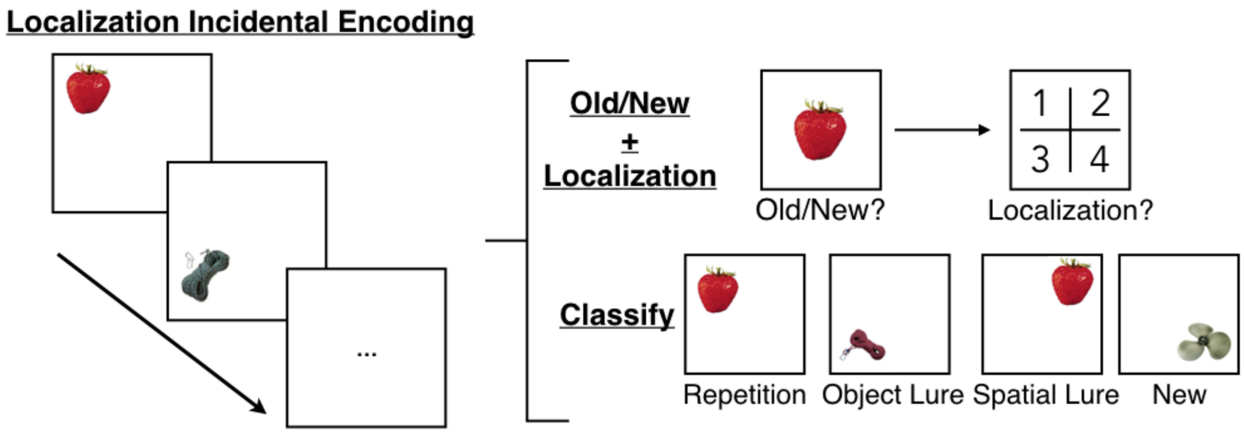

Figure 4. Illustration of typical localization VLTM methods. In a general incidental encoding task, participants see a stream of images of real-world objects and make some judgment about those objects (i.e. does this fit in a shoe-box?). Critically, every image is presented in one of four (or more) possible quadrants. After, they may be tested using several methods. Using Old/New and Localization tasks, participants are shown objects that were exactly the same as encoding (Old) and similar-looking or completely new images (New), and are asked to classify them accordingly. If they classify an object as "Old," they are then asked to indicate which of the four quadrants the image originally appeared in. In classification tasks, participants are shown either repetitions (old image, same location), object lures (similar images, same location), spatial lures (old image, different location), or new images. They are asked to classify the images accordingly.

What Brain Areas and Behaviors Are Involved in Source Judgments? Cansino and colleagues (2002) were interested in using this method to investigate what brain

regions may be involved in different memory processes beyond simple item recognition tasks. To accomplish this, participants first viewed images of real-world objects, and were asked to judge whether the objects were "natural" or "artificial." Critically, each image was presented in one of four quadrants in the display. After completing the task, participants were then administered a surprise test where previously shown images (old images) were mixed with completely new images. These images were shown in the center of the screen. Participants had to judge whether each image was old or new. They were instructed to press a single key if an image was new, and if an image was old participants indicated which position the image was presented during encoding using one of four keys. If a participant didn't know which quadrant an old image originated from, they were instructed to guess.

At test it was found that when classifying previously seen (Old) items, observers correctly identified 87% of items presented. However, 60.7% of these responses contained correct source responses, and 26.3% of responses contained incorrect source responses. This suggests that even when classifying Old and New objects in a typical retrieval task, memories for these items likely contain additional information beyond simply categorical or familiarity-based knowledge. Additionally, they observed via collected fMRI data that when recognizing an old object with a correct versus incorrect source judgment there was greater activity observed in the right hippocampus and left prefrontal cortex (Cansino et al., 2002). This suggests that

memories containing more information may elicit greater memory signals and decision-making coordination.

Potential Neurocorrelates of "What" and "Where" Memory. Source localization judgments have also been used to explore possible dissociations between object (what) and spatial (where) memories and their potential neural correlates. In the experiment, participants first completed an encoding task where images of real-world objects were presented in one of 31 possible locations on the screen for 3 seconds each. They were instructed to first judge whether the object was an "indoor" or "outdoor" object, and then whether the object appeared on the "left" or "right" relative to the center of the screen. Afterwards participants were given a surprise task, and were shown four possible trial types: repeated images (old images in the same location), lure images (similar images in the original object's location), spatial lure images (old images in a slightly different location), or new images (not shown during encoding). Participants were instructed to indicate whether an image showed "No Change," "Object Change," "Location Change", or "New" (Reagh & Yassa, 2014a).

Behaviorally, there was no difference in lure discrimination whether it was an object trial (i.e. similar image) or spatial trial (i.e. old image in slightly different location). This effect was consistent across both high similarity and low similarity stimuli. Neuroimaging data was also collected via fMRI, and demonstrated unique differences

based on lure type. It was observed that the lateral entorhinal cortex (LEC) was more engaged during object lure discrimination than during spatial lure discrimination, whereas the opposite pattern was observed in the medial entorhinal cortex (MEC). Additionally, the perirhinal cortex (PRC) was more active during correct rejections of object than spatial lures, whereas the parahippocampal cortex (PHC) was more active during correct rejections of spatial than object lures. Regardless of lure type, the dentate gyrus (DG) and subregion CA3 demonstrated greater activity during lure discrimination. Overall, this suggests two parallel but interacting networks in the hippocampus and related regions for managing object identity and spatial interference (Reagh & Yassa, 2014a).

Two-alternative forced choice test (2AFC)

A paradigm referred to as the two-alternative forced choice (2AFC) test has been primarily used to study the capacity of visual long-term memory. In a typical test, observers see two objects on the screen during test, one that they have seen before and another object they have not encountered previously. The other may be completely novel (Old-New comparison) or a similarly related lure (Old-Similar comparison) (Brady, Konkle, Alvarez & Oliva, 2008; Brady, Konkle, Oliva & Alvarez, 2009; Konkle, Brady, Alvarez & Oliva, 2010). The logic of this test is that it can tap into even "weak" memories. It makes sense that a 2AFC judgment is easier than other

kinds of responses because it's a binary response. Indeed, research has shown that given identical encoding, observers demonstrate better performance when given a 2AFC versus a standard Old/New judgment task (Cunningham, Yassa & Egeth, 2015).

Almost any strategy a memory system might use should be able to utilize the additional information provided by the new image in order to improve performance. Even an incredibly simple strategy would vastly improve its performance. For example, one could design a memory system that simply creates a color histogram of images it is presented (i.e. what color and how much of it is present). If that memory system was given an Old/Similar/New memory task, it would likely not have very good performance, especially in relation to similar images that were presented. However, 2AFC performance would be considerably better, unless two images presented happen to be similar in their color histogram.

Typically, 2AFC is conceptualized using a signal detection theory framework (Green & Swets, 1966; Loiotile & Courtney, 2015). The logic is that observers have a memory representation which creates a normally distributed signal in a "memory strength" space. At test, when observers are shown an old item, that item elicits a normally distributed signal, which due to noise or the strength of the memory it is being compared to may demonstrate more or less overlap with the original representation. If an observer was simply shown the old object, depending on the overlap and decision criteria (how liberal or conservative an observer is being with their

decisions) this may result in incorrectly identifying an old object as "new". However, by giving observers a foil in the 2AFC task, whether that foil is a completely new or similar-looking object, this gives observers a third normally distributed signal to assist in the comparison process. This third signal should be further away from the original memory trace than the signal provided by the old image, thus allowing observers to improve their performance and correctly identify the old image (Loiotile & Courtney, 2015; see Figure 5). This framework demonstrates from a modeling perspective why 2AFC tasks should be easier, and are also able to tap into "weaker" memories for items that may otherwise fail to be remembered or classified correctly in other types of memory tasks.

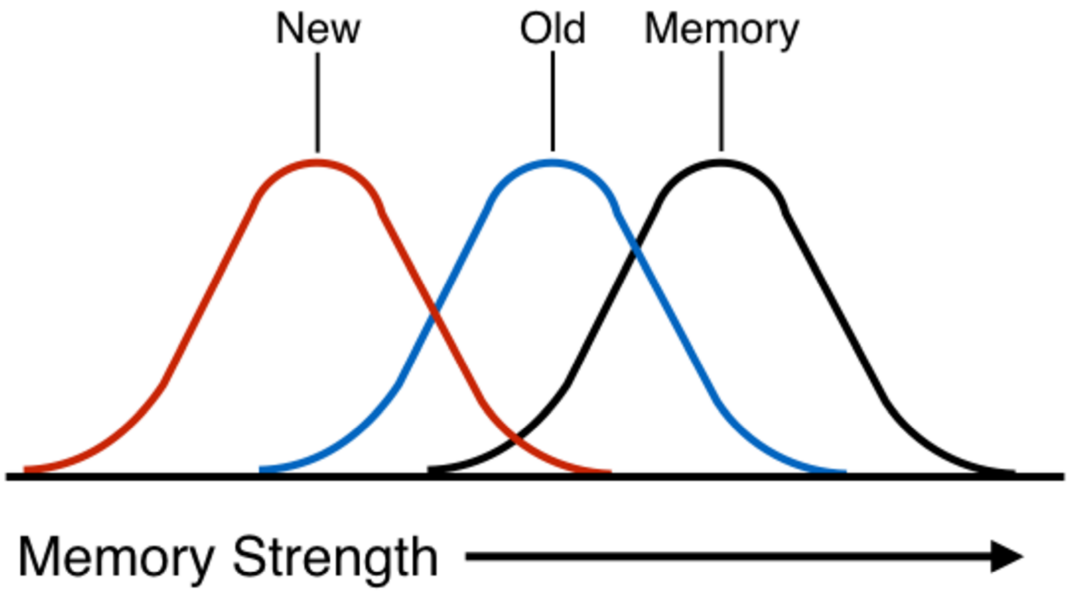

Figure 5. Visualization of 2AFC logic. Observers have a memory representation which creates a normally distributed memory signal in a "memory strength" space. At test observers are shown both an old and new image, which elicit their own normally distrusted signals of memory strength. While the old item signal may not completely overlap with the memory signal, it should share more overlap than the new item signal, and thus improve memory performance compared to if observers were only shown the old image.

In addition to traditional 2AFC tasks, which involve an old item paired with a new or similar-looking item, researchers have also expanded on the number of potential options, creating 3AFC and 4AFC type tasks. Generally, these tasks involve the addition of multiple similar-looking lures at test, in order to further evaluate the ability to discriminate between different kinds of lures. A potential limitation of 3AFC or 4AFC tasks is that adding lures may create interference and increase task difficulty (Holdstock et al., 2002). Additionally, when varying the kinds of lures available at test (such as providing two similar lures, or a similar lure and a completely new image), this creates conditions where the information available to an observer is not equivalent (Guerin, Robbins, Gilmore & Schacter, 2012). This means performance across different conditions cannot be directly compared to one another.

The Capacity of VLTM. Brady and colleagues (2008) sought to investigate the capacity of VLTM using a 2AFC method. Participants were presented 2,500 images of real-world objects for 3 seconds each and told them to remember all the details of each image. After completing this study portion, participants were then given a 2AFC task where they saw two images on the screen. One was a previously encountered image from the previous session, whereas the other was either a novel image, an exemplar of an object they had previously encountered, or an image of an object they had previously encountered in a new state (i.e. changed orientation). Participants were instructed to indicate which of the two images they had previously encountered.

Overall, performance was quite high, with significantly better accuracy for novel comparisons (92% correct) but still extremely accurate performance for state and exemplar comparisons (87-88% correct) (Brady, Konkle, Alvarez & Oliva, 2008). These results demonstrate that even when given very brief exposure of images, humans are able to remember thousands of objects (seemingly with no limit) with extremely high accuracy. Furthermore, they suggest that human's VLTM representations contain visual information necessary to assist in making difficult state and exemplar comparisons, beyond simply categorical or semantic knowledge of previous encounters.

Delayed Estimation (Continuous Report)

In order to estimate the fidelity of VLTM (i.e. the amount of information in memory) experiments have utilized a delayed estimation paradigm. At encoding participants observe objects embedded in unique colors. Then, at subsequent test, observers see grayscale versions of objects they observed previously and use a color wheel to indicate its original color. By taking the error in degrees between the response and the true value, researchers can create a distribution of long-term color memory responses and measure the standard deviation of the distribution to understand the fidelity of that representation (Brady et al., 2013).

The Precision of Information in VLTM Representations. Brady and colleagues (2013) used this method to understand the precision of color memory representations

across VWM and VLTM. In the study, researchers gave participants two separate tasks. In the VWM condition, participants saw three real-world objects simultaneously for 3 seconds, arranged in a circle around fixation. Participants were instructed to remember the color of all the objects. After a 1 second delay, one of the objects reappeared in gray scale, and participants could alter the color of the image using their mouse and were told to click the mouse when it matched the original color. In the VLTM condition, participants first underwent a study block viewing images sequentially for 1 second each, with a 1 second blank interval between images. Similar to the VWM condition, participants were instructed to remember the color of the object.

After the study block the color of the items was tested one at a time in a randomly chosen sequence and participants reported the image color using the same response mechanism used in the short-term memory condition. The precision of participants' memory representations was determined by calculating the distribution of the degree error of each response in color space (with larger mistakes represented by greater degree error). In the VWM condition at set size 3 (and above) participants' precision was 17.8 degrees, which did not significantly differ from the precision observed in the VLTM condition, 19.3 degrees (Brady et al., 2013). Therefore, it appears the precision of color representations across VWM and VLTM have equivalent limits, suggesting they may share or be constrained by similar processes.

1.5 Core Concepts of VWM

As previously discussed, VWM is typically thought of as the interface of multiple processes including perception, short-term memory, and attention (Baddeley & Hitch, 1974; Cowan, 2008). Due to this conception, researchers have generally described the function of VWM as supporting complex cognitive behaviors that require temporarily storing and manipulating information in order to produce actions (Baddeley, 2003; Ma, Husain & Bays 2014). In particular, a large body of research over the past decade has focused on the capacity limitations of VWM. As a result, many of the models proposed to explain VWM have focused on this limitation.

Fixed Slot Model

When trying to understand the inherent capacity limitations of VWM, a particularly influential model has been the "Fixed Slot Model". It suggests that VWM can only store a discrete number of integrated object representations. This model was proposed by a highly influential study conducted by Luck & Vogel (1997), who used a change detection task to quantify VWM capacity. In the task, participants were instructed to remember an array consisting of items of a single or conjunction of features (color, orientation, etc). After a brief delay (900ms), a test array was presented that was either identical to the previous array, or differed in terms of a single feature. Participants were instructed to indicate whether a change had occurred. Accuracy was

assessed as a function of the number of items in the stimulus array in order to determine how many items could be accurately maintained in VWM.

In a series of experiments, Luck & Vogel (1997) gave participants change detection tasks that varied the number of colored squares presented in an array (1-12). They observed that performance was at ceiling for arrays of 1-3 items, and then declined systematically as set size increased from 4-12 items. Overall, the average K value (estimations of VWM capacity) among participants was around 3-4 items. This finding led to the foundation of the slot model, which posits individuals can only store between 3-4 objects in VWM.

In addition to experiments consisting of arrays of single features, Luck & Vogel (1997) also presented participants with arrays consisting of a conjunction of features (i.e. lines differing in orientation and color). Participants completed a change detection task, but the researchers varied whether participants had to remember a single feature or a conjunction of features. For example, participants would see an array consisting of lines of different orientations and colors. In the color condition only a color change could occur, and participants were instructed to look for a color change. In the orientation condition only an orientation change could occur, and participants were instructed to look for an orientation change. And in the conjunction condition, either a color or orientation change could occur, and participants were instructed to remember both features of each item. Thus, in the conjunction condition, participants had to

remember 8 features but only four integrated objects. If VWM storage capacity is limited by individual features, performance should decline at lower set sizes in the conjunction compared to the single feature conditions. However, if VWM storage capacity is limited by integrated objects, then the same pattern of results should be observed throughout all three conditions.

Consistent with the latter case, they observed that VWM capacities were the same for single feature and conjunction items. Altogether, this provided the basis for the Fixed Slot Model that VWM capacity was constrained by slots of ~3-4 integrated objects. While further research has expanded upon these initial findings, it's also important to note that several experiments have failed to replicate the critical conjunction experiments (Wheeler & Treisman, 2002; Olson & Jiang, 2002). In particular, these studies have observed that storing two colors of one object was much more difficult (i.e. lower memory performance) than storing one color feature.

A key component of this model is that these VWM slots are considered "all-or-nothing" – an observer either remembers every object with the same fidelity (i.e. amount of information) within the capacity limit, or fails to remember the object completely. This all-or-nothing component is potentially problematic, as it suggests that an observer has the same amount of information per item whether they viewed single or multiple items. What if an observer sees two encoding displays in a typical

change detection task: one with a single image of an apple and another with four images of very similar looking apples? In each test array, there is either no change or one of the objects is replaced with another very similar looking image of an apple. According to the Fixed Slot Model, the precision of information available to the observer is the same in both conditions. It does not account for the potential interference four very similar objects may have in terms of their representations in VWM, or how they may affect the decision-making process when an observer decides whether a change has occurred.

Continuous Resource Model

Another model used to explain apparent limitations in VWM is referred to as the "Continuous Resource Model". Unlike the Fixed Slot Model, which defines capacity as being constrained by all-or-nothing slots of integrated objects, the Continuous Resource Model conceptualizes VWM capacity as information based and limited by a finite resource. Furthermore, this finite resource can vary across different representations. This unequal division of resources across representations can differ due to a variety of factors, such as top-down goals (i.e. attention) or the total information load of the display (i.e. set size) (Wilken & Ma, 2004; Bays & Husain, 2008).

Support for the Continuous Resource Model first came from Wilken and Ma (2004), who developed the continuous report method as a way of measuring the

fidelity, or amount of information, contained in VWM representations. In the continuous report paradigm, participants are instructed to remember an array consisting of items of a single feature (color, orientation, etc). After a brief delay (1.5 seconds), a square cue appeared centered at the location of one of the previously presented items. At the same time, a test probe is displayed in the center of the screen, which allows for continuous report of the probed feature. For example, in the color condition, a color-wheel containing all possible color values appears in the center of the screen, and participants indicate the color of the probed item by clicking a color on the wheel using a mouse. Responses are then reported as the degree error from the true color value, and a distribution can be made based on a participant's responses. The standard deviation (SD) of this distribution can then be used to estimate a participant's precision of color information in VWM.

In a series of experiments, Wilken and Ma (2004) varied the set size of displays, as well as the type of feature being probed in memory. Regardless of the type of feature probed, they observed that as set size increased, the precision of VWM representations decreased. However, even at large set sizes the distribution of responses was still centered around the true value of an item, and the precision (i.e. SD) was still well above chance. This led the researchers to conclude that individuals could store a continuous amount of information in VWM, but the precision with which an individual item was represented varied as a function of the total information load of

the display (i.e. set size). When set size is greater than four items, observers are able to store more than four items in memory. However, fewer resources are able to be allocated to each item. Given the constraints of a typical change detection task, an item may still be represented but there isn't enough information in order to make a successful comparison. This suggests that the ~3-4 object limit proposed by the Fixed Slot Model may simply be a behavioral artifact of the tasks being used to assess such limits.

A potential limitation of the research used to support the Continuous Resource Model is that variable precision has not been shown for holistic representations of objects. Studies in support of this model typically probe memory along a single feature dimension (i.e. color), even if observers are viewing images of real-world objects. It remains possible that representations for objects in VWM rely on different kinds of information, some of which may be variable but others that may not. For example, when an observer sees a single image of a real-world object, their representation of that object may contain categorical knowledge (i.e. teddy bear) in addition to other knowledge such as color (i.e. brown). The information of color available to the observer may be variable as a function of set size, but it is likely that categorical knowledge is not – observers either know the categorical identity of an object or they don't.

Work by Schurgin and Flombaum (2015) provides evidence this may be the case. In their task, participants saw two images of real-world objects in a display, that were then briefly masked, and participants then had to make a 2AFC judgment containing a previously seen object and a completely new object (indicating which of the two was the old object). Critically, they added image noise to stimuli at test by randomly scrambling up to 75% of the pixels in each image. They observed that VWM performance was unaffected by noise at test, with 0% noise and 75% noise demonstrating the same level of performance. The Continuous Resource model predicts that noise at test should make comparisons in memory harder, as observers would be comparing a noisy internal representation to a noisy external representation (i.e. test stimuli). Thus, when noise is added at test performance should decrease. In contrast, the Fixed Slot model would predict no performance decrease. Observers would have no noise in their internal representation, which would allow them to manage noise at test when making a comparison. Altogether, this work suggests that while resolution for memory could vary along a continuous resource for a single feature, such as color, the same may not be true along all dimensions, such as holistic object representations.

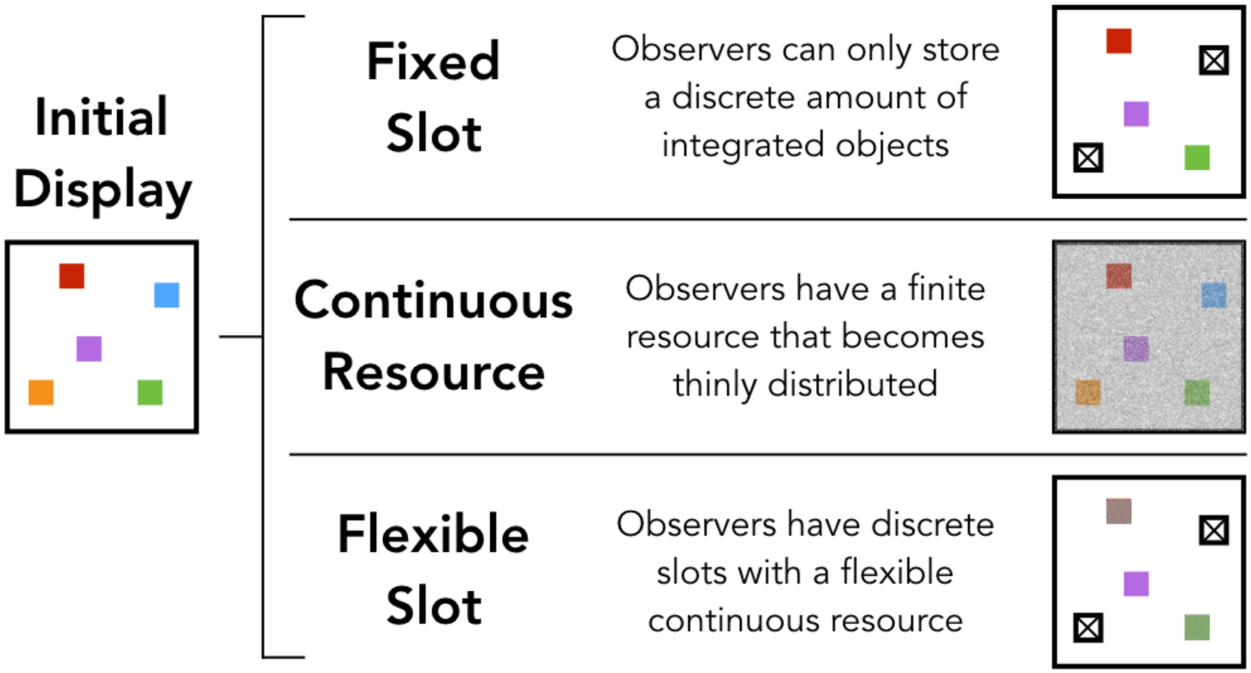

Figure 6. Illustration of VWM models. According to the Fixed Slot Model, VWM can only store a discrete amount of integrated objects. The Continuous Resource Model proposes that VWM has a finite resource that becomes more thinly distributed as the number of items in a display increases. And the Flexible Slot Model is an integration of the latter two, stating VWM has a discrete number of slots but that a finite resource can be flexibly allocated to each slot.

Flexible Slot Model

There exists contrasting evidence that may support either the Fixed Slot or Continuous Resource Model. The "Flexible Slot Model" provides a middle ground between the two, proposing that VWM is constrained to a maximum of ~3-4 representations, but that this capacity may also be limited by the amount of information load in the display. In short, VWM is constrained by "slots," but there is flexibility within the system of distributing limited resources across these slots.

One study in support of the Flexible Slot Model was conducted by Alvarez and Cavanagh (2004). They utilized a typical change detection task but varied the information load of displays by changing the type of stimuli presented. On each trial, 1-15 objects were presented for 500ms, followed by a brief delay (900ms), and then a test array. On half the trials one of the objects changed identity, and on the other half the displays were identical. Participants were instructed to indicate whether one of the objects had changed. Crucially, trials could contain stimuli pertaining to a single stimulus class that each differed in their visual complexity: line drawings, shaded cubes, random polygons, Chinese characters, and colored squares.

If VWM is limited by a fixed number of representations (i.e. slots), then performance should be equivalent across all stimulus categories, but if VWM capacity is limited by information load these estimates should vary across stimulus categories. After converting responses to K estimates of VWM capacity, Alvarez & Cavanagh (2004) observed there was varying capacity estimates across different stimulus classes, ranging from 1.6 for shaded cubes to 4.4 for colored squares. This provided support that VWM is limited in its number of representations (~4 objects), but is also limited by the amount of information (i.e. stimulus complexity) of what is being remembered.

There is disagreement as to whether these differences reflect storage limitations, therefore supporting a Flexible Slot Model, or rather reflect comparison errors made

during the decision-making process. It could be that items with higher visual complexity also have higher similarity to one another, and this will lead to greater errors at test even though overall memory capacity for items is the same. Awh and colleagues (2008) investigated this possibility using Alvarez and Cavangh's (2004) method and stimuli but with one critical change. When a change occurred at test, they could either be within-category changes (as in Alvarez & Cavanagh, i.e. a Chinese character replaced with a different Chinese character) or they could be cross-category changes (i.e. a Chinese character replaced with a line drawing). They observed that for within-category changes at test performance varied across stimulus categories, consistent with Alvarez and Cavanagh. Conversely, for between-category changes they observed no difference in performance relative to stimulus category. This suggests that variable performance across different kinds of stimuli may be due to differences in similarity, and not information load, in support of a Fixed Slot Model.

1.6 Common Methods in VWM

Change Detection Task

In order to understand the capacity of VWM many experiments have utilized a change detection task. In the paradigm, observers first see an array of items (whether a single feature, conjunction of features, or images of real-world objects) and are told to remember what they saw. The array then disappears and after a brief delay reappears

with either all the items exactly as before or with a single item having changed. Observers are asked to identify whether a change occurred or not. The general goal of this paradigm is that by varying the number of items in a display (2, 4, 8, etc) researchers can investigate what kinds of constraints may be limiting VWM capacity (Luck & Vogel, 1997; Vogel, Woodman & Luck, 2001, Xu, 2002; Vogel, Woodman & Luck, 2006; Awh, Barton & Vogel, 2007).

A variation of the change detection task uses a single probe report, showing only one item at test instead of the entire display (with or without a change). A potential advantage of single probe change detection tasks is that observers cannot use relational or summary statistical information (i.e. how much the overall display was "blueish" and whether the test display is different in the overall "blueishness") to inform their judgments. A potential disadvantage of this procedure is that if visual memory for certain stimuli (objects, scenes, etc) relies on relational information (i.e. position, layout, etc), probing a single item may remove important information and erroneously reduce performance. It's worth nothing that VWM capacity estimates have been found to be comparable whether they were based on the whole-display or a single-probe procedure (Luck & Vogel, 1997; Jiang, Olson & Chun, 2000).

Time Course of Consolidation in VWM. Vogel, Woodman and Luck (2006) were interested in understanding the time course of consolidation in VWM using typical change detection methods. By consolidation, they meant the process whereby VWM

becomes durable enough that it is not disrupted by new sensory inputs. The goal of the study was two-fold: to measure the time course of VWM consolidation and whether the rate of consolidation varied as a function of the number of items being consolidated. In the primary experiment, participants viewed arrays of between 1-4 colored squares. In order to interrupt consolidation, pattern masks at the location of the colored squares were introduced between 117-484ms from the initial onset. After a brief delay (controlled to be the same across different mask interruption timings), a test array was presented that either contained a change or no change. Participants were instructed to judge whether a change occurred. Across set sizes 2-4, performance declined when the delay between the initial encoding array and the onset of the mask was shorter, and this effect became larger as set size increased. Overall, the rate of consolidation was estimated to be ~50ms per item (although it may be faster, ~10ms per item; see Sperling, 1963). These results suggest that consolidation is limited in capacity (increasing in time as set size increases) but occurs quite rapidly.

How Object-Based is VWM? Change detection tasks have also been used to evaluate how object-based encoding in VWM may be affected by remembering objects with two features within the same dimension (i.e. two colors per item) or different dimensions (i.e. color and orientation per item). In one experiment, participants viewed arrays of mushroom-like objects that had two distinct parts: a cap and a stem. In the conjunction condition, 5 mushroom-like objects were shown, with

each cap and stem in an object being a distinct color from one another. In the feature condition, 10 separate parts (5 caps and 5 stems) of varying colors were displayed. Across both conditions participants were instructed to only monitor cap colors for a possible change, stem colors for a possible change, or both cap and stem colors for a possible change. Participants viewed an array, followed by a brief delay, and then a test array appeared with a possible change present. It was found there was no object based advantage (i.e. conjunction trials) for remembering one or both colors (Xu, 2002). This suggests that VWM does not encode visual arrays as integrated objects.

A follow-up experiment investigated this finding further, exploring whether potential object-based advantages existed for remembering two features along different dimensions. The experiment was similar to the first, but all the mushroom stems were a uniform green color. Instead of varying by color, they varied by orientation (either 45°, 90°, or 135° relative to horizontal). Again, participants viewed conjunction or feature displays, and had to monitor either cap changes (color), stem changes (orientation) or both. Interestingly, participants demonstrated better performance across all conditions in the conjunction relative to the feature display. This suggests that an object-based encoding advantage exists, but only for when an object contained two features along different dimensions (i.e. color and orientation). While VWM may operate using integrated objects (in line with the Fixed Slot Model), this is

likely constrained to when those objects contain features from different dimensions (Xu, 2002).

Olson and Jiang (2002) used a single probe change detection task to investigate a similar question of whether VWM was limited by the number of objects or features in a display. Participants first saw a display of 3 or 6 color items. After a brief delay, a single item reappeared that was either the same from the previous display or different. Participants indicated whether a change had occurred. All color stimuli were simple objects of the same type (inner square or the outer frame of a square). Crucially, there were two types of stimuli: feature (a single color, either inner square or outer frame) and conjunction (two colors, one in the inner square of the object and another in the outer frame of the object). This created three types of trials. A feature-feature trial consisted only of feature stimuli and at test a feature item was probed. A conjunction-conjunction condition consisted of both feature and conjunction stimuli, and at test a conjunction item was probed. And a conjunction-feature condition contained both feature and conjunction stimuli, and at test a feature item was probed.

It was observed that as set size increased, performance declined across all three trial types. However, performance was higher overall for the feature-feature condition than either of the two conjunction conditions, suggesting there was a cost to maintaining more colors in VWM even when those colors were bound to a single

object. Additional experiments involving color and orientation did observe that integrated objects may actually improve VWM capacity (Olson & Jiang, 2002). Overall, these results indicate that the predication of the Fixed Slot Model that VWM operates in units of integrated objects may be weaker than previously assumed. VWM may be limited by both the number of integrated objects, as well as the feature combinations of those objects.

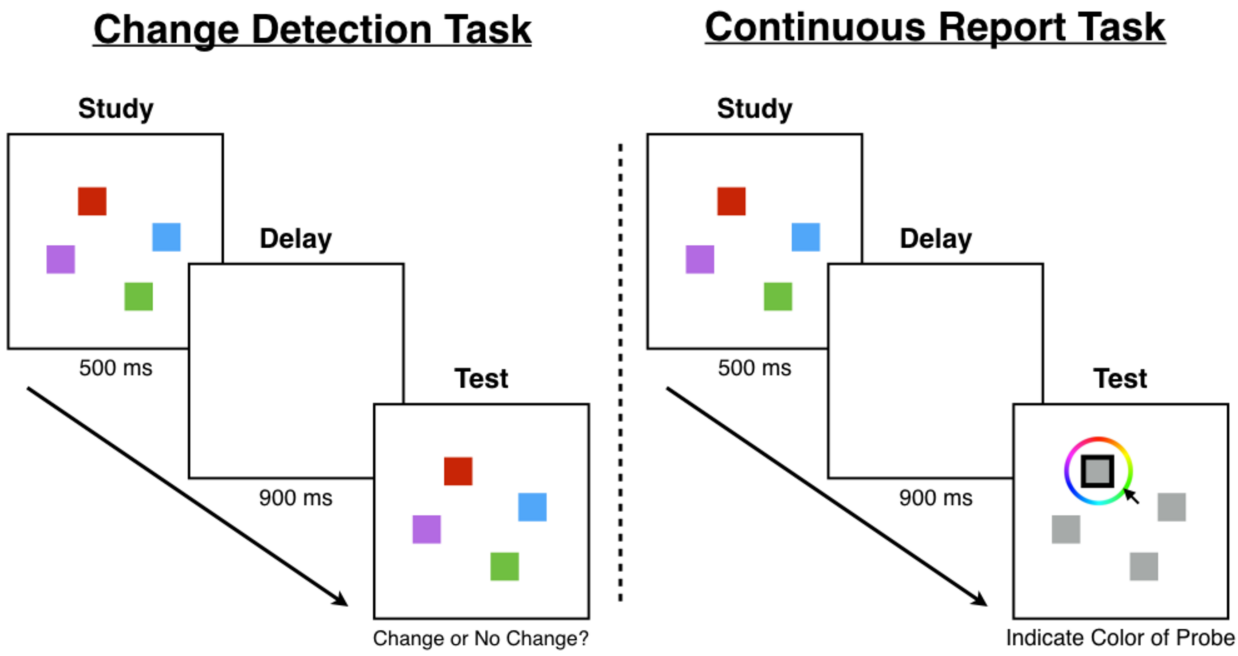

Figure 7. Illustration of typical VWM methods. In a change detection task, participants are briefly shown an initial array of items (colors, objects, etc). After a delay, they are then shown a test display. Half of the time the display is exactly the same, and whereas the other half one of the items in the display changed. Participants are told to indicate whether a change occurred. In the continuous report task, participants are briefly shown an initial array of items. After a delay, one of the previous items is probed, and participants indicate their memory for a single feature in a continuous space (i.e. report the previous color of the item using a color wheel).

Continuous Report

In order to measure fidelity of VWM, experiments have used a paradigm often referred to as continuous report. Generally, observers are presented an array of items of a single feature (i.e. color, orientation). After a delay, a single item position is probed (typically highlighted with a square), and participants indicate the specific feature of that item in a continuous space. So if the participant saw four colored squares, after a delay a single item was probed and participants indicated the color of the object using a continuous color wheel. This allows researchers to create a distribution of how precise the responses for items are (by taking the difference between responses and the true value), and see how these distributions vary as a function of different features or set sizes. For example, as set size increases a normal distribution with a larger standard deviation (SD) would indicate a less precise VWM representation of that feature. A limitation of change detection tasks is that they provide little information as to how well each individual object is remembered. However, continuous report allows for a measure that quantifies the amount of information available in memory for an item.

How Precise are VWM Representations? Zhang and Luck (2008) were interested in evaluating the role continuous resources and/or slots may play in explaining VWM capacity limitations using a continuous report method. In their experiment, participants

were presented with 1-6 color objects in an array. After a brief delay a single item was probed (using a highlighted square), and participants reported the color of the probed item using a continuous color wheel surrounding the display. In order to calculate accuracy, they took the error of each response to create a distribution around the true color value, and measured SD as a function of precision (i.e. amount of information available). They observed that as set size increased from 1-3 items, SD increased, but then remained constant as set size increased to 6. Further studies have collaborated these results (Brady et al., 2013).

Such results are not compatible with a pure resource model of VWM (which would predict SD's to increase across all set sizes), and instead suggest some variation of a "slot" model. Through a follow-up change detection experiment, Zhang and Luck (2008) proposed that a "Slots and Averaging" model (an amended version of the Fixed Slot Model) best fit the observed data in the continuous report experiment. The logic of this model is that an observer has a fixed amount of "slots" in VWM, able to store information in an all-or-nothing process of a given precision. The key difference between this model and the Fixed Slot Model is that these slots can be flexibly allocated across different items. So if an observer has a capacity of four slots and a display only contains one item, the observer can allocate all these slots to that single stimulus. By then averaging the slots together, the observer can obtain more precise estimates.

Bays, Catalao, and Husain (2009) also used continuous report to build upon this work, but while also taking into account both errors in color and location. Their experiment was similar to Zhang and Luck's (2008) – participants saw an array of 1-6 colored items, and after a brief delay a single item from the display was probed. Participants reported the color of the probed item by clicking on a color wheel. When modeling distributions of the responses to the targets, however, they also took into account possible responses to non-targets (i.e. location errors). They observed that responses to non-targets were not uniform (which would have reflected guessing), suggesting participants were making location errors during the task. When this error was taken into account (modeling distributions for both the target and non-target), they observed that precision decreased as the number of items in the display increased. This pattern of data is best explained by a common continuous resource distributed dynamically across the entire display, and not by a Slots and Averaging Model that constrains the number of items that can be represented.

1.7 Understanding Dichotomies in Memory

When looking at the memory literature what one notices very clearly is within a particular paradigm, memory system, or stimulus type we often talk about memory in dichotomies. We think of what memory is through the lens of a specific system,

method, or type of stimulus. When one zooms out, it becomes natural to ask to what extent are these dichotomies related to each other?

In fact, many of these dichotomies may be intimately related or even the same. For instance, the idea that people simply fail to represent information altogether is a component of some fixed capacity theories of VWM and some theories of recall in VLTM. In a typical change detection task, if an observer fails to detect a change both the Fixed Slot and Flexible Slot Model suggest the observer did not represent the changed item in memory. The failure at test is not due to an observer not having enough information to make the correct judgment, but rather that they had no information in order to make the correct judgment.

Along similar lines, if during a VLTM test an observer fails to identify a previously seen item, according to the DPSD model this may be the result of a recollection failure. When the observer saw the previously seen image at test, recollection failed to occur, so the observer had no information to inform their decision. Without any information related to the previously seen item, they incorrectly classified the image as "new". In both of these cases, fixed capacity theories of VWM and the DPSD model of VLTM say failures of memory are the result of the observer having no information available at test.

A natural question arises as to what extent are the Fixed/Flexible Slot Model and recollection describing the same kinds of phenomena but using distinct methods and theories? By utilizing methods across these fields, it may be possible to inform both perspectives. For instance, many VWM researchers studying the Fixed Slot Model (and in general) use some variation of a change detection task. What if methods typically adopted for the study of recollection were used in a VWM task? At test, observers could be asked to classify a test object as either Old or New, as well as make a source localization judgment. Would it be possible for an observer to incorrectly classify an old object, but make a correct source judgment? What would the implications of such results be for our understanding of VWM capacity?

There is also the idea that failures during memory tests are not the result of failing to represent information, but rather that the information available to observers to make decisions declines or becomes less precise under certain conditions. This is a notion shared by both the Continuous Resource Model of VWM and theories of familiarity in VLTM. In a typical change detection task, it may not be the case that observers fail to represent an item if they do not detect a change occurred. The Continuous Resource Model explains such performance failures as the result of a finite continuous resource that becomes more thinly distributed as the number of items in VWM increases. An observer may incorrectly report no change occurred, but they may still have some information of representation of that item. It was simply that they did

not have enough information about that item in order to make the required judgment (i.e. change or no change).

In VLTM, both the DSPD and CDP models of familiarity suggest that sometimes an observer may fail to correctly recognize a previously seen object but still have some information in memory about that object. An observer may have a familiarity signal related to a previously seen item. Familiarity operates via a signal detection based process, so even if the signal is in memory an observer may still fail to recognize a previously seen item depending on a variety of factors. The memory signal used in the comparison process could be weak, or the observer may have a particularly conservative decision criterion (avoiding all potential false alarms at the cost of recognizing some previously seen items). Regardless, this means an observer may fail to recognize an object, but still have some information about that object in memory.

Again, it becomes natural to ask to what extent are such theories describing the same kinds of phenomena? Are certain kinds of memory performance classified as "familiarity" in VLTM simply measuring memories with more thinly distributed resources? What happens in VLTM tasks as set size during encoding increases? Do observers transition from "recollection" based responses to "familiarity" responses? At high set sizes, can observers still demonstrate responses typically associated with recollection for certain items (consistent with the Fixed Slot Model of VWM), or does

performance reliably conform to familiarity-based responses (consistent with the Continuous Resource Model)? Interestingly, variable performance at high set sizes for recollection- and familiarity- based responses in VLTM would be analogous to the Flexible Slots Model, and entirely consistent with DPSD and CDP models of familiarity and recollection.

Pattern separation and pattern completion bear obvious resemblance to tolerance and discrimination. Tolerance refers to our ability to recognize something despite considerable changes in inputs across encounters. Pattern completion, which is the process by which varying inputs are recognized as belonging to the same source, seems to be one such process through which tolerance may be achieved. Discrimination refers to our ability to distinguish similar but distinct inputs from one another, and pattern separation is defined as the process of recognizing overlapping inputs as distinct. Given these definitions, it becomes abundantly clear that researchers studying VLTM in terms of pattern separation and pattern completion, and vision researchers studying object recognition in terms of tolerance and discrimination are investigating the same processes. But these two fields of research have operated largely independently of one another, devising their own methods and terminology to investigate the same phenomena. How then might methods from vision scientists studying object recognition inform long-term memory researchers studying pattern separation / completion? Do rules of perception that guide object identity also affect

long-term memory? And how might long-term memory processes affect our perceptual ability to recognize objects online?

1.8 Major Questions Remaining in the Literature

Despite generally being studied separately, VWM, VLTM, and object recognition research are clearly intimately related to one another. Indeed, many dichotomies established across different fields may be describing the same kinds of processes or behaviors. Given what these different areas have already established about the nature of memory, it becomes natural to ask what are the major questions remaining in the literature? As a result of this review, I have identified what I believe are the important unanswered questions for our understanding of memory.

What role does VWM play in supporting VLTM? Generally, VWM and VLTM have been studied separately and as a consequence, we don't have a good sense of how one system might support the other. On a basic level our current understanding is that information likely passes through VWM in order to enter VLTM. However, what if the function of one system is to support another? In particular, does VWM serve as a venue where long-term learning and integration of information takes place?

How do we characterize and directly compare VWM and VLTM? As a result of VWM and VLTM being studied independently of one another, many experiments utilize different methods that make them hard to directly compare. Do VWM and VLTM find

the same kinds of things challenging, or demonstrate similar limitations? In particular, how might these two systems handle tolerance and discrimination? There is a need to create methods, stimuli, and approaches that compare and contrast these systems, placing them on equal footing with each other. These should seek to characterize these systems beyond simply "better performance," but provide qualitative descriptions as well.

How does integration of information take place over time? Studies have tended to focus on the nature of memory following singular exposures (Shepard, 1967; Brady et al., 2008; Brady et al., 2009; Cansino et al., 2002; Konkle et al., 2010; Guerin et al., 2012; Brady et al., 2013; Kim & Yassa, 2013; Reagh & Yassa, 2014a; Cunningham, Yassa & Egeth, 2015). However, the formation of visual memories likely evolves over time and over repeated encounters with stimuli. How then do our memories acquire tolerance? And how does this tolerance change as we learn more about an object?

1.9 Outline of dissertation

The goal of this review is to give readers the necessary background into core concepts and methods of memory research that will assist with the questions addressed in this dissertation. Specifically, how does the mind acquire the representational qualities necessary to recognize past visual experience given new input? In chapter 2 I will explore how rules guiding the perception of objects in VWM (core knowledge /

spatiotemporal continuity) assist in the learning of representations in VLTM. In chapter 3 I will focus on how tolerance over the short-term is actually a feature of VWM that is used to support the construction of appropriately constrained representations in VLTM. In chapter 4 I will investigate how integration in memory takes place over time, by investigating how memory changes over multiple encounters with the same stimuli. And finally, in chapter 5 I will provide a general discussion of my results and their implications for understanding memory.

Chapter 2

Core Knowledge in VWM Supports Long-Term Learning

2.1 Synopsis

Generally, VWM and VLTM have been studied separately by distinct groups of researchers utilizing different methods and theories. As a consequence, we don't have a good sense of how one system might support the other. Our current understanding is only at the most basic level, suggesting that VWM simply acts as a passageway through which information must go through in order to enter VLTM. However, to the extent that the function of one system is to support another, this possibility has been largely ignored. VWM may not only serve as a passageway for information into VLTM, but as a venue where learning and integration of information takes place.

If integration does take place in VWM it's possibly supported by perceptual mechanisms, such as object kinematics. These are considered a component of 'Core Knowledge' and refer to the constraints from physics that can be used to determine when an object seen at one moment in time is likely to be the same individual as an object seen later (i.e. token). Object kinematics is widely known to affect the perception of objects (in infants, adults, and even non-human primates) over short periods of time.

I tested the hypothesis that human perception exploits expectations about object kinematics to limit the scope of association to inputs that are likely to have the same token as a source. In several experiments I exposed participants to images of objects, and I then tested recognition sensitivity. Using motion, I manipulated whether successive encounters with an image took place through kinematics that implied the same or a different token as the source of those encounters. Images were injected with noise at encoding (Experiment 1a) to evaluate how object kinematics might support integration, and with noise at encoding (Experiment 1b) to assess the robustness of the effect. In addition, I also controlled for potential confounding factors, altering the memory test (Experiment 2a) and evaluating a potential role of attention (Experiment 2b). I introduced variability through object orientation rather than noise (Experiment 3), and I included another manipulation of motion kinematics through smooth occlusion and disocclusion (Experiment 4). The basis of this chapter was recently accepted as a

forthcoming publication in *Journal of Experimental Psychology: General*.

2.2 Background

The problem of object recognition —whether accomplished by a person, animal, or computer system— is to recognize objects in the present on the basis of experience in the past. It is a computational problem because changes in viewpoint, orientation, and lighting conditions (among other things) can make the same object look different across encounters (Wallis & Bulthoff, 2001; Cox, Meier, Oertelt & DiCarlo, 2005; Cox & Dicarlo, 2008; DiCarlo & Cox, 2007; Rust & Stocker, 2010). In other words, an object can look very different from itself depending on how and when it is viewed, and this renders pixel-wise image comparisons an inadequate strategy for recognition (Logothetis & Sheinberg, 1996; DiCarlo, Zoccolan & Rust, 2012).

Yet humans possess relatively invariant and impressive recognition abilities, seemingly able to recognize thousands of objects (Shepard, 1967; Standing, Conezio & Haber, 1970; Biederman, 1987) following minimal exposure (Potter, 1976; Thorpe, Fize & Marlot, 1996) and despite changes to input properties. It remains unknown exactly how this is accomplished, and this remains one area in which artificial systems have yet to surpass humans (Pinto, Cox & DiCarlo, 2008; DiCarlo, Zoccolan & Rust, 2012; Andreopoulos & Tsotsos, 2013).

One known piece of the process involves a temporal association rule (Isik, Leibo

& Poggio, 2012). Over the short-term an experience that may otherwise be thought of as a single encounter with an object —perhaps lasting only a few seconds— may be better conceived as a collection of noisy encounters with varying input quality and structure. This creates an opportunity for learning. Cataloguing the ways that object-related inputs change in the short-term supplies a basis for knowing the type of variability to expect from an object in the future. Temporal association of this kind has been shown to support object recognition (Wallis & Bulthoff, 2001; Wallis, 2002; Cox et al., 2005) and to accommodate explicit computational algorithms (Isik et al., 2012).

But like many simple strategies, temporal association faces a frame problem (Dennett, 2006; Fodor, 1983). Merely correlating inputs that arrive in close succession will lead to the mixing of signals from different sources. Temporal association therefore requires a means to ensure that only signals from shared sources become associated during the process of encoding. One such means is to exploit eye movements. If an observer sees an object in her periphery and then saccades to foveate the object, she can learn by associating the two inputs. Psychophysical, neural, and computational evidence have shown that temporal association supports object memory through saccades, even demonstrating that invariant memory can be 'broken' —led astray— if an experimenter changes stimuli surreptitiously during the milliseconds between a saccade onset and termination (Cox et al., 2005; Li & DiCarlo, 2008; Isik et al., 2012; Poth & Schneider, 2016; Poth, Herwig, & Schneider, 2015). In these situations, an

observer is tricked; but in the real-world the strategy should be reliable nearly all the time. The target of a saccade and its terminus will usually be the same object.

I sought to investigate a second and unconsidered strategy for temporal association: that the visual system exploits the rules of object kinematics to limit the scope of association to inputs with the same individual object as their source. By 'the rules of object kinematics' I mean constraints from physics that can be used to determine when an object seen at one moment in time is likely to be the same individual as an object seen later. These are often referred to as the rules of spatiotemporal persistence (Scholl & Flombaum, 2010). Note that by 'the same individual' I mean 'the same exact exemplar' (or token), as opposed to 'an example of the same thing or kind of thing'.

Expectations about spatiotemporal persistence appear to support a great deal of visual and cognitive processing. They are known to constrain how children and infants evaluate and learn about the physical world (Baillargeon, Spelke & Wasserman, 1985; Spelke, Kestenbaum, Simons & Wein, 1995; Xu & Carey, 1996; Wilcox, & Baillargeon, 1998a; Wilcox, & Baillargeon, 1998b; Stahl & Feigenson, 2015). They also play a critical role in the motivation and theorizing underlying influential theories of midlevel object representation (Kahneman, Treisman & Gibbs, 1992). And they are known to influence the processing, categorization, and working memory representation of objects over short periods of time (on the order of seconds; Flombaum & Scholl,

2006; Yi et al., 2008; Flombaum, Scholl & Santos, 2009). More generally, these mechanisms are thought to be either innate or to arise early in human development, and preserved through evolutionary history for their obvious utility (Spelke & Kinzler, 2007). At least one previous study found effects of spatiotemporal understanding on the foraging behavior of a non-human species, the Rhesus monkey (Flombaum, Kundey, Santos & Scholl, 2004).

Yet expectations about spatiotemporal persistence have been scarcely considered as a mechanism of long-term learning (but see Stahl & Feigenson, 2015), particularly with respect to the challenge of object recognition. I hypothesized that the rules of spatiotemporal persistence should constrain long-term object recognition.

To investigate the hypothesis, I manipulated perceived object persistence by manipulating motion continuity. Object perception naturally involves attributions of token identity, attributions of 'which' as opposed to 'what.' As a tossed ball passes by, one perceives the ball's physical properties, and also that it is the same ball from moment to moment. This is despite the fact that the ball's projection on the retina mutates considerably over the ball's journey. Conversely, if you see two identical people walking past each other, the natural inference is that they are twins (because the same object cannot be in two places at once). This is despite the (nearly) identical physical appearances of the individuals.

To manipulate perceived token identity, I generated animations in which two objects moved in opposite directions and passed one another. The animations relied on apparent motion (Anstis, 1980; Chun & Cavanagh, 1997; Yi et al., 2008). In apparent motion, noncontiguous visual transients are perceived as instances of a single object moving continuously, provided that the transients occur close enough to one another in time and space. I will use the term 'stream' to refer to a perceived continuous motion path linking any number of transients. In my main experiment, each trial included two streams moving in opposite directions from one side of the display to another (Figure 8; dynamic demonstrations can be viewed online at jhuvisualthinkinglab.com/demos-schurgin-and-flombaum-stcontinuity). Each transient in a stream was usually the onset and then offset of a collage-like mask within a rectangular boundary (I will call these 'mask-transients'). But two transients in each display were the onset and offset of an image of a real-world object (I will call these 'image-transients').

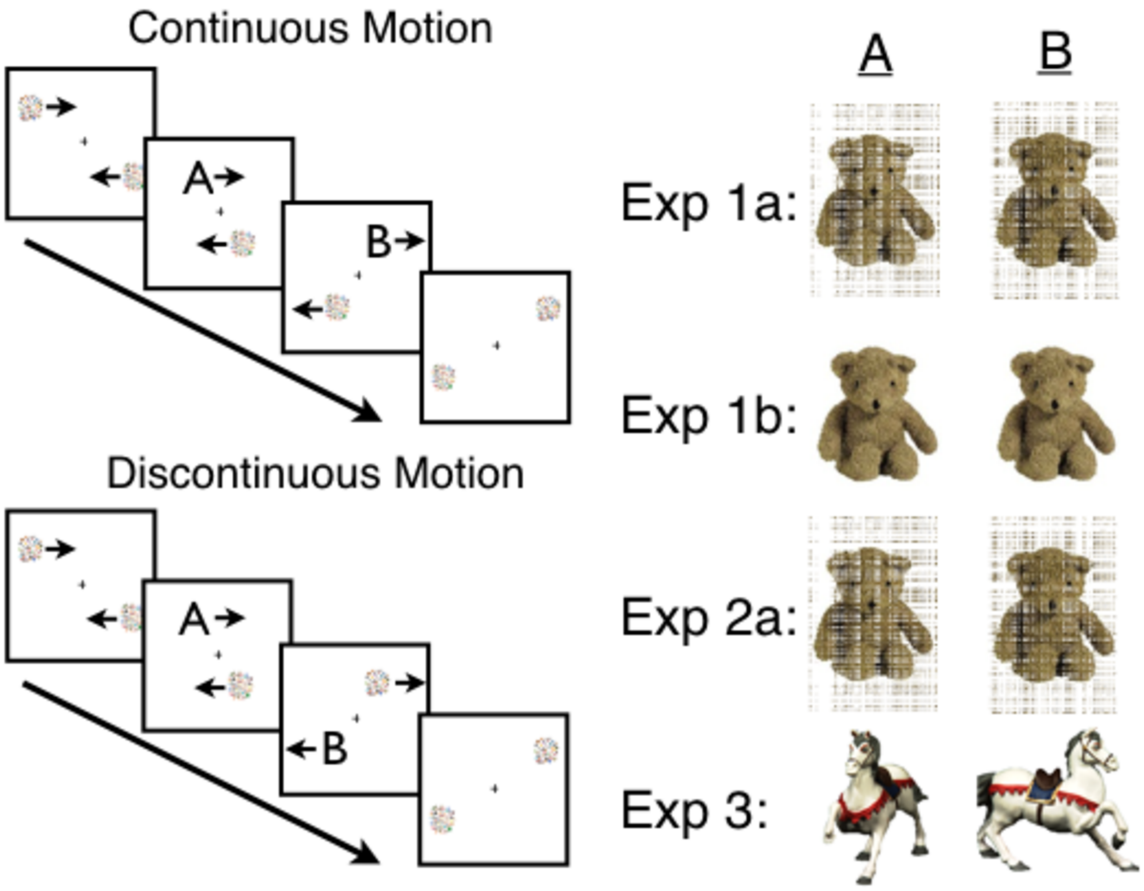

Figure 8. Procedure of the incidental encoding task. Each trial was made up of four frames, producing two apparent motion streams moving in opposite directions. Each stream comprised random noise masks (mask-transients) at their most eccentric initial and final positions during the first and last frames of a trial. [See Chun & Cavanagh, 1997, upon which these methods were closely modeled; see also Yi et al., 2008]. In the positions closer to fixation, two mask-transients were replaced by two appearances of an image (image-transients) in succession (i.e. in the the second and the third frame of a trial). The critical manipulation was whether the successive image appearances supplanted mask-transients in a single stream (continuous motion) or in different streams (discontinuous motion). In the figure, A and B designate the positions in which image-transients supplanted mask-transients for each of the two motion conditions. In Experiment 1a and Experiment 2a the image-transients were photos of an object embedded in independent noise during each of its appearances. In Experiment 1b the photos were noiseless (during encoding). In Experiment 3 the photos were of an object at one orientation and then of the same object at a different orientation. In this experiment the recognition test always showed a third, previously not shown orientation.

Within a trial, the two image-transients were always images of the same object. In half of the trials, the image-transients appeared in positions within the same stream (supplanting the mask-transients that would have appeared in those two positions otherwise). Thus, the images would be perceived as instances of the same token in two different places at different moments. I call this 'continuous motion'. In 'discontinuous motion' trials, the image-transients appeared sequentially in different streams, so that despite their similar (physical) appearances, they would be perceived as instances of two different tokens.

Exposure to images took place during an incidental encoding paradigm, typical of research on long-term memory (Brewer, Zhao, Desmond, Glover & Gabrieli; 1998; Kirchhoff, Wagner, Maril & Stern, 2000; Wittmann et al., 2005; Blumenfeld & Ranganath, 2006; Kim & Yassa, 2013). Each trial included two motion streams, as just described. The repeated image was unique in each trial. I subsequently probed memory for the presented images in a surprise test. My main prediction was that memory for images seen under continuous motion would be better than for images seen discontinuously, owing to temporal association mechanisms that are constrained by token assignment based on motion kinematics.

2.3 Experiment 1

Real-world experience with objects is often noisy and impoverished. To create such conditions in the laboratory, in Experiment 1a I randomly assorted 50% of the pixels in the images shown during encoding. Crucially, this image noise was inserted on both appearances of a given image in a trial, but assorted independently each time. This allowed me to investigate how observers integrate information from noisy experiences to build robust memories. In Experiment 1b, I replicated effects with noise injected at test, but not during encoding. The critical manipulation, as already described, was whether the images were presented continuously or discontinuously. I expected better recognition memory for images seen in continuous motion, that is, in a single apparent motion stream.

2.3.1 Experiment 1a: Integrating Noisy Inputs on the Basis of Object Kinematics

Methods

Participants. For each of the experiments reported my goal was to test and analyze results from approximately 20 participants. I sought 20 because a previously published exploratory study confirmed reliable effects and an effect size of $\eta_p^2 = 0.18$, using a sample size close to 20 (Schurgin, Reagh, Yassa & Flombaum, 2013). That study reported an experiment identical to the current Experiment 1, except that it included only noiseless presentations and testing of images. A power analysis on the results of

that experiment demonstrated that 18 participants would be sufficient, with a power of 0.95 and a 0.05 significance level.

I expected that engagement with the task would vary by subject and over the course of the several semesters during which experiments would be run. I therefore adopted a uniform exclusion criterion. Data from an individual subject was excluded from analysis if old-image classification accuracy (i.e. number correctly identified as old divided by number of old items tested) was more than two standard deviations below the group mean (with the subject included); or if more than 50% of responses were of a single response category (equal numbers of old, similar, and new images were presented during test).

A group of 20 Johns Hopkins University undergraduates participated in Experiments 1a. The results from four participants were excluded. All participants reported normal or corrected-to-normal visual acuity. Participation was voluntary, and in exchange for extra credit in related courses. The experimental protocol was approved by the Johns Hopkins University IRB.

Apparatus. Experiments took place in a dimly lit sound-attenuated room. Stimuli were presented on a Macintosh iMac computer with a refresh rate of 60 Hz. The viewing distance was 60 cm so that the display subtended 39.43° × 24.76° of visual angle.

Stimuli and Procedure. Stimuli were generated using MATLAB and the Psychophysics toolbox (Brainard, 1997; Pelli, 1997). All stimuli were presented within the full display frame of 39.43° × 24.76°. Participants completed a visual object recognition memory task that included two stages. During the first phase (incidental encoding), participants were shown 384 color images of real-world objects on a computer screen with the cover task of indicating whether an item onscreen was an "indoor" or "outdoor" item. They indicated their responses using the computer keyboard.

Items in the task were shown using a common apparent motion paradigm, wherein an item could be seen as part of a single apparent motion stream or as parts of two different apparent motion streams (see Chun & Cavanagh 1997 and Yi et al., 2008, upon which these methods were closely modeled). Each trial included four 200 ms frames with two images in each frame (Figure 8). In the first frame, the two images consisted of scrambled object parts, and the images were positioned in the periphery to bias the perceived motion directions in subsequent frames. In the second frame, the images approached fixation, and one of them turned into a real-world object. In the third frame, the images passed the central region, and one of them turned into the same real-world object that appeared in the previous frame. Finally, in the fourth frame, the images moved to peripheral positions. Thus, in all trials, participants saw a single real-world object repeated twice, but either along the same stream (continuous

motion) or between streams (discontinuous motion). Participants were told that they would see two images of the same object during each trial. I also instructed participants to do their best to maintain fixation in the center of the display, a strategy which I suggested would maximize performance. But I did not enforce fixation.

Experiment 1a manipulated the presentation of the real-world object in the second and third frame so that 50% of the pixels were randomly scrambled.

To enhance the perception of two distinct apparent motion streams, 20% of the trials contained only mask-transients in in all four frames. Participants were instructed to press the space bar if a trial contained no image of an object. Trails were self-paced and began when the participant pressed space to continue.

During each trial of the second phase (surprise retrieval), participants viewed a single image presented on the screen. The images were those presented in the previous task (old images), objects similar but not identical to ones in the previous task (similar images), and completely new objects that were not in the previous task (new images). For each image, participants were instructed to indicate whether it was "old," "similar" or "new." Participants were instructed to classify an image as old only if it was identical to the image they saw previously. They were also told that similar images would be categorically similar to an image they saw previously, but something may have changed. During the instructions for the surprise test, participants were shown examples (not from the exposure set) of old, similar, and new images to make these

instructions clear. Images were shown for 3000 ms each, with a 500 ms inter-stimulus interval. Participants could only make a response when an image was onscreen, and only the first response made was recorded. Overall, participants did not respond on 3.6% of trials in Experiment 1a and 4.7% in Experiment 1b.

Data Analysis. I did not include responses that were faster than 200 ms, which accounted for a total of 0.66% of trials in Experiment 1a.

Results

Cover Task Performance During Encoding. During incidental encoding participants were asked to classify the image that appeared in each trial as either 'Indoors' or 'Outdoors.' To determine whether motion kinematics affected responses on this cover task, I computed the probability of a given classification as a function of motion type. There was no significant difference for continuous (indoor classifications, M = 61.6%) compared to discontinuous motion (indoor classifications, M = 60.4%), $t(383) = 0.25$, $p = 0.8$.

Recognition from Memory During Test. To de-confound discrimination ability from potential response biases, I analyzed recognition performance using a signal detection measure d_a (Green & Swets, 1966; Loiotile & Courtney, 2015). Each trial of the surprise test included either one of the objects shown during encoding, a novel object that was not shown previously, or a similar object: an object in the same

category as and bearing some resemblance to one of the objects actually shown during encoding. Participants reported whether they thought an object was 'old,' 'new,' or 'similar,' i.e. previously seen, unseen, or resembling something seen, respectively. With d_a I could compute both a Novel-Old discrimination index and a Similar-Old discrimination index.

I observed significantly better Novel-Old discrimination for objects encountered along spatiotemporally continuous (d_a = 1.25) compared to discontinuous (d_a = 0.98) paths, $t(15) = 3.00$, $p < 0.01$, Cohen's D = 0.51 (Figure 9). A similar trend was observed for Similar-Old discriminations, although this effect did not reach significance. (Continuous d_a = 0.27; discontinuous d_a = 0.17, $t(15) = 1.65$, $p = 0.12$, Cohen's D = 0.37).

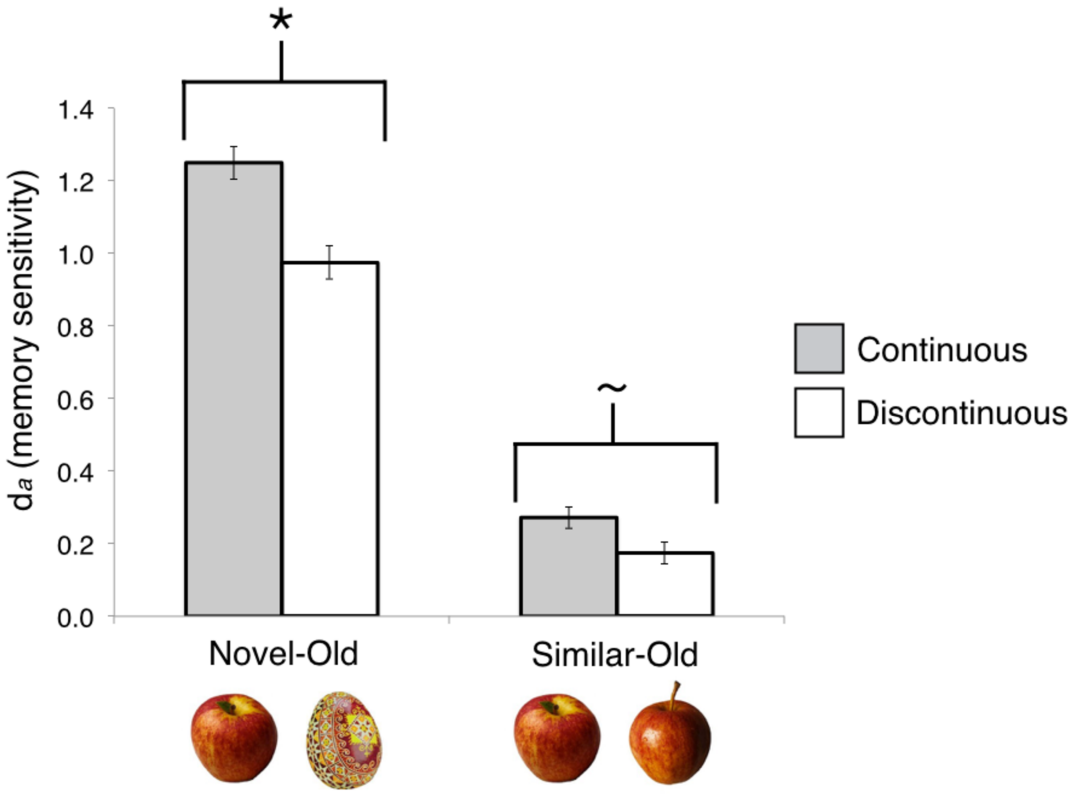

Figure 9. Results of Experiment 1a (n = 16). For both Novel-Old and Similar-Old comparisons, participants' discrimination performance was better for objects encountered along continuous as opposed to discontinuous paths. * designates $p < 0.01$, ~ designates $p = 0.12$. Error bars represent within-subject error (to remove between-subject variability, see Cousineau, 2005).

Memory strength analysis. To further investigate the extent to which exposure through spatiotemporal continuity supports the acquisition of more robust memory, I performed a memory strength analysis of the kind often used in research on long-term memory (Wickens, 2002, Ch. 3; Wixted, 2007). The analysis assumes that recognition decisions are based on the strength of a memory signal in relation to a decision criterion (Egan, 1958; Ratcliff, Sheu, & Gronlund, 1992). I conceived of the analysis as follows: each time an object image was encountered during encoding it added

strength to a memory trace for that object. Was greater strength added when the second encounter followed the first under continuous motion (implying a single object token)? I could answer the question by modeling the strength of memory traces fit to classifications made during the surprise test.

The model assumes that in a continuous space of memory strength, test stimuli would elicit normal distributions centered around different means (μ) with unequal variance (Wixted, 2007). The distribution of responses elicited by New images was specified to be centered at 0 (i.e. on average, failing to elicit a memory signal since those images had not been seen previously) with a standard deviation (SD) of 1. For Old images, I created two distributions coded by motion continuity (continuous vs. discontinuous). These distributions had their means as free parameters, and a shared SD, also a free parameter. The distribution of memory strengths elicited by similar images also included a mean and SD as free parameters. To classify responses in the model, I specified free parameters for two ordered decision criteria (λ). Within a given distribution any sample above λ_1 would be classified as "Old," samples between λ_1 and λ_2 would be classified as "Similar," and samples less than λ_2 would be classified as "New". In total, this gave the model seven parameters to fit the data.

For simplicity of reporting the present model used averaged data across all participants as its input, although I also employed a hierarchical version to account for

individual data. (Results were virtually identical with those of the aggregate model). The model simulated the averaged data for 10,000 iterations across two chains.

The results are shown in Table 1. Figure 10 plots the results graphically. The key effect is that the memory traces evoked by continuously presented items were stronger than were the memory traces evoked by discontinuously presented items. This difference was significant ($p<.01$); over the 10,000 simulations, the 99% confidence interval for the difference between μ (Continuous) and μ (Discontinuous) was 0.44 to 0.12 (thus larger than 0). This analysis adds to reported sensitivity comparisons by showing how differences in memory strength between continuously and discontinuously shown memoranda could have produced the particular patterns of response observed.

Parameter outputs for model of Experiment 1a.

Parameter	Mean	Standard Deviation	97.5% CI
μ (Continuous)	1.50	0.06	1.61-1.38
μ (Discontinuous)	1.21	0.05	1.32-1.11
SD (Cont. & Disc.)	1.50	0.07	1.36-1.64
μ (Similar)	1.10	0.04	1.17-1.03
SD (Similar)	1.13	0.05	1.04-1.23
λ_1	1.61	0.04	1.63-1.54
λ_2	0.55	0.02	0.60-0.51

Table 1. Outputs of parameter values for the memory strength model of Experiment 1a. Old items were shown either continuously or discontinuously, and were thus fit by distributions with different means. The main result is that the fitted μ for continuously shown items were larger than for discontinuously shown ones, implying stronger memory traces on average.

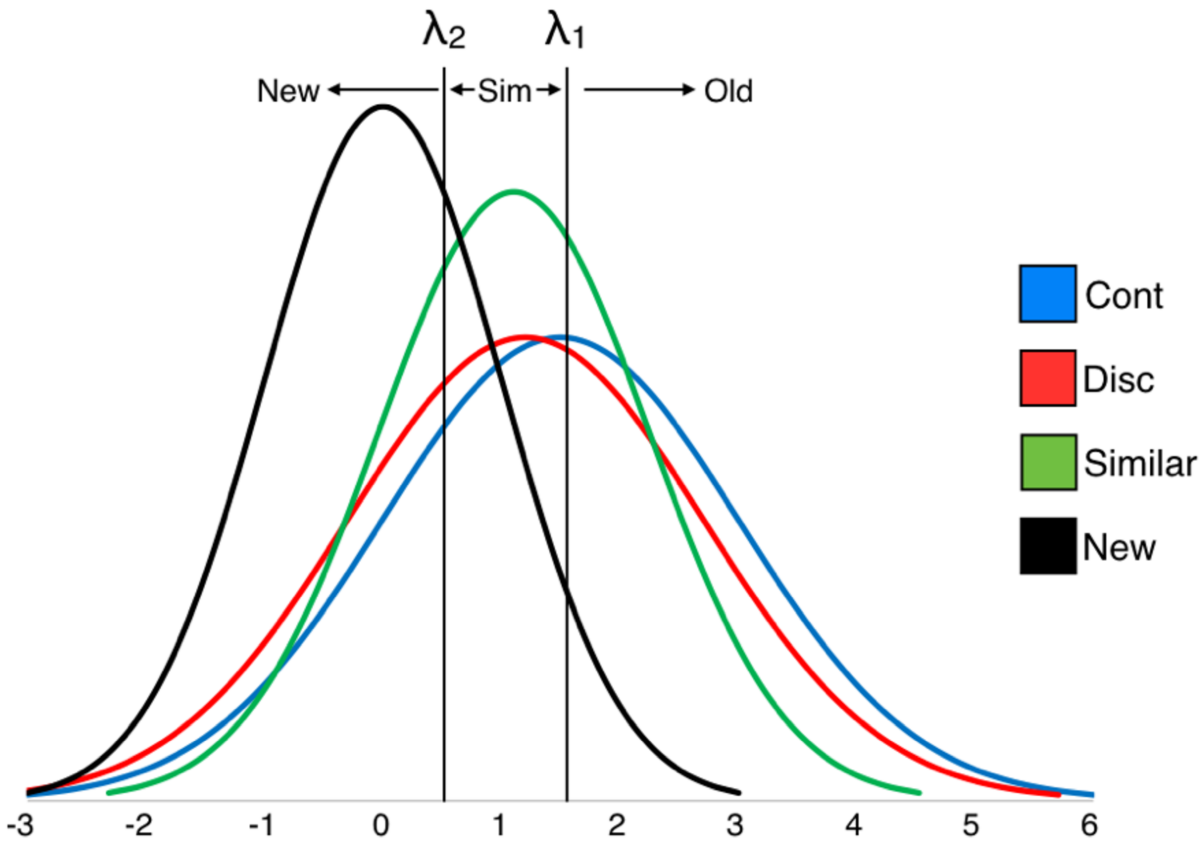

Figure 10. Fitted model distributions for Experiment 1a. Under the assumption that new items evoke virtually no memory trace, those (shown in black) were assigned a μ of 0 and a SD of 1. Distributions for continuously presented items (blue) were fit, along with distributions for discontinuous items (red) and similar items (green). Continuous items elicited stronger memory traces, on average.

2.3.2 Experiment 1b: Continuity Breeds Robust Tolerance to Noise at Test

Experiment 1b was designed to replicate the main results of Experiment 1a, with a small methodological modification. In this case, images were presented without noise during encoding. But at test, images were embedded in either 25%, 50%, or 75% noise. In Experiment 1a, therefore, exposure to images was impoverished and the opportunity to recognize those images took place under relatively better conditions. In

Experiment 1b, this was reversed: exposure conditions were better than test conditions, which varied from slightly noisy to very noisy.

Methods

Participants. A separate group of 20 Johns Hopkins University undergraduates participated in Experiment 1b. All participants reported normal or corrected-to-normal visual acuity. Participation was voluntary, and in exchange for extra credit in related courses. The experimental protocol was approved by the Johns Hopkins University IRB.

Stimuli, Apparatus and Procedure. All methods were identical to those in Experiment 1a, with the following exceptions: At encoding, participants viewed a total of 360 color images of objects without noise. At retrieval, participants classified images of objects, the pixels of which were randomly scrambled by 25%, 50%, or 75%. The three levels of noise were equally distributed across old, similar, and new images.

Data Analysis. I did not include responses that were faster than 200 ms, which accounted for a total of 1.58% of trials in Experiment 1b.

Results

Recognition from Memory During Test. Experiment 1b replicated the effects observed in 1a. A between-subjects ANOVA revealed a main effect of motion continuity on Novel-Old discrimination, $F(1, 19) = 19.12$, $p < 0.01$, $\eta_p^2 = 0.50$, and also a

main effect of noise level, $F(2, 38) = 26.23$, $p < 0.01$, $\eta_p^2 = 0.58$, with no interaction between the two $F(2, 38) = 0.19$, $p = 0.83$, $\eta_p^2 = 0.01$. As noise level increased, performance decreased. Across all three noise levels there was a benefit for motion continuity (Figure 11). Planned comparisons contrasting continuous and discontinuous exposure were significant at all three noise levels (all $p < .05$). As in Experiment 1a, effects on Similar-Old discriminations were in the same direction, but failed to reach significance.

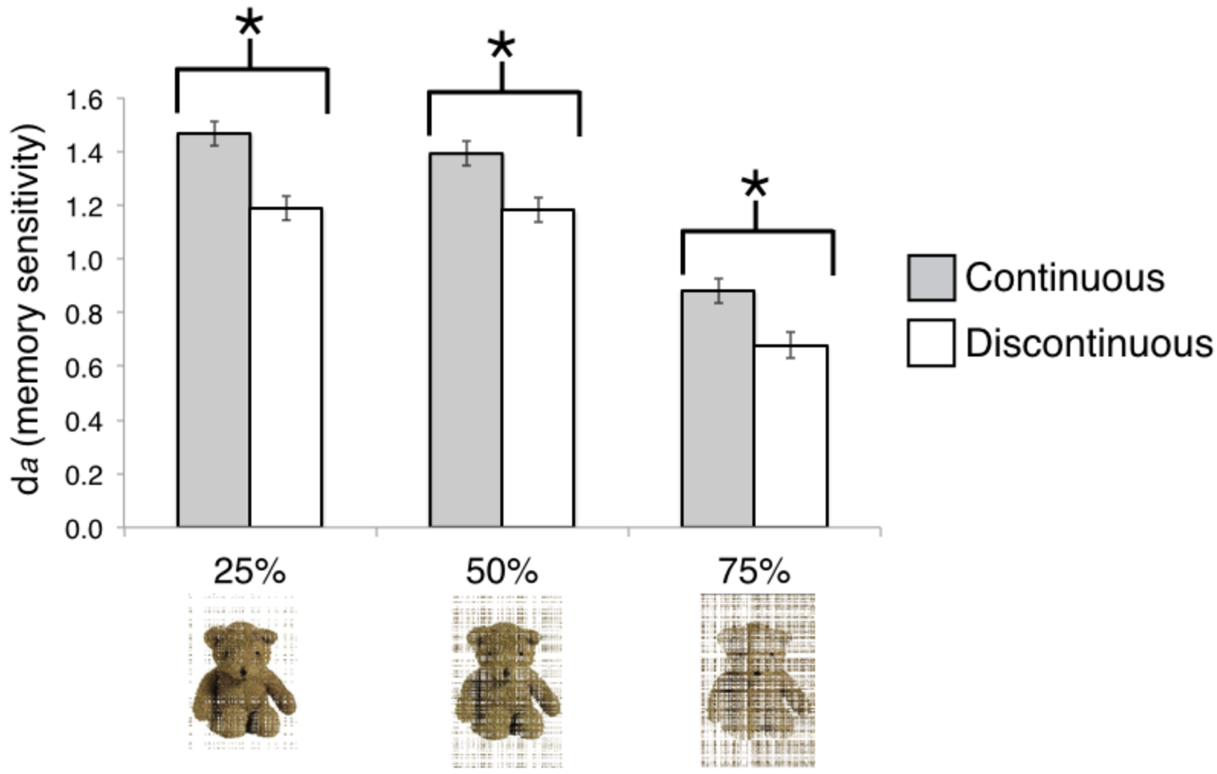

Figure 11. Results of Experiment 1b (n = 20). At test, participants judged whether an image was old, similar, or new relative to objects encountered during encoding. Each image was embedded in either 25, 50, or 75% noise. Across all three noise levels discrimination of Old objects was better when the object was encountered along a continuous (compared to a discontinuous) path. * designates $p < 0.05$. Error bars represent within-subject error.

Memory Strength Analysis. I again applied a memory strength model to analyze the results of Experiment 1b. The same model was applied here as in Experiment 1a, but it was applied three times, once to the data from each noise condition at test. Table 2 shows μ parameter values for continuous and discontinuous images as a function of noise level, along with 95% confidence intervals for their differences. Again, continuous motion elicited stronger memory traces than discontinuous (all $p<.05$).

Experiment 1b also provided an interesting test of concept for the memory strength model. Additional noise at test should cause an image to evoke a weaker memory trace (when it was seen previously). Indeed, mean μ values as a function of noise level declined, particularly for items with 75% noise injected compared to 50%. This effect is less a prediction of the model, and more a useful demonstration that the model generally captures effects that should impact the strength of an evoked memory signal.

Together, the novel effects reported in Experiment 1 support the hypothesis that token identity in the short-term is exploited to support long-term object memory. These experiments do not identify the exact mechanisms by which token identity is tracked and then imposes its constraints. I discuss potential mechanisms, including the role of attention, in the General Discussion under the heading, 'Potential Mechanisms

and the Role of Attention.' Experiment 2b will additionally address the potential role of attention.

Subset of outputs for model of Experiment 1b.

Parameter	Noise Level 25%	Noise Level 50%	Noise Level 75%
μ (Continuous)	1.57	1.55	1.15
μ (Discontinuous)	1.25	1.33	.93
95% confidence μ (Cont-Disc)	.51 to .13	.41 to .028	.45 to .012

Table 2. μ values as a function of noise level for continuously and discontinuously presented images, along with 95% confidence intervals for their differences. Continuously presented items elicited stronger memory signals when represented at test across all noise conditions. [Full parameter outputs for this and all reported experiments can be obtained at http://www.jhuvisualthinkinglab.com/demos-schurgin-and-flombaum-stcontinuity].

2.4 Experiment 2

Experiments 1a and b investigated an impact of motion kinematics during encoding, and for this reason, I utilized an old/similar/new procedure that has been used previously in studies of long-term memory investigating encoding effects (Bakker

et al., 2012; Kim & Yassa, 2013; Rentz et al., 2013; Stark, Yassa, Lacy & Stark, 2013). Because the effects reported are novel, and involve a kind of manipulation not typical in research on long-term memory (i.e. perceived motion), I was motivated to replicate the effects with a forced choice measure of long-term memory. According to some theories, old/similar/new judgments with a single test image index different processes than forced choices between two or more images. Thus, Experiment 2a was designed to replicate the effects in Experiment 1a while using a forced choice procedure to demonstrate the reproducibility and pervasiveness of these effects. Experiment 2b used the methods of 2a to serve as a control investigating an alternative attention-driven account of the results.

2.4.1 Experiment 2a: Replication with a Forced Choice Test

Methods

Participants. A new group of 21 Johns Hopkins University undergraduates participated in Experiment 2a. The results from two participants were excluded from Experiment 2a following the exclusion criteria from Experiment 1. All participants reported normal or corrected-to-normal visual acuity. Participation was voluntary, and in exchange for extra credit in related courses. The experimental protocol was approved by the Johns Hopkins University IRB.

Stimuli, Apparatus, and Procedure. All methods were identical to those in Experiment 1a, with the following exceptions: At incidental encoding, participants were shown 368 images of real-world objects. At retrieval, participants performed a two alternative forced choice (2AFC) task. In the center of the screen were two objects, only one previously shown during incidental encoding. Participants had to indicate which of the two they had encountered during encoding. Half of these trials were Novel-Old comparisons, in which a previously shown object was paired with a new, categorically distinct object. The other half of these trials were Similar-Old comparisons, in which a previously shown object was paired with a similar-looking object from the same category. Each pair of images was present on the screen until the participant made a response.

Data Analysis. I did not include responses that were faster than 200 ms, a total of 3.6% of trials.

Results

Cover Task Performance During Encoding. As in Experiment 1, I found no significant difference in the proportion of outdoor/indoor image judgments during encoding as a function of motion continuity, $t(367) = 1.38$, $p = 0.17$.

Recognition from Memory During Test. I observed a main effect of motion continuity for Novel-Old comparisons, $t(18) = 2.24$, $p = 0.038$, Cohen's $D = 0.40$.

Performance was better when objects were encountered along a spatiotemporally continuous (M = 76.7%) versus discontinuous path (M = 72.9%).

For Similar-Old comparisons, I observed a similar trend, $t(18) = 1.9$, $p = 0.07$. Performance was better when objects were encountered along a spatiotemporally continuous (M = 59.2%) versus discontinuous path (M = 55.5%), Cohen's D = 0.57.

Figure 12 plots the results graphically.

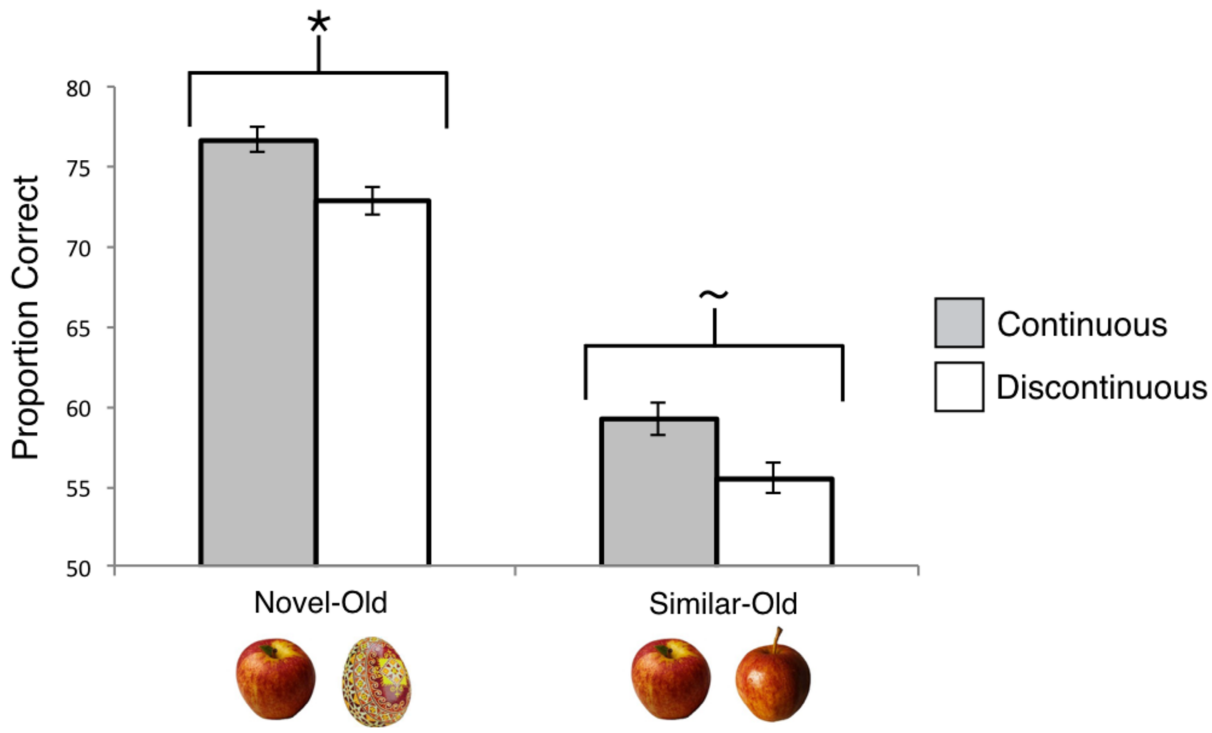

Figure 12. Results of Experiment 2a (n = 19). For both Novel-Old and Similar-Old comparisons, participants' discrimination performance was better for objects encountered along continuous as opposed to discontinuous paths. * designates $p < 0.01$, ~ designates a trending difference ($p = 0.07$). Error bars represent within-subject error.

2.4.2 Experiment 2b: Controlling for the role of attention

A critical feature of my theorizing around the three experiments presented so far is that continuous motion —and consequent token perception— enhance memory performance by supporting the integration of independent encounters with the images shown. A salient set of questions about the mechanisms underlying such a process concerns the role of attention: might the mechanism of integration be an attentional bias induced by the initial appearance of an image in one of the streams? I will discuss this possibility in the General Discussion, because it seems likely and it is fundamentally consistent with my view that Core Knowledge interacts with psychological mechanisms broadly, including attention. Via Experiment 2b, however, I sought to control for an alternative account by which attention could conceivably act to produce the effects without also supporting integration over time.

Specifically, I was concerned that continuous motion merely biases spatial attention, causing the second image to be seen —and thus better remembered— without promoting integration of the encounters. Suppose that the first encounter with an image is often barely perceived or processed, because its location is unpredictable and its presence fleeting. Then that encounter may fail, on its own, to produce any substantive memory trace. But if attention is attracted by the first object, it could track the stream in which that object appears, which means it could more effectively capture the second appearance within continuous streams, while missing (or processing less so)

any second appearance that falls in the discontinuous stream. In other words, perhaps the benefit I observed is simply one of perceiving and processing a single image during its second appearance, not one of integrating two encounters, an effect having nothing to do with integration over token identity.

To control for this possibility, in Experiment 2b I placed two distinct objects in each trial. I manipulated motion continuity as I had previously, and then in the second phase of the experiment I tested memory for the second images shown during encoding (using the 2AFC methods of Experiment 2a). The alternative attention account just outlined predicts that memory for the second images within continuous streams would be better than for the second images within discontinuous streams. But in this case the effect would not be attributable to integration across encounters, since the relevant encounters always involved distinct images. This would be a problem since the modulation in memory performance would have nothing to do with integration of object knowledge from multiple observations. In contrast, a null result would suggest that continuity does not merely promote memory for images that follow other images in a continuous stream, and therefore, that the effects already observed arise in virtue of integration per se.

Because my point of view predicts a null result for this experiment, while the alternative predicts a significant effect, I used methods here that I hoped would produce good performance overall, maximizing the possibility of observing an effect

and reducing the possibility of observing floor performance. Specifically, I used the 2AFC test of Experiment 2a, a test that generally provides for better performance than old/similar/new classification. For the same reason, I included only noiseless images at both encoding and test. I also restricted testing to old vs. new images because similar discriminations in the previous three experiments only produced trending effects and generally low performance. Finally, the first image in each trial of incidental encoding was always the same picture of an apple. This was done to reduce the chances that the image would be integrated with the second image in each trial (which would produce poor memory representations), and to make it highly predictive with respect to the appearance of the second image. To the extent that the first image is a cue towards tracking continuity, I sought to make it as reliable and recognizable as possible. Figure 13 schematizes the methods of this experiment.

Methods

Participants. A new group of 21 Johns Hopkins University undergraduates participated in Experiment 2b. All participants reported normal or corrected-to-normal visual acuity. Participation was voluntary, and in exchange for extra credit in related courses. The experimental protocol was approved by the Johns Hopkins University IRB.

Stimuli, Apparatus, and Procedure. The experiment was identical to 2a, with the following exceptions. For both continuous and discontinuous trials, the first image of a

real-world object was always the same: an apple. Participants were instructed to judge whether the object that followed the apple was an "indoor" or "outdoor" object. No noise was present in the images during encoding. At test, the 2AFC task exclusively included Novel-Old comparisons.

This experiment was preregistered with aspredicted.org, available permanently at https://AsPredicted.org/85qv5.pdf. This experiment was conducted following a round of reviews, and thus after all of the other reported experiments. My lab recently adopted preregistration as standard practice, though this was after the other experiments had already been completed.

Data analysis. I did not include responses that were faster than 200 ms, a total of 0.06% of trials.

Results

Cover Task Performance During Encoding. As in Experiments 1 and 2a, I found no significant difference in the proportion of outdoor/indoor image judgments during encoding as a function of motion continuity, $t(367) = 0.94$, $p = 0.35$.

Recognition from Memory During Test. I failed to observe a main effect of motion continuity for Novel-Old comparisons, $t(20) = 0.82$, $p = 0.92$, Cohen's $D = 0.03$. Performance was indistinguishable statistically regardless of whether an image appeared in the second position during encoding along a spatiotemporally continuous

(M = 86.1%) versus discontinuous path (M = 85.9%). Figure 6b shows these results graphically.

Learning to ignore? What if in the control experiment, the repeated and recognizable apple caused participants to learn to ignore motion continuity, and this is why performance in the two conditions became similar (in contrast with Experiment 2a)? To rule out the possibility I analyzed performance over the course of the experiment in four blocks. Learning to ignore would predict a continuity effect early on in the experiment, and no effect only later. I found overall that performance slightly decreased as the experiment progressed: 87.63% in block 1, 87.50% in block 2, 85.38% in block 3, and 83.34% in block 4. The difference between block 1 and 4 was significant, $t(20) = 2.50$, $p = 0.02$. But I did not find a continuity effect in any individual block (all p's > 0.64).

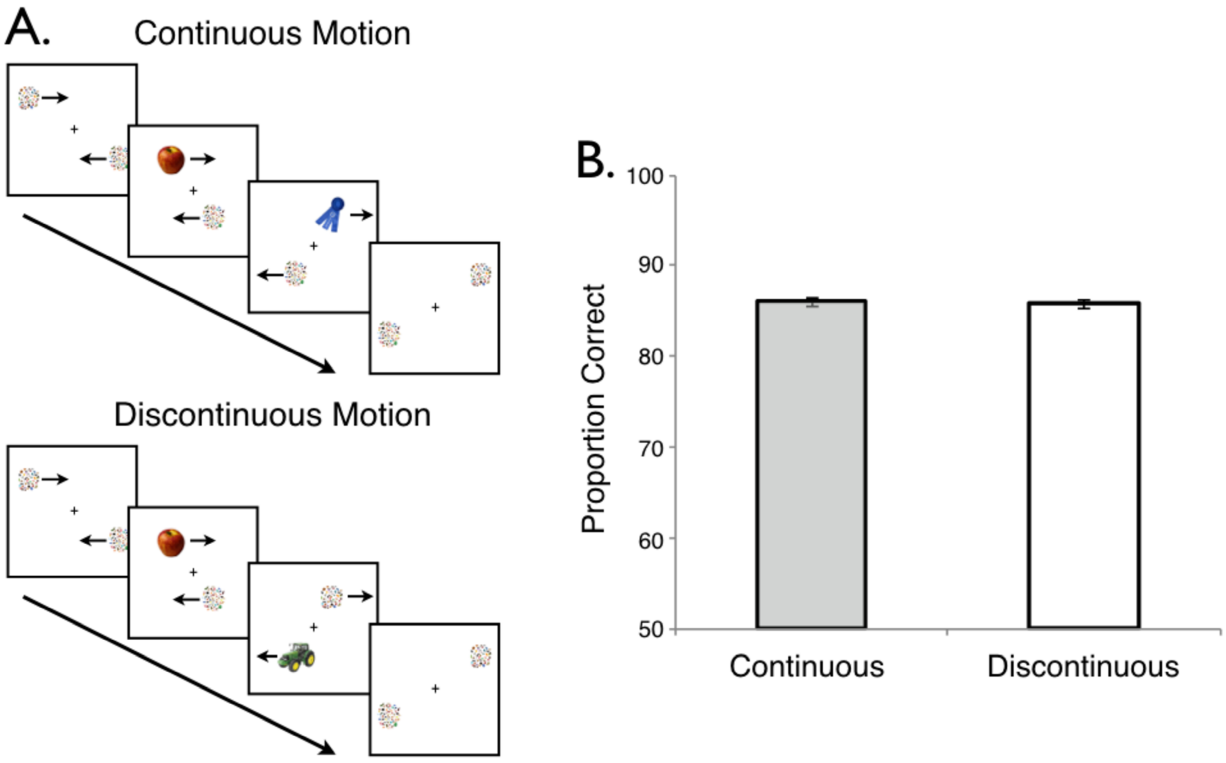

Figure 13. Methods and Results of Experiment 2b (n = 21). For Novel-Old comparisons, participants' discrimination performance was indistinguishable for images following an apple along continuous and discontinuous paths. Error bars represent within-subject error. These results are inconsistent with an account in which continuity merely promotes memory for images in the second position of a continuous path.

2.5 Experiment 3: Integrating Inputs with Different Orientations

Changes to an observer's viewpoint —which result in changes to an object's orientation relative to the observer— exert radical changes on the inputs to the visual system from a given object. Consider the two images of a horse shown in Figure 1: the two images are extremely different in the particular stimulations that they produce on the retina. Yet we easily recognize the horse in the two images as the same —an ability often called viewpoint invariance, or tolerance. Experiment 3 was designed to directly

test whether viewpoint invariance is supported by spatiotemporal continuity. Objects moving through space will naturally appear in changing orientations relative to an observer. Knowing that an object is nonetheless the same token can support integration in the service of invariance.

Experiment 3 was identical to Experiment 1, with the following exceptions. Each presentation of an object during encoding comprised two images of the objects at two different orientations (without image noise injected). At test, 'Old' objects were shown at a third, different orientation. To simplify the instructions to participants, this experiment included only 'New' foils, thus requiring that participants judge each image shown as either 'Old' or 'New.'

Methods

Participants. A new group of 30 Johns Hopkins University undergraduates participated in Experiment 3. I initially recruited 20 participants, but found that my exclusion criteria demanded the removal of six. The effects with fourteen were significant, but I decided to test an additional 10 participants to verify the findings considering so many exclusions. This resulted in one more participant meeting my exclusion criteria, for a total of seven participants removed and 30 tested total. Below I report results for the 23 included participants as well as the full set of 30, in the interest of transparency. All participants reported normal or corrected-to-normal visual acuity.

Participation was voluntary, and in exchange for extra credit in related courses. The experimental protocol was approved by the Johns Hopkins University IRB.

Stimuli, Apparatus, and Procedure. In Experiment 3, the stimuli and procedure were the same as in Experiment 1 with the following exceptions: At encoding, participants judged whether the object was more "square" or "round." This change was made because the stimulus set used included an overwhelming number of objects that would have been rated as "indoor" objects. Participants judged a total of 316 color images of objects (taken from Geusebroek, Burghouts & Smeulders, 2005). At retrieval, participants saw images that were either a new object or an old object from the incidental encoding phase but shown at a completely new orientation. They were asked to classify the images at test as "Old" or "New".

Data Analysis. At retrieval, responses were only recorded if the image was still present on the screen. Overall, participants did not respond on 1.5% of trials. Additionally, I did not include responses that were faster than 200 ms, which resulted in an additional 0.3% of trials being excluded from the reported analyses.

Results

Cover Task Performance During Encoding. As in Experiments 1-2, I found no significant difference in the proportion of square/round image judgments during encoding as a function of motion continuity, $t(157) = 1.02$, $p = 0.31$.

Recognition from Memory During Test. Requiring only an Old or New judgment meant that I could compute d' as a measure of sensitivity. As in Experiment 1, performance was better when images appeared along continuous as opposed to discontinuous paths (d' = 1.35 vs. d' = 1.22, t(22) = 2.67, p = 0.01, Cohen's D = 0.23). When including all 30 participants, results were still significant (d' = 1.24 vs. d' = 1.14, t(29) = 2.34, p = 0.026, Cohen's D = 0.19). Thus, even when recognizing an object at a third, never-before-seen viewpoint, spatiotemporally continuous motion supported better recognition performance (Figure 14).

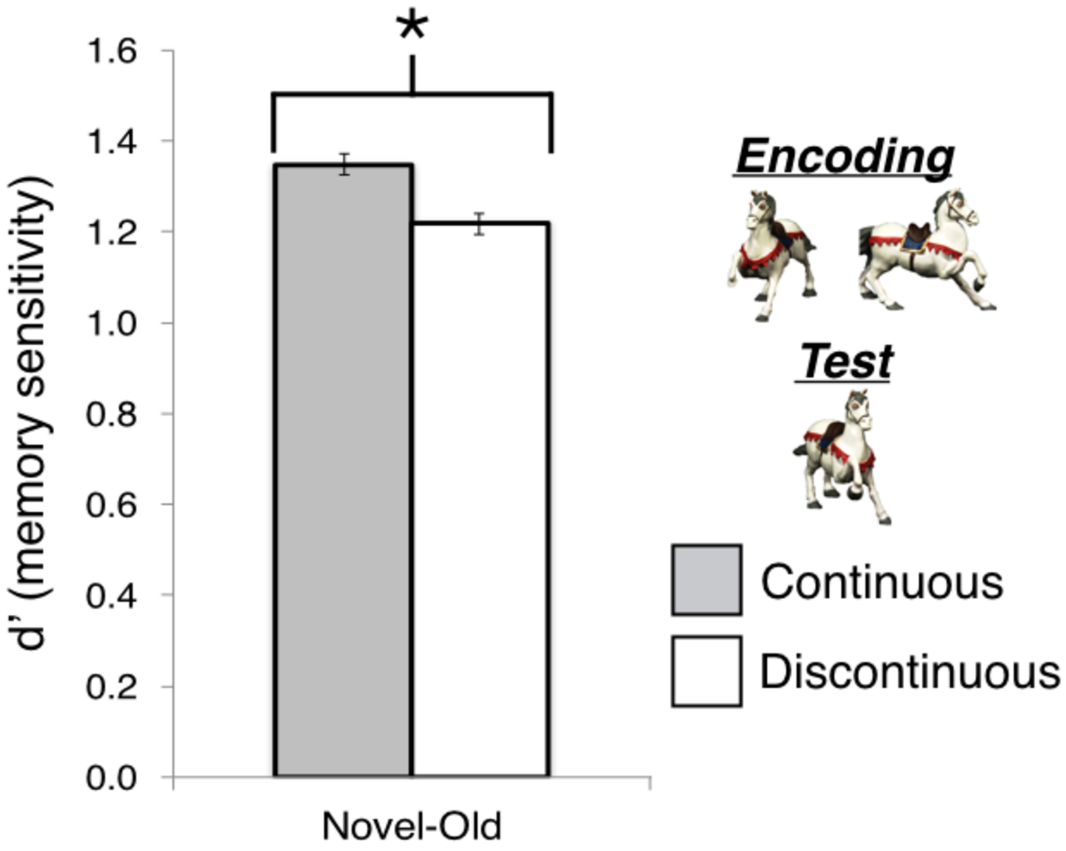

Figure 14. Results of Experiment 3 (n = 23). Participants' memory sensitivity was better for objects encountered along continuous as opposed to discontinuous paths. * designates $p < 0.05$. Error bars represent within-subject error.

Memory Strength Analysis. For this experiment, I again performed a model-based memory strength analysis. Recall that during presentation each exposure to an object was at a different orientation. The theory in this case is therefore that continuity (compared to discontinuity) would breed integration across these orientation differences, thus producing stronger memory signals when the objects were seen during test at a third orientation. This experiment did not include similar foils or similar judgments. Thus, it could be modelled with fewer free parameters, specifically the μ and SD for continuous and discontinuous old object distributions, and a single λ criterion. Average μ estimates for continuously perceived objects (1.84) were higher than for discontinuously perceived objects (1.57). The 95% confidence interval for the differences between the condition estimates was .66 to .01.

2.6 Experiment 4: Integrating Noisy Inputs through Occlusion and Disocclusion

Research has shown that infants and human adults track object persistence via multiple cues, and that they track object persistence despite occlusion, the complete disappearance of an object caused by another object in-between it and the viewer (Burke, 1952; Spelke et al., 1995; Xu & Carey, 1996; Yi et al., 2008). In Experiment 4 I

sought to conceptually replicate the results of Experiment 1 while employing a different manipulation of motion continuity.

Again, the experiment began with an incidental encoding phase, and again each trial included a single image that appeared twice. Here, however, the appearances of the images followed two episodes of complete occlusion. In the continuous motion trials, the images appeared from behind the same occluder (a pillar drawn into the image). While in the discontinuous motion trials the images appeared from behind two different occluders, each presented on opposite sides of the display. And crucially, the image never fully traversed the space between the occluders. In both conditions, motion was fast in the periphery and slower at fixation so that the majority of the presentation of each image was directly near fixation regardless of the motion continuity condition. Figure 15 schematically illustrates these presentations.

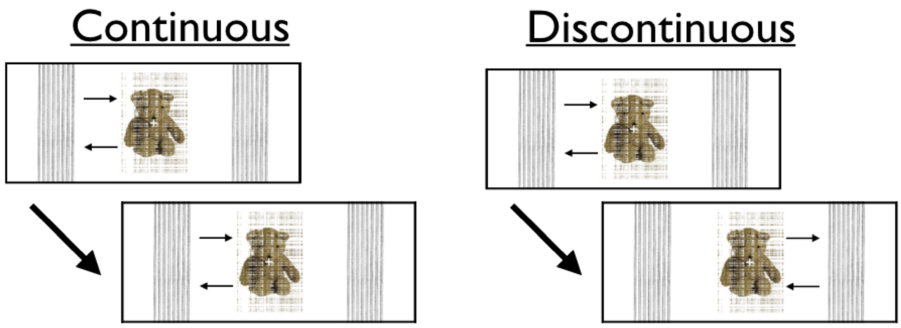

Figure 15. Procedure of the incidental encoding task used in Experiment 4. In each trial, an image of an object (embedded in noise) appeared from behind an occluder, moved towards the center of the display and then retreated to its origin, where it became occluded. This full trajectory lasted for one second, and it was repeated twice in a trial. The manipulation was whether the second appearance was from behind the same occluder as the first (continuous motion) or from the occluder on the opposite side of the display (discontinuous motion).

Methods

Participants. A new group of 22 Johns Hopkins University undergraduates participated in Experiment 4. The results from two participants were excluded. All participants reported normal or corrected-to-normal visual acuity. Participation was voluntary, and in exchange for extra credit in related courses. The experimental protocol was approved by the Johns Hopkins University IRB.

Stimuli, Apparatus, and Procedure. In Experiment 4, the procedure was the same as in Experiment 1, except as follows. During each trial of incidental encoding an image appeared twice under dynamic occlusion. Participants judged the object as either 'indoors' or 'outdoors' (as in Experiment 1). The encoding display contained a vertical column on each side of fixation (each occupying a space of 6.4° and starting 10.6° from fixation on either side). An image embedded in 50% noise appeared from behind one of the two columns, moved to fixation, reversed motion direction to return to its origin column, and it then disappeared by naturalistic occlusion behind that same column. Critically, the same image appeared a second time embedded again in (independent) 50% noise. But the second appearance could be either from behind the same (continuous) column as the original appearance, or the one on the opposite side of the display (discontinuous).

Subjects were instructed to fixate the central cross throughout the experiment, although eye movements were not monitored. The images always exited an occluder at a speed of 38.72°/s. During each approach to fixation, the speed (S_i) in each frame of subsequent motion (i) was equal to:

$$S_i = S_{i-1} - \left(0.64 - \left(\frac{0.02i}{4}\right)^2\right)$$

This produced a speed at fixation of 3.78°/s (30 frames of motion). Returning to its origin the object moved through the same trajectory in reverse. Thus, movement was very fast in the periphery and slower at fixation so that the majority of the presentation of each image was directly near fixation.

Throughout the incidental encoding task participants viewed a total of 320 unique color images (i.e. there were 320 trials of incidental encoding). During the retrieval test, participants were asked to classify the objects as "Old" or "New". It was explained that they should only classify old images as "Old," and that new or similar looking objects should be classified as "New." The proportion of trials that contained old, similar and new images were exactly the same (160 images per image type). Compared to Experiments 1-3, I changed the response type (removing the similar response) because performance with that response type tended to be poor and to make the instructions more straightforward for participants. Each image was present on the screen until the participant made a response.

Data Analysis. I excluded from analysis responses that were faster than 200 ms, a total of 1.3% of all trials.

Results

Cover Task Performance During Encoding. As in Experiments 1-3, I found no significant difference in the proportion of outdoor/indoor image judgments during encoding as a function of motion continuity, $t(319) = 0.80$, $p = 0.43$.

Recognition from Memory During Test. Participants demonstrated extremely high performance for Old images overall, $d' = 2.44$. This advantage compared to the previous experiments was likely caused by the greatly increased encoding time in Experiment 4, with each image remaining present for a total of 2000 ms (compared to 400 ms in Experiments 1-3). As in the previous three experiments, I found a significant performance benefit for Old images that had been encoded under continuous motion ($d' = 2.48$) compared to objects encountered under discontinuous motion ($d' = 2.41$), $t(19) = 2.20$, $p = 0.04$ Cohen's $D = 0.13$.

Memory Strength Analysis. I also applied a memory strength model to this experiment. estimated μ values were much higher than before —again, probably because of the greatly increased exposure time and concomitant performance. But μ estimates for continuously presented images were on average larger than for discontinuously presented ones (mean of the estimates 2.37 v 2.30, respectively). And

the 95% confidence interval for the mean estimate of the parameter differences slightly fell below 0, having a range between .22 and -.031. The smaller effects in this experiment make sense given the increased exposure times and higher performance, and they are also consistent with previous work by Yi and colleagues (2008) on the effects of motion continuity, wherein the same apparent motion manipulation used in the previous experiments had larger and more reliable effects than the same occlusion manipulation used in this experiment.

2.7 Discussion

The problem of object recognition is often characterized as one of invariance, storing memories that can be used to recognize an object despite drastic changes at the level of basic inputs. Any given object can induce a wide array of retinal stimulations because of factors as basic as lighting conditions, viewer motion, and rotation. To meet this challenge one strategy involves temporal association — integrating noisy experiences during one encounter to build representations that tolerate variability at later encounters. My results demonstrate that temporal association is a component of human object memory, but critically, one that is supported and constrained by core principles of spatiotemporal object persistence. Tracking an object at the level of a 'token' allows an observer to leverage changes in appearance over the short-term to produce appropriate discrimination in the long-term.

Two important features of my experiments are the use of both low-level image noise (Experiments 1 and 3) and an orientation manipulation (Experiment 2), as well as the use of two rather different manipulations of object continuity (occlusion and apparent motion). These different manipulations produced results that varied in magnitude, but importantly, the similarity in the kinds of effects obtained make many lower-level factors unlikely as explanations. Specifically, I obtained effects with very short exposures (400 ms) and with relatively longer exposures (2 s); I obtained the effects when the relevant stimuli could be construed as competing with other stimuli in the display for encoding (Experiments 1-3), and when they appeared alone in the display (Experiment 4); I obtained effects for stimuli that could be easily fixated, regardless of continuity condition (Experiment 4), as well as stimuli that might have been viewed often outside of the fovea (Experiments 1-3). This convergence suggest that the effects are not exclusive to taxing encoding settings or the opposite, to more comfortable ones. Nor are they exclusive to instances in which eye movements cannot be made quickly enough to directly pursuit a moving object —in Experiments 1-3, each image appeared for only 200 ms, followed immediately by the next image while a saccade takes at least 200 ms— nor to settings in which eye movements could easily adjust following an incorrectly anticipated image appearance —In Experiment 4, each image remained present and smoothly moving for one second, and it was always the only image in the display. I also obtained the same effects using a variety of testing

procedures and analyses, including alternative forced choice, old/similar/new judgments, signal detection measures, and memory strength modeling. Effects of spatiotemporal continuity on object encoding and subsequent memory appear, therefore, to reveal a general mechanism for associating inputs based on tracked token identity.

Potential Mechanisms and the Role of Attention

Future research will be required to characterize the exact nature of the mechanisms involved in the process of using token identity to constrain object memory. It seems to me relatively uncontroversial to believe that the relevant mechanisms operate without intention to encode and remember nor even the intention to track, in any explicit sense.

One promising candidate for a supporting mechanism is visuospatial attention. Attention is known to constrain learning and memory in a variety of specific cases (Chun & Turk-Browne, 2007, Kawachi & Gyoba, 2006; Scholl & Pylyshyn, 1999). But there are several different ways that attention could play a role in the experiments, each with different implications.

In the context of Experiment 2b, a control, I discussed one alternative account of my results that would appeal to attention. This involved an attentional bias towards continuous motion, but one that does not promote integration across encounters. Note

that even that account has built into it a reliance on spatiotemporal continuity. And because Experiment 4 (dynamic occlusion) was so different in kinematics and structure from the other experiments, a consistent alternative account would require several ad-hoc adjustments to explain all my results. (For example, Experiment 4 involved 2000 ms exposure durations compared to only 400 ms exposures in the other experiments, making it unlikely that a second object appearance could be completely missed, even given an attentional bias). More importantly, though, Experiment 2b dispatched this alternative account empirically, and so I turn now to another attentional account that seems both more plausible and includes a role for integration through motion.

An image appearing in one place could conceivably produce an expectation of an image in a second, trajectory-extrapolated place, directing attention automatically, or at least without volition necessary, to that second location. On this view, spatiotemporal expectation would be conceived as built into the operation of attention, producing a sort of filter by which association is naturally biased towards signals that are likely to have arrived from the same source. Crucially, the view is one that involves association, not merely more exposure. Viewing the role of attention in this way might suggest that token identity is not tracked explicitly or intelligently, but more implicitly and ballistically. However, the argument that the relevant expectations are hardwired, as opposed to representation-dependent, becomes less appealing considering the abundance of evidence that attention has been shown to have the

right expectations about token identify in very different contexts (see the discussion called 'Core Knowledge Supports Cognition', below). I am comfortable with either interpretation, to be sure, noting that the important contribution at this moment is that token identity, in practice, affects object memory.

An alternative to an attentional account would be that even with diffuse or unfocused attention, integration of signals is constrained after entering memory encoding. For example, whether a second image is related to a first via pattern completion or pattern separation in the medial temporal lobe (MTL) could be a 'switch' that token identity acts upon (Yassa & Stark, 2011). Indeed, these two computations in the MTL seem to bear an uncanny resemblance to the opposing challenges of tolerance and invariance often discussed in the context object recognition. Thus, it is surprising that long-term visual memory and object recognition are often treated as separate topics.

At present my results are insufficient to adjudicate between a more memory- or attention-based account of the influence of token identity. In lieu, I therefore emphasize that the results are not mutually exclusive, and that there are extant reasons to anticipate that both hold some truth. Importantly, both hypotheses suggest an interaction between the underlying mechanisms of perception and memory, topics which are often investigated in isolation.

Implications for "What" and "Where" in Object Recognition

Similarly, my results suggest that human object recognition depends on the coordination of 'what' and 'where' representations of objects. These representations are often ascribed to ventral and dorsal processing streams, respectively, with research on object recognition generally focused on the former. In my experiments, continuous motion supported the assimilation of information about object appearances, suggesting that investigating coordinated activity in these systems could lead to advances. Research on long-term visual memory rarely involves manipulations designed to engage token processing mechanisms. Most visual long-term memory studies display only static stimuli in fixed positions on a screen (Brady, Konkle & Alvarez, 2011; Guerin et al., 2012; Kim & Yassa, 2013; Stark et al., 2013). If memory mechanisms are optimized for, expect, and/or depend upon at least some object motion and related stimulus variability then static objects may fail to fully engage typical encoding mechanisms. Integrating this knowledge into subsequent research could lead to advances in our understanding of long-term memory.

The Importance of Token Identity Beyond Perception

My results also suggest that clues to object recognition may generally be found in research more typically framed in terms of attention and perception. For example, in the domain of object file research (using a paradigm known as object reviewing) some

studies have investigated a small set of appearance transformations that occasionally preclude token ascription. It has also been found that under certain circumstances similarities in physical appearance engender token ascriptions, despite a lack of any obvious cues to spatiotemporal continuity (Mitroff & Alvarez, 2007; Richard, Luck & Hollingworth, 2008; Hollingworth & Franconeri, 2009; Moore, Stephens & Hein, 2010). Presumably any expectations about the physical appearances of objects that are strong enough to guide token perception will play an important role in scaffolding long-term recognition and categorization.

Implications for Machine Learning

These results suggest a remedy to the problem of unsupervised learning in object recognition. In machine learning, object recognition is often trained with supervision —an oracle explicitly conveys which exemplars in a training set belong to the same categories or are instances of the same object (Andreopoulos & Tsotsos, 2013). Thus, machine learning algorithms demand prior knowledge (and often, a great deal of training). What could play the role of supervision in human object recognition, especially in very young children? Tracking of token identity can be thought of as supplying a sequential set of test stimuli with feedback to a learning program. If one knows that an object at time point B is the same token as the one at time point A, then one can ask whether the image observed at B is consistent with the representation

observed at A. If it is, then it should reinforce the representational structure acquired at A. If it is not, it should lead to an adjustment of the representation and a new prediction about what should be encountered at time point C.

This is largely the point that has been made in research on how eye movements can support learning with a temporal association rule: it is a good bet that the intended target of a saccade and its final terminus are instances of the same token (Cox et al., 2005; Li & DiCarlo, 2008; Isik et al., 2012). Perhaps learning through saccades is a particular case of a more general strategy of learning about appearance by tracking token identity. I am not aware of any research in young children or infants that has investigated temporal association through saccades in the interest of object memory. But a great deal of research shows that children and infants track token identity over time, largely by relying on spatiotemporal constraints (Baillargeon, Spelke & Wasserman, 1985; Spelke, Kestenbaum, Simons & Wein, 1995; Xu & Carey, 1996; Stahl & Feigenson, 2015). Thus, learning about object appearance through token assignment is a viable mechanism for early learning in the service of object recognition.

Core Knowledge Supports Cognition

Finally, my results demonstrate that mechanisms underlying 'Core Knowledge' may be a critical component of object recognition and long-term memory. This is consistent with the broader literature, which has long established that Core Knowledge

constrains memory, learning and perception across the lifespan (e.g. Van Marle & Scholl, 2003; Cheries et al., 2008; Stahl & Feigenson, 2015). Here I have focused on knowledge of spatiotemporal persistence (at times, called 'continuity') in the context of object motion and occlusion. Core Knowledge also includes an understanding of object cohesion (Xu & Carey, 1996; Noles, Scholl, & Mitroff 2005; Cheries et al., 2008), the specific dynamics of occlusion (Scholl & Pylyshyn, 1999) expectations about balance (Baillargeon & Hanko-Summers, 1990), and possibly even expectations about entropy (Newman, Keil, Kuhlmeier & Wynn, 2010), among other things. In many of these instances infants and adults not only possess the relevant knowledge, but the corresponding expectations also appear to govern reasoning and learning. For example, there is evidence that infants reason about the number of objects in a scene based on spatiotemporal expectations (Xu & Carey, 1996; Wilcox, & Baillargeon, 1998a; Wilcox, & Baillargeon, 1998b). Violations of object cohesion have been shown to corrupt infants' expectations about more versus less (Cheries et al., 2008), while the same kind of cohesion violations have been shown to disrupt object file priming (Noles, Scholl, & Mitroff, 2005). Violations of naturalistic occlusion impair multiple object tracking in human adults (Scholl & Pylyshyn 1999). And in adults, the tendency to classify otherwise similar physical events as involving occlusion or containment influences which features are more easily remembered in a working memory paradigm

(Strickland & Scholl, 2015). Thus, Core Knowledge appears to be a set of constrains on cognition with a domain general reach.

Surprisingly though, long-term memory and object recognition have not been considered previously as domains influenced by Core Knowledge. A recent study by Stahl and Feigenson (2015) is perhaps the exception, where violations of expectations in infants were shown to influence associations between objects and features. Given those results and the ones presented here, the natural hypothesis that follows is that Core Knowledge in general plays an important role in long-term visual memory. Indeed, the motivation for my experiments was that expectations about persistence create an opportunity for appropriately constrained memory (i.e. memory that balances tolerance and discrimination); if one knows that an object is still the same, one can learn just how different that object can look form itself.

Similar opportunities may arise by applying expectations about other aspects of Core Knowledge. For example, one challenge in object recognition is to distinguish between objects that are distinct, but touching or otherwise contiguous in a two-dimensional image plane. Research on balance and support suggests that even young infants have expectations about when one object could conceivably support a contiguous one, showing surprise when an object does not topple over under certain relationships (Baillargeon & Hanko-Summers, 1990). Knowing something about when contiguous parts of an image are distinct objects should affect long-term

representations of those objects. Broadly then, I hope that Core Knowledge is integrated into research on how visual long-term memories are acquired and encoded because that knowledge may supply solutions to many of the computational challenges facing object recognition.

Conclusion

Altogether, perhaps the main conclusion to draw is that tolerance in long-term object recognition arises from a kind of over-tolerance in shorter-term object recognition. Confidence about token identity based on kinematics lends flexibility to perceptual and working memory mechanisms so that changing inputs can become amalgamated.

Chapter 3

Short-Term Tolerance Supports Long-Term Recognition

3.1 Synopsis

This chapter is motivated to address two related questions that arise from my review. The first question is one of characterizing VWM and VLTM. This is in part a descriptive challenge, as these systems have been studied using different methods and tasks making them hard to directly compare. At the most basic level, we can ask which system has better performance? This isn't a very interesting question, as we might guess that utilizing the same methods VWM would demonstrate better performance. However, there's another question that is much more interesting: do VWM and VLTM find the same kinds of things challenging? In qualitative ways, do they face the same kinds of limitations?

The second motivation is perhaps a more particular way of phrasing the questions posed above. In particular, I want to understand how these two systems handle tolerance and discrimination. By tolerance, I am referring to the challenge of recognizing previous input despite change across encounters, and by discrimination I am referring to the ability to distinguish similar but distinct input. This is in many ways a question that one arrives at through thinking about both systems in the context of object recognition. In object recognition, a major unresolved question is where does tolerance come from? How are we able to grow tolerant memories? However, if you're only focused on one memory system (VWM or VLTM), you're not looking at the growth process. Tolerance needs to be investigated across the continuum of memory, in both the short- and the long-term.

I hypothesized that features of visual memory over the short-term support the construction of appropriately constrained and diagnostic long-term memories. Specifically, VWM should be more tolerant of subsequent variability at test than VLTM. To investigate this issue, I developed a novel paradigm equating encoding and test across different memory types. Participants were given a single encounter with two objects, one tested immediately in VWM and another to be tested later in VLTM. At test, VWM performance was robustly tolerant to variability at test (whether through injecting noise into images or by altering their orientation). In contrast, VLTM performance suffered linearly as more variability was introduced into test stimuli.

Additionally, I replicated these general differences in a single trial adaptation of my experiment, demonstrating poorer performance in VLTM is not the result of interference from remembering hundreds of test items. Further experiments controlled for a variety of high- and low-level alternative explanations. Altogether, my results demonstrate that VWM exhibits greater tolerance in order to integrate information into VLTM. Thus, part of the function of VWM is to support future object recognition ability.

3.2 Background

Visual object recognition is an exceptionally challenging computational problem whenever pixel-level image analysis is an inadequate solution (Logothetis & Sheinberg, 1996; DiCarlo, Zoccolan & Rust, 2012). For humans, many organisms, and for general-purpose artificial systems (AI), variability at the pixel-image level arises inescapably because of changes in viewpoint, orientation, and lighting (Wallis & Bulthoff, 2001; Cox et al., 2005; Cox & Dicarlo, 2008; DiCarlo & Cox, 2007; Rust & Stocker, 2010). Yet we humans are remarkably good at recognizing objects. We can even recognize objects when they are partially occluded, when they are simplified as line drawings, and when we have only encountered other exemplars from a particular category, not the specific object to be recognized (Biederman, 1987; Rust & Stocker, 2010). Object recognition is a rare case in which humans are still competitive with (and in many cases surpass) AI. How we acquire our ability remains a mystery.

I sought to investigate object recognition in human observers by contrasting abilities over the short-term and over the long-term. In this way, I could precisely characterize how humans are able to *acquire* object representations tolerant to the kinds of variability described above. Surprisingly, the majority of object recognition research has focused on defining the potential content of our representations (Biederman, 1987; Biederman & Gerhardstein, 1993; Scholl, 2001; Feldman, 2003) as opposed to how we acquire them. As a result, there remains almost no research describing how humans learn to recognize objects outside of developmental research investigating infant and toddler abilities to perceive and identify objects (Spelke, 1990; Spelke, Kestenbaum, Simons & Wein, 1995; Xu & Carey, 1996; Xu, Carey & Quint, 2004; Cheries, Wynn & Scholl, 2006; Stahl & Feigenson, 2015).

In what follows I provide one of the first descriptions detailing the learning process behind human object recognition. Specifically, I investigated the novel hypothesis that memory for objects over the short-term should be highly tolerant – accepting large amounts of variability – in order to act as a venue for integration. This is because over the short-term the visual system can rely on the expectation that objects in a scene will remain the same and can therefore discount potential differences in appearance to learn more about what it saw. In turn, this learning process can be used to construct appropriately constrained representations to be

utilized in the long-term that are both tolerant to variability but discriminating (i.e. able to reject objects that look the same but are not).

I tested this prediction by investigating the tolerance of visual working memory (VWM) and long-term memory (VLTM) in several experiments. In the experiments participants were initially exposed to two (or more) real-world objects in a trial. After a brief delay, one of the objects was paired with a new object and the task was to indicate which was in the recently seen set. After completing the visual working memory task, a long-term memory test probed the previously untested objects against foils using the same testing procedure. Thus, both objects were encoded in exactly the same way, and were tested in exactly the same way – the only difference was whether an object representation was probed in either VWM or VLTM. By utilizing this procedure and manipulating the variability of stimuli, I examined the information available to these systems and their tolerance to changes across different viewing conditions. I find that VWM demonstrates greater tolerance than VLTM – it more readily recognizes objects as the same despite inputs that differ considerably in appearance – consistent with my proposed learning process for object recognition.

3.3 Experiment 5: VWM Robustly Tolerant to Noise

The primary goal of Experiment 5 was to directly evaluate my hypothesis that VWM should be more tolerant to variability than VLTM. To accomplish this, I created a novel paradigm and ran Experiment 5a to establish baseline performance – participants saw two objects in a VWM task, and their memories of those object were tested either in immediately (VWM) or later in the long-term (VLTM; see Figure 1). Then, in Experiment 5b, I injected noise into test stimuli during the test encounters by randomly scrambling a percentage of pixels (25-75%) in the images. I expected that less scrambled images would be recognized more easily, allowing me to assess the relative tolerance of VWM and VLTM.

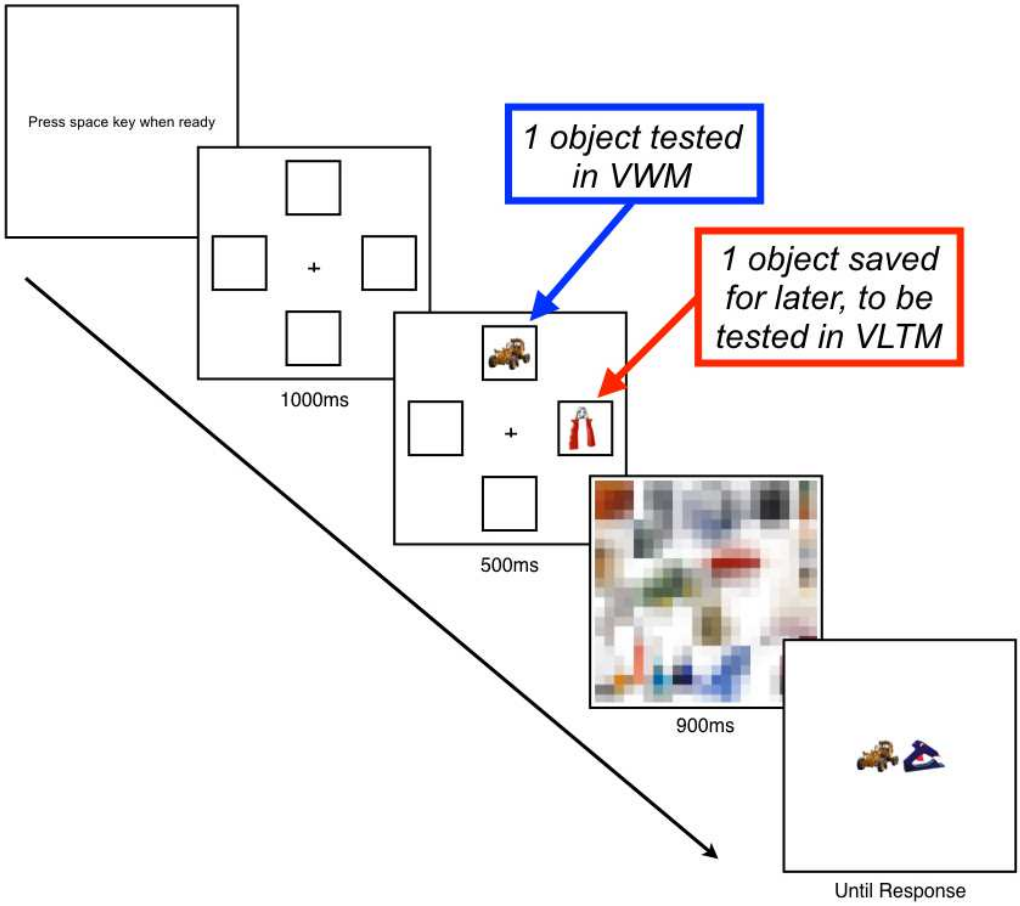

Figure 16. Illustration of general experimental procedure for Experiments 5-7. Two objects appeared briefly in one of four locations. One object was tested immediately in VWM using a 2AFC procedure. The other object in the display was saved to be tested later in VLTM in exactly the same way.

Methods

Participants. A group of 18 Johns Hopkins University undergraduates participated in Experiment 5a and a separate group of 20 in Experiment 5b. In both experiments the results from one participant was excluded due to noncompliance with the instructions. All participants reported normal or corrected-to-normal visual acuity. Participation was voluntary, and in exchange for extra credit in related courses. The experimental protocol was approved by the Johns Hopkins University IRB.

Apparatus. Experiment 5 took place in a dimly lit sound-attenuated room. Stimuli were presented on a Macintosh iMac computer with a refresh rate of 60 Hz. The viewing distance was 60 cm so that the display subtended 39.43° × 24.76° of visual angle.

Stimuli and Procedure. Stimuli were generated using MATLAB and the Psychophysics toolbox (Brainard, 1997; Pelli, 1997). All stimuli were presented within the full display frame of 39.43° × 24.76°. In Experiment 5a, participants first completed a VWM test. In each trial of the experiment, participants briefly saw two real-world objects in one of four possible locations (there were also set size one trials, to provide a baseline performance for VWM). The display was masked, and the task was to maintain the two pictures in VWM. At test, participants faced an alternative forced (2AFC) judgment involving a randomly selected object from the encoding display and a new object, with the task of identifying the old object (Figure 16). After 180 trials of this task participants faced a surprise VLTM test. On each trial, the previously untested object from each of the encoding displays was paired with a new object, and participants reported the one that was 'old,' that is, the one that had appeared at some point in the encoding phase.

Experiment 5b was identical to Experiment 5a, with the following exceptions: during the VWM task, objects were embedded in 75% noise. In the VLTM task, objects were embedded in either 25%, 50%, or 75% noise.

Results

In Experiment 5a, I observed that performance in VWM was quite high, with an average of 97.5% (SD = 1.7%) correct at set size two. Performance for the other object in the array, when tested in VLTM, was considerably worse, with participants averaging 81.7% correct (SD = 6.3%), $t(16) = 5.08$, $p < .001$. Thus, despite objects being encoded and tested in exactly the same way, VWM demonstrated better performance than VLTM.

Subsequently, in Experiment 5b I observed that VWM performance was unaffected by noise at test, even when images were embedded in 75% noise. Participants averaged 97.2% correct (SD = 2.24%), which was not significantly different than performance observed in Experiment 1a, $t(34) = 0.42$, $p = 0.68$.

In contrast, VLTM performance was significantly affected by noise, $F(2, 36) = 20.28$, $p < 0.001$, $\eta_G^2 = 0.53$. As noise increased there was a clear drop in performance, with 78.5% correct (SD = 7.6%) at 25% noise, 74.3% correct (SD = 8.8%) at 50% noise, and 67.7% (SD = 7.6%) at 75% noise (see Figure 17). Thus, as the amount of noise at test increased, VLTM performance decreased in a linear fashion.

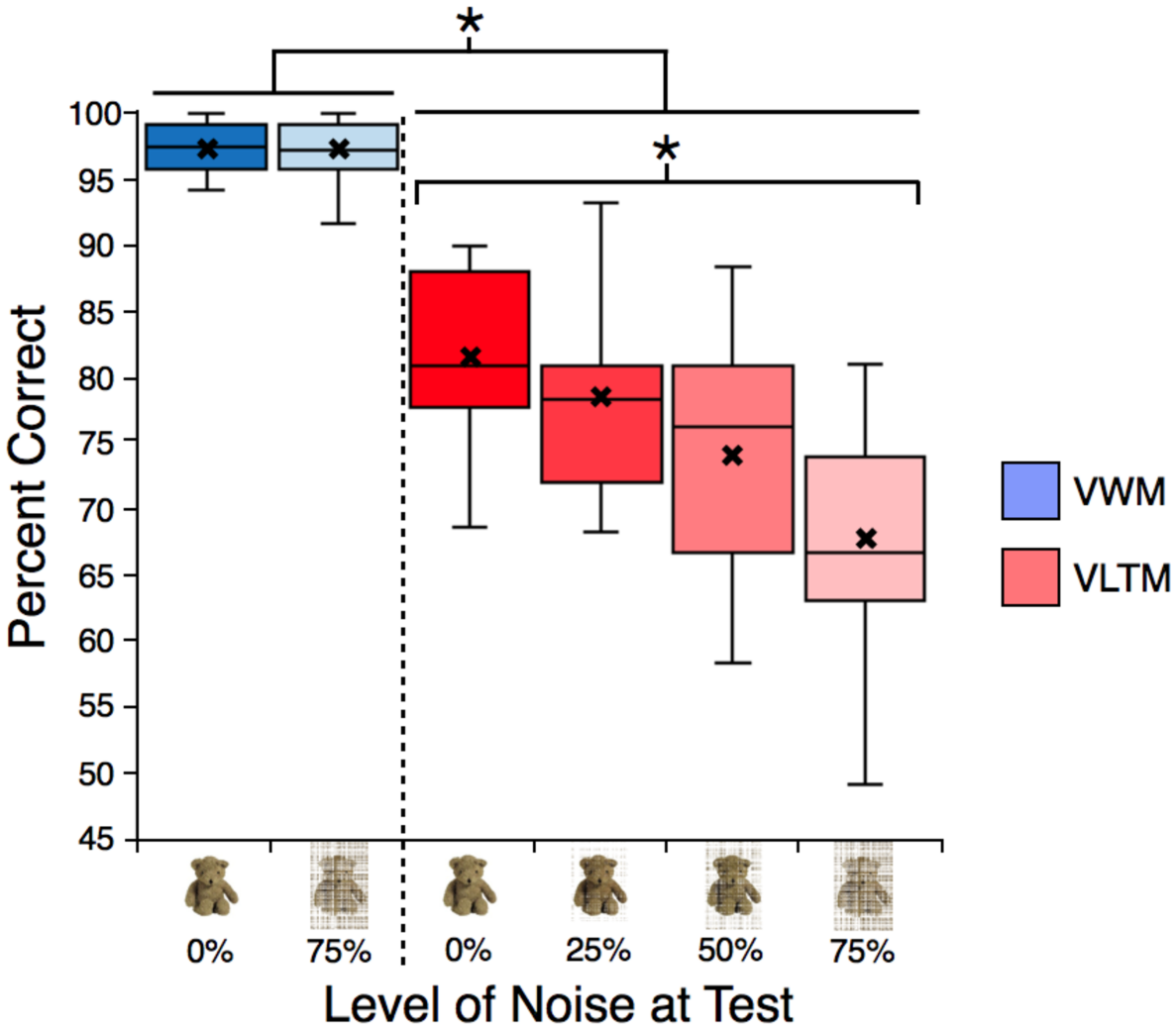

Figure 17. Results of Experiment 5a & 5b. The horizontal line within each box indicates the median, the boundaries of the box indicate the 25th and 75th percentile, and the whiskers indicate the highest and lowest values of the results. The "X" marked in the box indicated the mean. Across both experiments, memory performance was unaffected by noise level at test in VWM, with comparable performance across 0-75% noise. In contrast, VLTM was greatly affected by noise at test, with a linear decrease in performance observed across 0-75% noise.

3.4 Experiment 6: VWM Tolerance Extends to Orientation

In Experiment 5, I investigated how VWM and VLTM respond to noise at test in a literal way by injecting noise into test stimuli by randomly scrambling their pixels. More

generally, I am interested in investigating how resilient are VWM and VLTM object representations in the face of variability. And a common kind of variability that object recognition must address across encounters is orientation variability. In the real-world, we constantly recognize objects across changes in viewpoint, which introduces drastic changes to surface feature information (shape, color, line, texture, etc) into inputs arising from the same object between encounters (a key insight of the object recognition literature; DiCarlo & Cox, 2007; Pinto, Cox & DiCarlo, 2008; Rust & Stocker, 2010).

I therefore sought in Experiment 6 to evaluate how VWM and VLTM differed across recognizing a previously seen object either at the same or a completely new orientation relative to encoding. Experiment 6 was identical to Experiment 5, with the following exceptions: each presentation of an object during encoding was shown from a specific orientation. At test, "old" objects were shown either at the original encoding (relative to encoding) or a third, never-before-seen orientation (different).

Methods

Participants. A new group of 24 Johns Hopkins University undergraduates participated in Experiment 6. The results from one participant was excluded due to noncompliance with the instructions. All participants reported normal or corrected-to-normal visual acuity. Participation was voluntary, and in exchange for extra credit in

related courses. The experimental protocol was approved by the Johns Hopkins University IRB.

Stimuli, Apparatus and Procedure. All methods were identical to those in Experiment 5a, with the following exceptions: during the VWM task, participants were shown color images of objects from a specific orientation (taken from Geusebroek, Burghouts & Smeulders, 2005). At test, participants made a 2AFC judgment, but the old object was shown from the same or a different orientation. After 79 trials of this task, participants faced a surprise VLTM test. Again, at test participants made a 2AFC judgment, but the old object was shown either from the same or a different orientation.

Results

VWM performance was not significantly different for recognizing a previously seen object at the same (M = 96.11%) or different (94.50%) orientation, $t(22) = 0.77$, $p = 0.45$. VLTM performance at the same orientation averaged 66.67% correct, whereas recognition at a different orientation was significantly lower averaging 61.76%, $t(22) = 2.21$, $p = 0.038$ (see Figure 18). Thus, when recognizing an object at a new, never-before-seen orientation VLTM performance suffered, whereas VWM was incredibly tolerant to these changes.

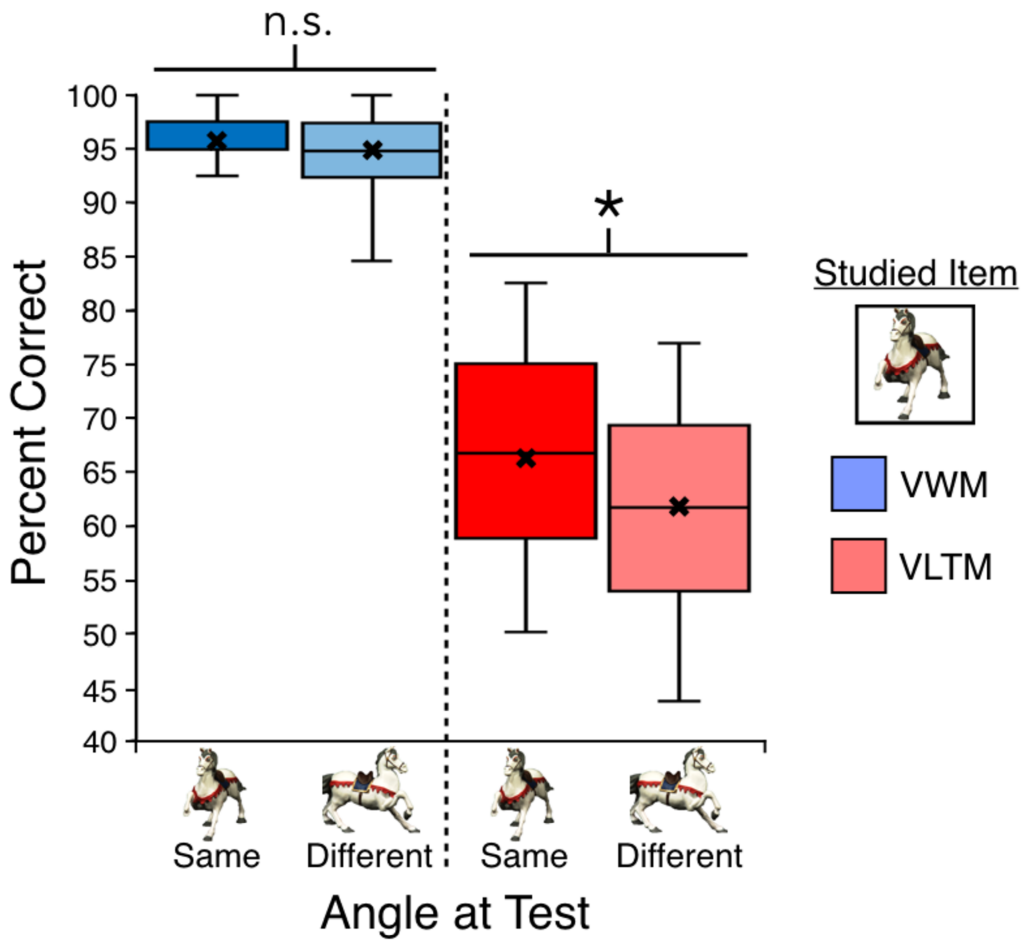

Figure 18. Results of Experiment 6. The horizontal line within each box indicates the median, the boundaries of the box indicated the 25th and 75th percentile, and the whiskers indicate the highest and lowest values of the results. The "X" marked in the box indicated the mean. VWM performance was equivalent across recognizing a previously seen object whether at the same or a new (different) orientation. In contrast, VLTM performance was significantly worse when recognizing a previously seen object at a different orientation than the one it originally appeared in.

3.5 Experiment 7: The Possible Role of VLTM Interference

While one of the goals of the previous experiments was to evaluate VWM and VLTM performance on equal footing, it remains possible that the tests in my paradigm are not equivalent. In a VWM trial participants must only remember two objects, whereas

over the long-term they must remember hundreds of previously encountered objects. This amount of load in long-term memory might create interference, thus resulting in poorer performance relative to VWM and possibly contributing to VLTM's lack of tolerance. For example, in verbal memory it's been found that recall performance for a single syllable (with load) does not degrade over time. However, as additional syllables are loaded into memory, performance decreases as the retention interval increases, suggesting that poorer long-term performance for verbal memory is the result of interference from holding multiple items in memory (Keppel & Underwood, 1962).

To address this possibility, I ran a one trial version of Experiment 5. Participants received only one trial of a VWM task, and then participated in a completely unrelated study that contained no images of real-world objects. Afterwards, they then received one trial of the surprise long-term test. If the poorer performance observed in VLTM was a result of interference from recalling multiple items, I should observe equivalent one-trial performance between VWM and VLTM.

Methods

Participants. A new group of 60 Johns Hopkins University undergraduates participated in Experiment 7. All participants reported normal or corrected-to-normal visual acuity. Participation was voluntary, and in exchange for extra credit in related courses. The experimental protocol was approved by the Johns Hopkins University IRB.

Stimuli, Apparatus and Procedure. All methods were identical to those in Experiment 5a, with the following exceptions: Participants received only one VWM trial. After completing this trial, they would then complete an unrelated study, which did not involve any stimuli containing real-world images of objects. After completing this study (approximately 40 minutes), they were then given one trial of the surprise VLTM test. Objects encoded and tested were drawn randomly from 378 possible images, and the "new" images in the 2AFC test were drawn randomly from a separate catalogue of 378 images.

Results

Performance across both VWM and VLTM was similar to previous experiments, with participants averaging 98.33% correct (SD = 1.67%) and 76.67% correct (SD = 5.51%), respectively. Astonishingly, this suggests performance was consistent with my previous experiments with only a single trial (see Figure 19). The poorer performance over the long-term in visual memory is not the result of being exposed to hundreds of potential memory items. To my knowledge, this is the first single trial experiment investigating object recognition, and provides a strong confirmation for the distinctive features of long-term recognition already established in the literature.

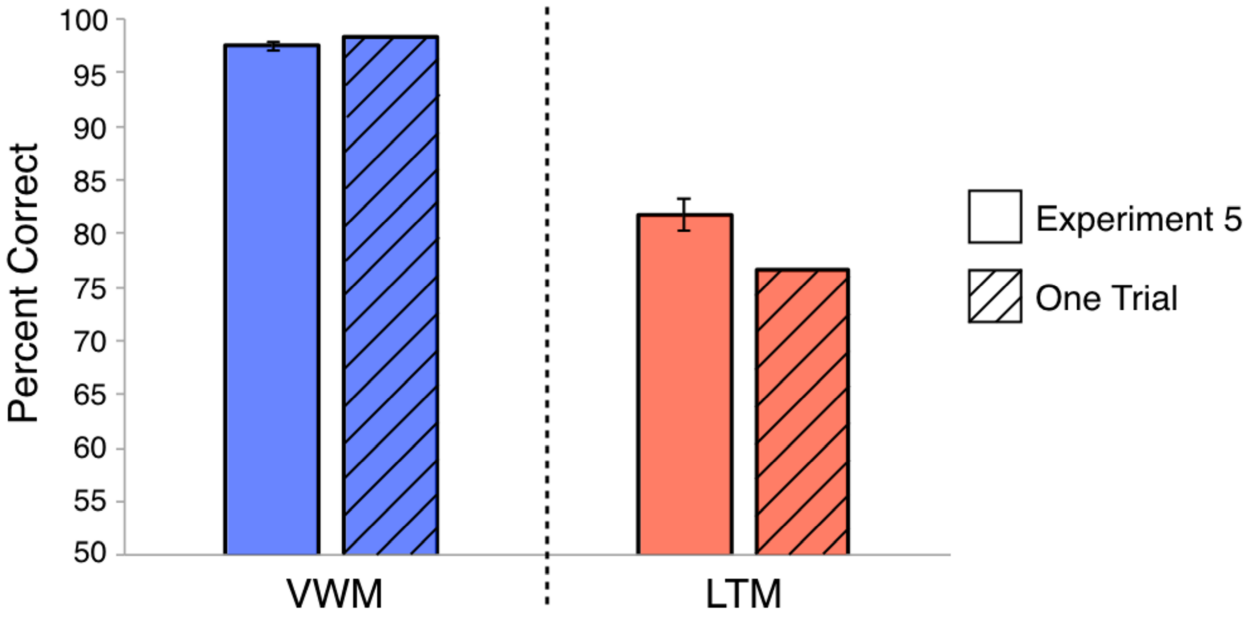

Figure 19. Results of Experiment 7, comparing one-trial performance to the baseline observed in Experiment 5. Even with a single trial, VLTM was significantly worse than VWM (consistent with Experiment 5), suggesting the present results cannot be explained by potential interference in long-term memory.

3.6 Other Controls

There are potentially several alternative explanations for the present results. In order to address the robustness of my results, as well as provide strong evidence in favor of my hypothesis that VWM demonstrates greater tolerance than VLTM, I ran several additional control experiments detailed below.

Ceiling Effects. Considering VWM performance was quite high throughout the present experiments, it remains possible that the greater tolerance demonstrated by VWM may be the result of ceiling effects. To evaluate this possibility, I ran a control where participants completed a VWM task at set size 4 (SS4) and were given test objects embedded in either 0% or 75% noise. Overall, performance at SS4 when test

images were shown without noise was 83.3%, which was significantly worse than SS2 in Experiment 1 (M = 97.5%), $t(35) = 11.24$, $p < 0.001$. In addition, I observed no significant difference for whether test images were embedded in 0% (M = 83.3%) or 75% (M = 83.6%) noise at test, $t(19) = 0.17$, $p = 0.87$. This suggests that the greater tolerance observed in VWM is not the result of potential ceiling or set size effects.

The Role of Color Information. Another potential explanation of the greater tolerance demonstrated by VWM is that participants may be utilizing trial-by-trial color information to improve performance at test, which is not a viable strategy during long-term memory tests. To evaluate this possibility, I replicated Experiment 1 but with completely grayscale images. Despite all stimuli being presented without color information, performance was remarkably similar to Experiment 1. Even when given 75% noise at test, participants averaged 96.8% correct (SD = 1.7%) in VWM. This was not significantly different than what was observed in the Experiment 1, $t(30) = 0.67$, $p = 0.51$.

VLTM performance followed a similar pattern as well, with a linear decrease in performance observed as noise increased from 25-75%. I observed no significant differences comparing individual noise levels to Experiment 5 (all p's > 0.60). Thus, utilizing color information in the short-term cannot account for the stark differences in performance observed in previous experiments. Additionally, these results suggest

color information may not be an important or critical component to the whole-object representations utilized in either VWM or VLTM.

Test Interference. It remains possible that the poorer performance observed in VLTM is not a reflection of the system per se, but rather the result of observers discarding trial information after being tested in VWM. To evaluate this possibility, I gave observers a VWM task where half of the time participants were tested after seeing an encoding display, whereas for the other half participants received no test. Participants were told during the VWM task that their memories for those items would later be tested.

I found no significant difference in VLTM for whether observers had seen that object in a display with a short-term test or no test. VWM performance was in line with previous experiments (98.6% correct, SD = 1.1%). For VLTM performance, there was no significant difference between whether a short-term test appeared after an array (M = 67.2%, SD = 1.4%) compared to whether there was no test (M = 70.1%, SD = 2.0%), $t(19) = 1.70$, $p = 0.11$. Overall participants averaged 68.6% correct in VLTM. This suggests that the worse performance observed in VLTM was not due to interference caused by the VWM task. Additionally, the present study also controlled for the surprise nature of the long-term test contributing to reduced performance.

3.7 Discussion

I found that VWM was robustly tolerant to variability at test, whereas VLTM was much more discriminating. What do these data say about the nature of human object recognition abilities? It is known that humans are capable of recognizing thousands of objects despite drastic changes to input properties across encounters (Wallis & Bulthoff, 2001; Cox et al., 2005; Brady et al., 2008; Cox & Dicarlo, 2008; DiCarlo & Cox, 2007; Rust & Stocker, 2010). It is theorized this tolerance is supported through temporal association, via mechanisms such as exploiting expectations of eye movements (Cox et al., 2005; Li & DiCarlo, 2008; Isik et al., 2012; Poth & Schneider, 2016; Poth, Herwig, & Schneider, 2015) or object physics (Schurgin et al., 2013; Schurgin & Flombaum, 2017). The present results demonstrate a natural prediction from these important theories: VWM is highly tolerant in order to act as a venue for subsequent integration. Further work will be necessary to understand the dynamics of this integration, but the results presented here offer a foundation describing the learning process supporting object recognition.

Pivoting Towards the Process of Object Recognition

A central concern of object recognition research has been focused on understanding the relative tolerance of our representations. One of the earliest theories to supply a description for producing this tolerance was Biederman's

Recognition-by-Components (RBC) theory. It suggests that object recognition is supported by a basic component, geons, which are derived from non-accidental properties in 2D images. Geons are viewpoint invariant, so they can be recognized despite variability across inputs and subsequently matched to object representations (Biederman, 1987; Biederman & Gerhardstein, 1993). While there remain several limitations of RBC theory, geons do provide a descriptive vocabulary for how to use raw materials to recognize previously seen and new objects.

This descriptive approach has been applied broadly throughout the object recognition literature, even in the context of designing artificial intelligence (A.I.) object recognition systems. Traditionally, programmers would create a potential description of an object and then assess how successful that description was in recognizing another object. These descriptions were typically defined using specific surface feature information (texture, creating shape skeletons), along with other secondary factors, and laid the foundation for designing object recognition A.I. systems (Belongie, Malik & Puzicha, 2002; Shotton, Winn, Rother & Criminisi, 2006; Zhang, Marszalek, Lazebnik & Schmid, 2007).

Recently, A.I. object recognition systems have seen tremendous advances through deep learning neural networks (Krizhevsky, Sutskever & Hinton, 2012), even surpassing human-level performance under specific conditions (He, Zhang, Ren & Sun, 2015). In traditional programmed A.I. models there is no real learning – the

representational structure is derived by computer scientists who then implement it in the code. The general approach of deep learning neural networks it that computer scientists implement a learning algorithm, but the actual structure of the representation acquired may remain unknown. The power and success of deep learning neural networks is through understanding the potential learning process of object recognition, not necessarily the representational content.

I argue a similar pivot should be made in the psychological literature. Currently psychologists have focused on testing kinds of representations and algorithms to see if they can produce tolerance, similar to traditional A.I. models. But little work has considered how such representations are acquired. Scientists studying object recognition should be asking how do object representations develop – How are they *learned*? For example, what if RBC theory is how humans recognize objects. How then are geons learned? How do we learn to build an object out of its constituent geons? There remains very little research on how we learn to recognize objects, outside of developmental research investigating infant and toddler abilities to perceive and identify objects (Spelke, 1990; Spelke et al., 1995; Xu & Carey, 1996; Xu, Carey & Quint, 2004; Cheries, Wynn & Scholl, 2006; Stahl & Feigenson, 2015).

Characterizing the Learning Process of Object Recognition

In the present experiments, I have made no proposal as to what kind of representation may produce tolerance. I assume they exist, and have asked a different but related question – how are these representations learned? Specifically, how tolerant are these representations initially?

Recent theories suggest part of the encoding process supporting object recognition over the long-term involves integrating information about an object over brief encounters (Wallis & Bulthoff, 1999; Cox et al., 2005; Cox & DiCarlo, 2008; Schurgin & Flombaum, 2017). Humans appear to temporally associate information to build object representations through saccades (Cox et al., 2005; Li & DiCarlo, 2008; Isik et al., 2012; Poth & Schneider, 2016; Poth, Herwig, & Schneider, 2015), as well as through exploiting expectations about object physics (Schurgin et al., 2013; Schurgin & Flombaum, 2017). This suggests that tolerance in the long-term may ultimately depend on experience.

A natural prediction that follows these important theories is that memory over the short-term for objects should actually be highly tolerant – accepting a large amount of variability – in order to act as a venue for integration. This is because over the short-term the visual system can rely on the expectation that objects in a scene will remain the same. Thus, it can discount potential differences in appearance to integrate information over time to create appropriately constrained representations in long-term

memory. The logic of this novel hypothesis is surprisingly similar to how deep learning neural networks operate – in a sense they are high tolerant before being trained.

I observed through a series of experiments that when given a single exposure to an object VWM demonstrates remarkable tolerance to variability at test, whereas VLTM is much more discriminating. This was true whether variability was introduced in the form of noise (Experiment 5) or orientation (Experiment 6), and could not be explained by set size limitations, ceiling effects, utilizing color information, the surprise nature of the long-term test, or VWM tests causing interference for subsequently tested items (see Controls). Moreover, I provided a single trial adaptation of my experiment demonstrating the robustness of the present results and that the poorer performance observed in VLTM cannot be explained through potential interference from observers being exposed to hundreds of previous-seen objects (Experiment 7).

The current data provides evidence that the tolerance demonstrated by VLTM is in fact inherited from an even more highly tolerant VWM system. This greater tolerance demonstrated by VWM is likely the result of another challenging facing object recognition over the long-term: constructing object representations with discriminatory power — the ability to reject as matches objects that look the same but are not. As shown across all the present experiments, object representations appear to become more discriminating as they consolidate into long-term memory.

Implications for VWM and VLTM

The present results have implications for both VWM and VLTM, and more importantly, for how they may relate. Previous psychophysical, neurological, and modeling evidence has found clear distinctions separating VWM and VLTM into dissociable systems (Cowan, 2008; Brady, Konkle & Alvarez, 2011). However, this does not mean that they are completely independent of one another and do not interact. Indeed, working memory maintenance is a critical step for long-term encoding (Ranganath, Cohen & Brozinsky, 2005), information from VLTM may affect VWM performance (Brady, Konkle & Alvarez, 2011; Curby & Guathier, 2007; Curby, Glazek & Gauthier, 2009), and the precision of color information in VWM and VLTM is similar under certain circumstances (suggesting a common constraint; see Brady et al., 2013).

My data proposes we should seek to characterize these systems, and their relationship to one another, by their functions as opposed to their time course. Generally, VWM or VLTM have been distinguished by the time scale over which the memory takes place (Squire, 2004; Cowan, 2008; Brady, Konkle & Alvarez, 2011). While time may constrain systems in different ways it is not sufficient to understand all possible differences and similarities between the systems. If we know the function of a system, we can clearly identify the specific challenges facing such functions, even if we do not know the specific solutions to these challenges a system may choose to employ.

Here I have shown that one of the functions of VWM is to support VLTM. Specifically, VWM should be highly tolerant of variability to learn what are the stable and diagnostic features of an object. This is in many ways counterintuitive – VLTM is typically described in terms of its tolerance, whereas VWM is often evaluated and characterized using change detection tasks, which are tests of discrimination. It suggests that what some may have interpreted as limitations or noise in VWM may actually be features to support subsequent long-term memories.

I have also supplied an ideal paradigm for researchers interested in investigating the relationship between VWM and VLTM. Previous research has remained inherently limited, as they have either focused on memory for a single feature and rarely equated encoding and testing conditions across memory type. My current paradigm not only places VWM and VLTM on equal footing, so they can be directly compared, but also allows researchers to holistically characterize visual memory for objects.

Conclusion

By completely equating the encoding and test across VWM and VLTM, I found that VWM demonstrated remarkably greater tolerance to variability at test than VLTM. These results suggest that VWM is tolerant in order to integrate information that will appropriately constrain representations in the long-term. This raises important questions about the learning process supporting object recognition. In particular, my

results suggest that part of the function VWM is to support subsequent object recognition ability.

Chapter 4

How does integration take place over time?

4.1 Synopsis

Another obvious question arises when one starts to look at the typical methods utilized by researchers studying memory: experiments tend to focus on the nature of visual long-term memory following singular exposures (Shepard, 1967; Brady et al., 2008; Brady et al., 2009; Cansino et al., 2002; Konkle et al., 2010; Guerin et al., 2012; Brady et al., 2013; Kim & Yassa, 2013; Reagh & Yassa, 2014a; Cunningham, Yassa & Egeth, 2015). However, the formation of visual memories likely evolves over time and over repeated encounters with stimuli. Your memory of a best-friend, a favorite childhood toy, the view from a favorite window, presumably do not go unchanged after your first encounter with each. Instead, each encounter likely leaves a trace on the nature of your

memory. And importantly, every experience you have with a stimulus will not be exactly the same.

How then does integration take place over time? I sought to further understand the growth process of memory representations by examining how memories change after re-exposure. In the context of object recognition, I can ask where does tolerance come from? How does it change as we learn more about an object? We do know that seeing an object more than once improves later memory performance for that object (Hintzman, 1976; Reagh & Yassa, 2014b), but many basic questions remain about how visual long-term memory evolves over multiple experiences with stimuli. I sought to expand our basic understanding of VLTM from this point.

4.2 Experiment 8: Variable Quality Across Encounters

Initially, I considered the fact that experiences with objects are likely to vary in quality. Therefore, asked whether memory benefits when a lower quality encounter either precedes or follows a higher quality one? It may be that the last encounter an observer has with an object has the greatest bearing on subsequent long-term memory performance. In contrast, it may be that when observers are first given a high quality encounter this creates conditions better able to assimilate information in the future, resulting in more robust memories. Finally, it may also be the case that our long-term

memory system is able to integrate information ideally, so it does not matter whether a lower quality encounter either precedes or follows a higher quality one.

In order to address this question, I varied the amount of information available when observers encounter an object twice during the encoding phase of a recognition experiment. I did this by inserting either a low or high amount of noise into each encounter through randomly scrambling a percentage of the pixels in the image (low = 40% noise, high = 70% noise). This created four dual-encounter conditions (in addition to two single encounter conditions): both presentations in low noise, both presentations in high noise, an initial presentation in high and then low noise (high-low), and an initial presentation in low and then high noise (low-high). I predicted that performance would be better in the low-high than the high-low conditions, as the initial encounter in low noise would provide a better representation to integrate subsequent (noisy) information.

Methods

Participants. A group of 22 Johns Hopkins University undergraduates participated in Experiment 8. The results of 2 participants were excluded due to responding randomly (i.e. at chance performance). All participants reported normal or corrected-to-normal visual acuity. Participation was voluntary, and in exchange for extra

credit in related courses. The experimental protocol was approved by the Johns Hopkins University IRB.

Apparatus. Experiment 8 took place in a dimly lit sound-attenuated room. Stimuli were presented on a Macintosh iMac computer with a refresh rate of 60 Hz. The viewing distance was 60 cm so that the display subtended 39.43° × 24.76° of visual angle.

Stimuli and Procedure. Stimuli were generated using MATLAB and the Psychophysics toolbox (Brainard, 1997; Pelli, 1997). All stimuli were presented within the full display frame of 39.43° × 24.76°. Participants completed a visual object recognition memory task that included two stages. During the first phase (incidental encoding), participants were shown 300 color images of real-world objects on a computer screen with the cover task of indicating whether an item onscreen was an "indoor" or "outdoor" item. They indicated their responses using the computer keyboard. Each image was on-screen for 2000 ms, with a 500 ms ISI. Responses were only recorded if the stimulus was on-screen.

All images of objects were shown once or twice across the entire task (organized into two randomized blocks containing the same images, unknown to the participants). And the images were embedded in noise by randomly scrambling a percentage of their pixels. I varied the amount of noise embedded in each image to be low (40%) or high (70%). This created six memory quality conditions: a single presentation in low

noise (Single Low), a single presentation in high noise (Single High), when both presentations of an object were embedded in high noise (High-High), when both presentations were embedded in low noise (Low-Low), when the first presentation was embedded in low noise and the second presentation was in high noise (Low-High), and when the first presentation was in high noise and the subsequent presentation was in low noise (High-Low). Participants completed a total of 600 trials (60 each for images shown once, 120 each for images shown twice).

During each trial of the second phase (surprise retrieval), a previously viewed object from encoding was paired with a new object, and participants reported the one that was 'old,' that is, the one that appeared at some point in the encoding phase. For half of the trials the new image was a completely new, categorically distinct object (Old-New comparison) and for the other half of trials the new images was a categorically similar-looking object (Old-Similar comparison). Images were presented on-screen until participants made a response.

Data Analysis. I did not include responses that were faster than 200 ms, which accounted for a total of 0.08% of trials in Experiment 8.

Results

A between-subjects ANOVA revealed a main effect of distance on Old-New comparisons, $F(5, 95) = 24.71$, $p < 0.01$, $\eta_p^2 = 0.57$. Performance generally increased

from single high exposure (M = 80.8%), single low exposure (M = 89.1%), high-high (M = 86%), high-low (M = 91.5%), low-high (M = 93.7%), to low-low (M = 95.5%) conditions (see Figure 20). To verify the significance of this pattern, I ran a linear contrast which revealed a significant effect, $F(1, 19) = 86.91$, $p < 0.01$, $\eta_p^2 = 0.82$. Planned comparisons found a trending difference between High-Low and Low-High conditions, $t(19) = 1.86$, $p = 0.07$. I also observed a significant difference between High-High and High-Low conditions, $t(19) = 2.95$, $p < 0.01$, and a trending difference between Low-High and Low-Low conditions, $t(19) = 1.76$, $p = 0.09$.

A between-subjects ANOVA revealed a main effect of distance on Old-Sim comparisons, $F(2, 36) = 12.36$, $p < 0.01$, $\eta_p^2 = 0.99$. Again, performance increased from single high exposure (M = 61.9%), single low exposure (M = 65.9%), high-high (M = 70.2%), high-low (M = 74%), low-high (M = 77.8%), to low-low (M = 82.9%) conditions. To verify the significance of this pattern, I ran a linear contrast which revealed a significant effect, $F(1, 19) = 82.55$, $p < 0.01$, $\eta_p^2 = 0.81$. Planned comparisons found a trending difference between High-Low and Low-High conditions, $t(19) = 1.88$, $p = 0.07$. I also observed a trending difference between High-High and High-Low conditions, $t(19) = 2.03$, $p = 0.06$, and a significant difference between Low-High and Low-Low conditions, $t(19) = 2.61$, $p = 0.02$.

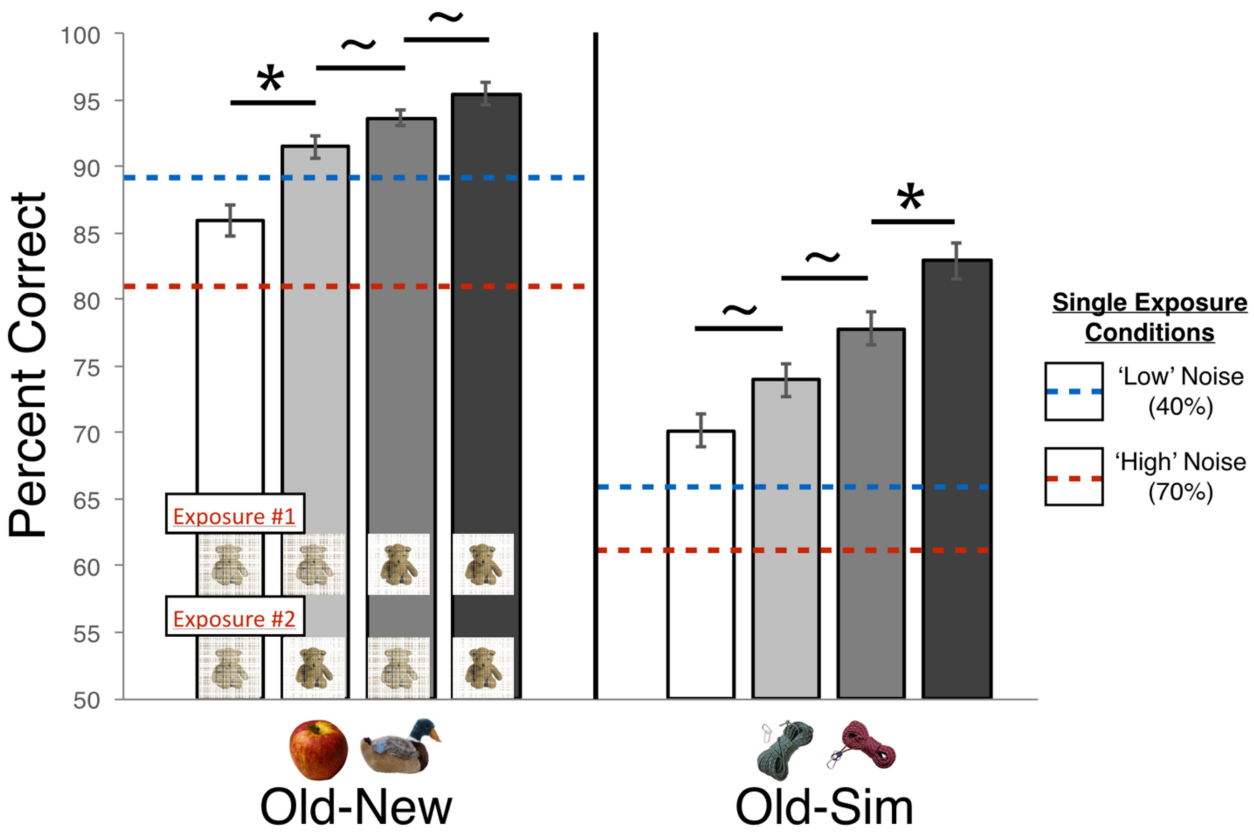

Figure 20. Results of Experiment 8. A linear contrast across both Old-New and Old-Sim comparisons revealed performance increased from High-High (white), High-Low (light gray), Low-High (dark gray), to Low-Low (black). Additional contrasts were conducted, verifying trending (~ $p < 0.1$) or significant (* $p < 0.05$) effects. Error bars represent within-subject error.

While the current results demonstrating better performance in the Low-High compared to High-Low condition are consistent with my integration hypothesis, there remain limitations to this interpretation. In particular, it may be that the noise level for encounters in high noise was too high. Under this account, the better performance for Low-High was not the result of integration per se, but that first encountering an object under better conditions allowed for future recognition under high noise. In the High-Low condition, observers were not able to recognize the object during its first

encounter, and were simply relying on the last encounter. If this is the case, I should then see comparable performance across High-Low and the single encounter low conditions. For Old-New comparisons, I observed a trend that objects encountered under the High-Low (M = 91.5%) condition led to better performance than a single low noise encounter (M = 89.1%), $t(19) = 1.78$, $p = 0.09$. For Old-Sim comparisons, I observed a significant effect that High-Low (M= 74%) was better than a single low noise encounter (M = 65.9%), $t(19) = 3.46$, $p < 0.01$. Thus, it appears the poorer performance of High-Low compared to Low-High is not the result of observers simply relying on information from the last encounter.

Overall, across both Old-New and Old-Sim comparisons, I found a linear increase in performance from single high, single low, high-high, high-low, low-high to low-low conditions. Consistent with my hypothesis, for both Old-New and Old-Sim comparisons, performance was significantly better for objects initially encountered in low and then high noise compared to those encountered first in high and then low noise. This suggests that, even though the most recent exposure to the object involved a degraded image (high noise), memory performance was better due to the strength of the initial representation that supported the subsequent integration of this degraded input.

4.3 Experiment 9: Variable Orientation Across Encounters

In Experiment 8 I demonstrated that when seeing an object multiple times the quality of encounters matters for long-term memory, with an initial high-quality encounter creating a better representation to integrate subsequent (noisy) information. Another kind of variability observers must deal with when encountering the same object across multiple encounters are changes in orientation– the same object will rarely (if ever) be seen from the exact same orientation as before.

I sought to investigate what happens when observers see an object twice from the same or two different orientations. Then, at test I probed their memories of an object using a previously encountered image (i.e. same orientation) or an image of a previous object from a new, never-before-seen orientation (i.e. different orientation). I hypothesized that when observers see an object from two different orientations across encounters, this should create an opportunity to construct a more tolerant representation. Thus, I should observe better performance for recognizing an object at a new orientation for objects that were encountered along two different (as opposed to two of the same) orientations.

Methods

Participants. A new group of 20 Johns Hopkins University undergraduates participated in Experiment 9. All participants reported normal or corrected-to-normal

visual acuity. Participation was voluntary, and in exchange for extra credit in related courses. The experimental protocol was approved by the Johns Hopkins University IRB.

Stimuli, Apparatus and Procedure. All methods were identical to those in Experiment 8, with the following exceptions: At encoding, participants judged whether the object was more "square" or "round." This change was made because the stimulus set used included an overwhelming number of objects that would have been rated as "indoor" objects. Participants viewed a total of 152 color images of objects without noise (taken from Geusebroek, Burghouts & Smeulders, 2005). Each object was shown twice (again organized into two randomized blocks, unbeknownst to participants), but either from the same orientation (same orientation condition) or two different orientations (different orientation condition).

During the surprise test, participants only made Old-New comparisons. However, the image of the old object was either the same as encountered during encoding (Original-New), or the object shown from a completely new orientation (Different-New).

Data Analysis. I did not include responses that were faster than 200 ms, which accounted for a total of 0.03% of trials in Experiment 9.

Results

For Original-New test performance, I failed to find a difference across same orientation (M = 90.92% SD = 9.83) and different orientation (M = 90.26%, SD = 7.10) conditions, $t(19) = 0.39$, $p = 0.70$. And for Different-New test performance, I failed to find a difference across same orientation (M = 87.4%, SD = 9.00) and different orientation (M = 87.8%, SD = 9.86) conditions, $t(19) = 0.23$, $p = 0.82$. Thus, it appears that encountering an object twice from two different orientations did not create more tolerant memories.

When comparing different test performance across each condition, I did find a significant effect that in the same orientation condition performance was significantly better for Original-New than Different-New tests, $t(19) = 2.27$, $p = 0.04$. A similar trend was observed in the different orientation Condition, $t(19) = 1.99$, $p = 0.08$. This suggests that overall, performance was slightly better across both same and different orientation conditions at recognizing an object at a previously seen orientation.

Across both types of tests, I failed to find a difference across seeing an object twice from the same or two different orientations. While performance for recognizing an object at a previously seen orientation was slightly better than recognizing an object at a new orientation, performance was actually quite high for both types of recognition test (~88-90%). This suggests that, in both the same and different orientation conditions, participants were able to build robustly tolerant visual long-term memories.

It appears our ability to construct tolerant memories is not dependent on whether we encounter objects multiple times from the same or different orientations.

4.4 Experiment 10: Variable Distance Across Encounters

While the quality or orientation between encounters may affect subsequent memory performance, another important factor is the variable amount of time (or distance) between encounters. Spacing effects have been explored in the context of episodic memory (Melton, 1970; Zhang & Byrne, 2016), but little is known about how variable distance across encounters may affect *visual* long-term memory — is it better for a second encounter with an object to take place immediately, or with some time in-between?

I explored this question by showing objects twice during an incidental encoding paradigm, but by varying the amount of distance between repetitions. Objects were either repeated after a brief delay (2-back, 3 seconds between encounters) or a longer delay (10-back, 30 seconds between encounters). I predicted that subsequent performance at test would be better for items presented in the 10-back compared to 2-back conditions, as this would allow for more time to consolidate initial representations in memory.

Methods

Participants. A new group of 21 Johns Hopkins University undergraduates participated in Experiment 10. The results of 2 participants were excluded due to responding randomly (i.e. at chance performance). All participants reported normal or corrected-to-normal visual acuity. Participation was voluntary, and in exchange for extra credit in related courses. The experimental protocol was approved by the Johns Hopkins University IRB.

Stimuli, Apparatus and Procedure. All methods were identical to those in Experiment 8, with the following exceptions: At encoding, participants viewed a total of 360 color images of objects without noise. Each image was shown either once (120 images), twice after a brief delay (120 images, 2-back), or twice after a longer delay (120 images, 10-back), for a total of 600 trials during encoding.

Data Analysis. I did not include responses that were faster than 200 ms, which accounted for a total of 0.23% of trials in Experiment 10.

Results

A between-subjects ANOVA revealed a main effect of distance on Old-New comparisons, $F(2, 36) = 11.15$, $p < 0.01$, $\eta_p^2 = 0.99$. Follow-up analyses revealed that 10-back performance (M = 96.1%, SD = 3.8) was significantly better than 2-back performance (M = 93.9%, SD = 5.4), $t(18) = 2.44$, $p = 0.02$. In addition, 2-back

performance was significantly better than a single exposure (M = 91.1%, SD = 8.2), t(18) = 2.68, p = 0.01 (see Figure 21).

A between-subjects ANOVA also revealed a main effect of distance on Old-Sim comparisons, F(2, 36) = 12.36, p < 0.01, η_p^2 = 0.99. Follow-up analyses revealed a trending difference between 10-back (M = 80.2%, SD = 7.6) and 2-back (78.4%, SD = 6.5) performance, t(18) = 1.39, p = 0.18. In addition, 2-back performance was significantly better than a single exposure, (M = 73.5%, SD = 7.6), t(18) = 3.13, p < 0.01.

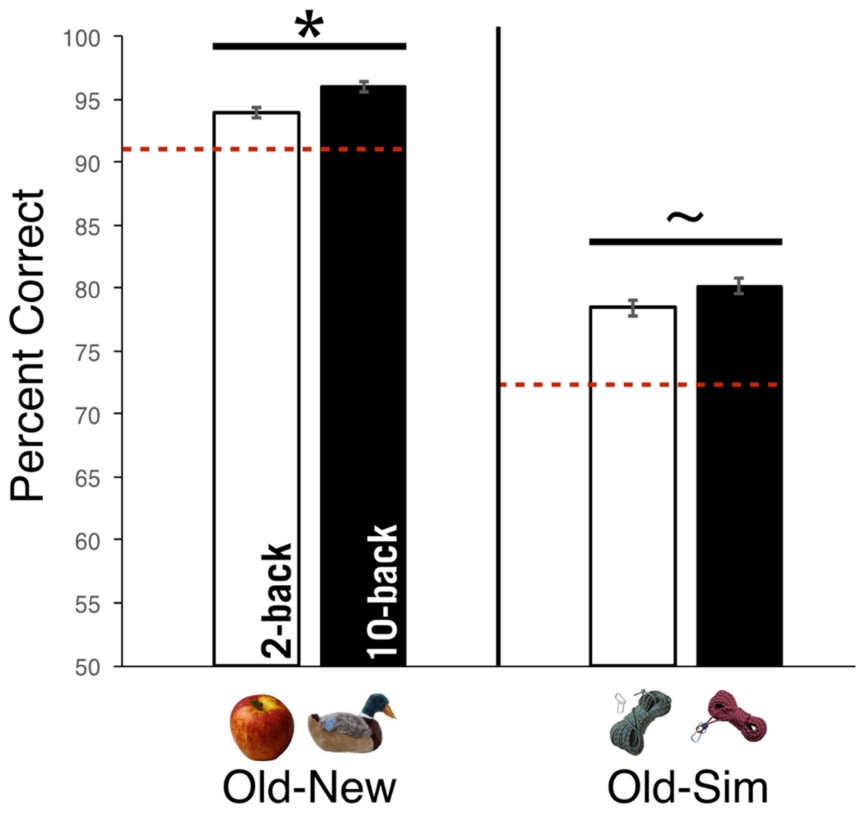

Figure 21. Results for Experiment 10. Across both Old-New and Old-Sim comparisons, performance increased from the single encounter condition (red dashed-line) to the 2-back condition (white), and was best at the 10-back condition (black). * designates p < 0.05, ~ designates p < 0.20. Error bars represent within-subject error.

Overall, performance was better for objects encountered twice (2-back or 10-back) compared to once (single exposure). More importantly, performance was significantly better for objects encountered in the 10-back compared to the 2-back condition. This suggests that when additional time between encounters was introduced subsequent memory performance benefited. This longer delay likely allowed for further consolidation of the initial representation, which was then able to better integrate subsequent encounters.

4.5 Discussion

While the majority of research investigating visual long-term memory has been concerned with singular experiences, the formation of our visual memories in the real-world develops over time and over repeated encounters with stimuli. I sought to expand our understanding of the growth process of memory through examining how memories change after re-exposure. What are the potential variables that improve our memory strength or tolerance over time?

I observed that both the quality of an observer's initial encounter with an object (Experiment 8) as well as the time between encounters (Experiment 10) affect memory performance. Interestingly, I failed to find that encountering an object multiple times at different orientations creates more tolerant memories (Experiment 9). The present

findings offer a foundation identifying several factors that change the quality of our visual memories over repeated exposures.

Implications for Memory Consolidation

There remains much that is unknown about the consolidation process of representations into visual long-term memory. Previous research has shown that working memory maintenance appears to be critical step to subsequent long-term consolidation (Ranganath, Cohen & Brozinsky, 2005). However, other factors such as the total number of items to be remembered (whether 20 or 360 items) or increasing exposure past 1 second (to either 3 or 5 seconds) appear to have no effect on subsequent long-term memory performance (Brady et al., 2013).

In Experiment 10, I demonstrated that when given two encounters of the same object, memory performance is better when additional time in introduced between encounters. This is largely consistent with the learning literature in other domains, which has demonstrated additional consolidation time generally improves performance (Melton, 1970; Smolen, Zhang & Byrne, 2016). A limitation of the present research is that only two variable distances were explored (3 vs 30 seconds). It remains unknown whether additional consolidation time between encounters improves performance further, or at what time point memory performance plateaus.

Recently, research has shown different effects of variable distance in other types of visual memory tasks. For instance, when observers are given a continuous recognition task where they see a long sequence of object images and are tasked with detecting repeated images, memory performance declines as the distance between repeated items increased (Singh, Oliva & Howard, 2017). Initially, these results may seem incompatible with the present findings, but the confines and computational challenges facing a continuous recognition task are quite unique. Generally, in long-term memory (or object recognition) observers are not consciously attending to a stream of stimuli presented in a specific order with the sole task of identifying variably spaced repeats. Such a task requires additional cognitive components beyond memory, such as executive functions and working memory. Visual long-term memory is typically described as a passive, automatic system (Squire, 2004; Brady, Konkle & Alvarez, 2011), and therefore these results are likely specific to the constraints associated with recognizing repeated images in a continuous timeline.

Insight into Tolerance

Recent theories suggest part of the encoding process supporting object recognition over the long-term involves integrating information about an object over brief encounters (Wallis & Bulthoff, 1999; Cox et al., 2005; Cox & DiCarlo, 2008; Schurgin & Flombaum, 2017). Humans appear to temporally associate information to

build object representations through saccades (Cox et al., 2005; Li & DiCarlo, 2008; Isik et al., 2012; Poth & Schneider, 2016; Poth, Herwig, & Schneider, 2015), as well as through exploiting expectations about object physics (see Chapter 2; Schurgin & Flombaum, 2017). This suggests that tolerance in the long-term may ultimately depend on experience.

While tolerance may be learned through experience in the short-term, the present results suggest that building tolerant representations does not depend on multiple encounters with an object over longer periods of time. In Experiment 9, observers constructed robustly tolerant memories regardless of whether they saw an object multiple times from the same or different orientations. This demonstrates that memory representations tolerant to changes in orientation can be rapidly extracted without much exposure to variability across multiple encounters.

Initially this may appear somewhat surprising, but is ultimately consistent with previous research investigating tolerance in human object recognition. When briefly flashed a photo, observers can rapidly (<300 ms) recognize objects despite never having viewed the photograph before (Potter, 1976). This result is consistent whether observers are shown color photographs or simplified line drawings of objects, suggesting tolerance for recognizing objects can be rapidly extracted for even abstract stimuli (Biederman, 1987; Biederman & Ju, 1998). Considering how rapidly humans can recognize objects despite considerable variability to inputs (including line drawings

devoid of any other surface feature information), it makes sense that tolerance does not depend on learning through multiple prolonged encounters with an object.

Conclusion

These results have broad implications for researchers studying memory, learning, and object recognition. They demonstrate the importance of moving past singular experiences to investigate the integration of multiple encounters, and they begin to elucidate the nature of the mechanisms that integrate new input into long-term memories. Specifically, they suggest that these mechanisms function more effectively given an opportunity to consolidate initial encounters, consistent with many other types of non-visual learning (Melton, 1970; Smolen, Zhang & Byrne, 2016). Additionally, the results suggest that integration works best when an initial encounter supplies a high-quality basis to assimilate with subsequent experiences.

Chapter 5

General Discussion

5.1 Summary

Humans have a remarkable ability to recognize visual objects following limited exposure, and despite drastic changes to inputs across encounters. This capacity is sometimes called 'tolerance' or 'invariance.' How we acquire this ability remains a mystery, and it remains an area in which artificial systems have yet to match human performance.

In my dissertation, I explored how the mind acquires the representational qualities necessary to recognize objects. First, I demonstrated that mechanisms supporting the perception of objects in VWM (core knowledge / spatiotemporal continuity) are also critical to the construction of representations in VLTM (Chapter 2: Experiments 1-4). Next, I demonstrated that VWM was much more tolerant to

variability at test than VLTM, suggesting VWM acts a venue for integrating information into long-term memory (Chapter 3: Experiments 5-7). Finally, I extended my experiments beyond single exposures to further understand the construction process of visual memories over repeated encounters. Across two experiments, I found that both the initial quality of an encounter and the amount of space between encounters affects consolidation and future performance in VLTM (Chapter 4: Experiments 8, 10). I also discovered that the visual system can rapidly and efficiently construct tolerant VLTM representations without variable input across encounters (Chapter 4: Experiment 9).

5.2 Towards a unified approach

Through trying to define "systems" of memory, many fields of research have become fixated on only investigating a specific subset of cognitive mechanisms likely related to our memory abilities. For example, I have shown how mechanisms typically associated with the perception of objects may be critically important to the formation of visual memories (Chapter 2: Experiments 1-4). Considering most neuroscientists studying long-term memory attempt to discard potential influences of perception, they are likely failing to fully engage learning systems involved in the formation of visual memories.

A more fruitful approach to the study of memory would be to create conditions that facilitate collaboration and communication between different fields and types of

researchers. Part of the strength of the results and discoveries in the present dissertation came through leveraging theories and methods from multiple fields. In Chapter 2, I took insight from vision scientists studying object perception and applied it to a typical neuroscience memory experiment. And in Chapter 3, I applied the logic of recent approaches used in machine learning and deep learning neural networks to advance our understanding of visual memory. Future research would greatly benefit from utilizing insights from different fields in order to solve shared questions pertaining to visual memory and cognition more broadly.

5.3 Concluding remarks

In short, my work on visual memory has uncovered new mechanisms for how humans acquire the representational qualities necessary to support invariant recognition, and it has provided a unifying framework to relate traditionally disparate fields of vision scientist studying object perception, neuroscientists studying long-term memory, and engineers designing artificial intelligence object recognition systems. In this dissertation, I have provided several novel paradigms investigating how our visual memories are acquired and change over time. Altogether, the current results provide strong evidence in favor of an alternative approach to memory research, moving away from defining the content and representational structures supporting memory with an

emphasis towards understanding the process through which our memories *acquire* such qualities.

References

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

Andreopoulos, A., & Tsotsos, J. K. (2013). 50 years of object recognition: Directions forward. *Computer Vision and Image Understanding*, *117*(8), 827-891.

Anstis, S.M. (1980). The perception of apparent movement. *Philosophical Transactions of the Royal Society of London B, 290*, 153-168.

Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, *18*(7), 622–8.

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature reviews neuroscience, 4*(10), 829-839.

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of learning and motivation, 8*, 47-89.

Bakker, A., Krauss, G. L., Albert, M. S., Speck, C. L., Jones, L. R., Stark, C. E., ... & Gallagher, M. (2012). Reduction of hippocampal hyperactivity improves cognition in amnestic mild cognitive impairment. *Neuron*, *74*(3), 467-474.

Baillargeon, R., & Hanko-Summers, S. (1990). Is the top object adequately supported by the bottom object? Young infants' understanding of support relations. *Cognitive Development*, *5*(1), 29-53.

Baillargeon, R., Spelke, E. S., & Wasserman, S. (1985). Object permanence in five-month-old infants. *Cognition, 20*(3), 191-208.

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, *321*(5890), 851-854.

Bays, P. M., Catalao, R. F., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of vision*, *9*(10), 7-7.

Belongie, S., Malik, J., & Puzicha, J. (2002). Shape matching and object recognition using shape contexts. *IEEE transactions on pattern analysis and machine intelligence*, *24*(4), 509-522.

Biederman, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological review, 94*(2), 115.

Biederman, I., & Gerhardstein, P. C. (1993). Recognizing depth-rotated objects: evidence and conditions for three-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human perception and performance, 19*(6), 1162.

Blumenfeld, R. S., & Ranganath, C. (2006). Dorsolateral prefrontal cortex promotes long-term memory formation through its role in working memory organization. The *Journal of Neuroscience, 26*(3), 916-925.

Botvinick, M. M., & Plaut, D. C. (2006). Short-term memory for serial order: a recurrent neural network model. *Psychological review*, *113*(2), 201.

Bowles, B., Crupi, C., Pigott, S., Parrent, A., Wiebe, S., Janzen, L., & Köhler, S. (2010). Double dissociation of selective recollection and familiarity impairments following two different surgical treatments for temporal-lobe epilepsy. *Neuropsychologia*, *48*(9), 2640-2647.

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences, 105*(38), 14325-14329.

Brady, T. F., Konkle, T., Oliva, A., & Alvarez, G. A. (2009). Detecting changes in real-world objects The relationship between visual long-term memory and change blindness. *Communicative & Integrative Biology, 2*(1), 1–3.

Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of vision,* 11(5), 4-4.

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.

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

Brewer, J. B., Zhao, Z., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. (1998). Making memories: brain activity that predicts how well visual experience will be remembered. *Science,* 281(5380), 1185-1187.

Broadbent, D. E. (1957). A mechanical model for human attention and immediate memory. *Psychological review*, *64*(3), 205.

Burke, L. (1952). On the tunnel effect. *Quarterly Journal of Experimental Psychology*, *4*(3), 121-138.

Cansino, S., Maquet, P., Dolan, R. J., & Rugg, M. D. (2002). Brain activity underlying encoding and retrieval of source memory. *Cerebral Cortex, 12*(10), 1048-1056.

Cheries, E. W., Mitroff, S. R., Wynn, K., & Scholl, B. J. (2008). Cohesion as a constraint on object persistence in infancy. *Developmental science*, *11*(3), 427-432.

Chun, M. M., & Cavanagh, P. (1997). Seeing two as one: Linking apparent motion and repetition blindness. *Psychological Science, 8*(2), 74-79.

Chun, M. M., & Turk-Browne, N. B. (2007). Interactions between attention and memory. *Current Opinion in Neurobiology*, *17*(2), 177-184.

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

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in brain research, 169*, 323–338.

Cox, D. D., Meier, P., Oertelt, N., & DiCarli, J. J. (2005). 'Breaking' position-invariant object recognition. *Nature Neuroscience, 8*, 1145-1147.

Cox, D. D., & DiCarlo, J. J. (2008). Does learned shape selectivity in inferior temporal cortex automatically generalize across retinal position?. *The Journal of Neuroscience, 28*(40), 10045-10055.

Cunningham, C. A., Yassa, M. A., & Egeth, H. E. (2015). Massive memory revisited: Limitations on storage capacity for object details in visual long-term memory. *Learning & Memory, 22*(11), 563-566.

Curby, K. M., & Gauthier, I. (2007). A visual short-term memory advantage for faces. *Psychonomic bulletin & review, 14*(4), 620-628.

Curby, K. M., Galzek, K., & Gauthier, I. (2009). A visual short-term memory advantage for objects of expertise. *Journal of Experimental Psychology: Human Perception and Performance, 35*(1), 94.

Dede, A. J., Squire, L. R., & Wixted, J. T. (2014). A novel approach to an old problem: Analysis of systematic errors in two models of recognition memory. *Neuropsychologia*, *52*, 51-56.

Dennett, D. C. (2006). The Frame Problem of AI. Philosophy of Psychology: *Contemporary Readings*, 433.

Diana, R. A., Yonelinas, A. P., & Ranganath, C. (2007). Imaging recollection and familiarity in the medial temporal lobe: a three-component model. *Trends in cognitive sciences*, *11*(9), 379-386.

DiCarlo, J. J., & Cox, D. D. (2007). Untangling invariant object recognition. *Trends in cognitive sciences, 11*(8), 333-341.

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition?. *Neuron, 73*(3), 415-434.

Egan, J. P. (1958). Recognition memory and the operating characteristic. *USAF Operational Applications Laboratory Technical Note*.

Feldman, J. (2003). What is a visual object?. *Trends in Cognitive Sciences*, *7*(6), 252-256.

Flombaum, J. I., Kundey, S. M., Santos, L. R., & Scholl, B. J. (2004). Dynamic object individuation in rhesus macaques a study of the tunnel effect. *Psychological Science*, *15*(12), 795-800.

Flombaum, J. I., & Scholl, B. J. (2006). A temporal same-object advantage in the tunnel effect: facilitated change detection for persisting objects. *Journal of Experimental Psychology: Human Perception and Performance, 32*(4), 840.

Flombaum, J. I., Scholl, B. J., & Santos, L. R. (2009). Spatiotemporal priority as a fundamental principle of object persistence. *The origins of object knowledge*, 135-164.

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. MIT press.

Fortin, N. J., Wright, S. P., & Eichenbaum, H. (2004). Recollection-like memory retrieval in rats is dependent on the hippocampus. *Nature*, *431*(7005), 188-191.

Gabrieli, J. D., Fleischman, D. A., Keane, M. M., Reminger, S. L., & Morrell, F. (1995). Double dissociation between memory systems underlying explicit and implicit memory in the human brain. *Psychological Science*, *6*(2), 76-82.

Geusebroek, J. M., Burghouts, G. J., & Smeulders, A. W. (2005). The Amsterdam library of object images. *International Journal of Computer Vision*, *61*(1), 103-112.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics* (Vol. 1). New York: Wiley.

Guerin, S. A, Robbins, C. A, Gilmore, A. W., & Schacter, D. L. (2012). Retrieval Failure Contributes to Gist-Based False Recognition. *Journal of Memory and Language, 66*(1), 68–78.

Guerin, S. A., Robbins, C. A., Gilmore, A. W. & Schacter, D. L. (2012). Interactions between visual attention and episodic retrieval: dissociable contributions of parietal regions during gist-based false recognition. *Neuron*, *75*(6), 1122-1134.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *Proceedings of the IEEE international conference on computer vision*, 1026-1034.

Hintzman, D. L. (1976). Repetition and memory. *Psychology of learning and motivation, 10*, 47-91.

Holdstock, J. S., Mayes, A. R., Roberts, N., Cezayirli, E., Isaac, C. L., O'reilly, R. C., & Norman, K. A. (2002). Under what conditions is recognition spared relative to recall after selective hippocampal damage in humans? *Hippocampus*, *12*(3), 341-351.

Hollingworth, A., & Franconeri, S. L. (2009). Object correspondence across brief occlusion is established on the basis of both spatiotemporal and surface feature cues. Cognition, 113, 150166.

Isik, L., Leibo, J. Z., & Poggio, T. (2012). Learning and disrupting invariance in visual recognition with a temporal association rule. *Frontiers in computational neuroscience*, 6.

Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology Learning Memory and Cognition*, *26*(3), 683-702.

Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., & Moore, K. S. (2008). The mind and brain of short-term memory. *Annual Review of Psychology*, *59*, 193–224.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive psychology*, 24(2), 175-219.

Kawachi, Y., & Gyoba, J. (2006). A new response-time measure of object persistence in the tunnel effect. *Acta Psychologica*, *123*(1), 73-90.

Kensinger, E. A., Garoff-Eaton, R. J., & Schacter, D. L. (2006). Memory for specific visual details can be enhanced by negative arousing content. *Journal of Memory and Language*, *54*(1), 99-112.

Keppel, G., & Underwood, B. J. (1962). Proactive inhibition in short-term retention of single items. *Journal of verbal learning and verbal behavior*, *1*(3), 153-161.

Kim, J., & Yassa, M. A. (2013). Assessing recollection and familiarity of similar lures in a behavioral pattern separation task. *Hippocampus, 23*(4), 287-294.

Kirchhoff, B. A., Wagner, A. D., Maril, A., & Stern, C. E. (2000). Prefrontal–temporal circuitry for episodic encoding and subsequent memory. *The Journal of Neuroscience, 20*(16), 6173-6180.

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *Journal of Experimental Psychology: General*, *139*(3), 558.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 1097-1105.

Li, N., & DiCarlo, J. J. (2008). Unsupervised natural experience rapidly alters invariant object representation in visual cortex. *Science*, *321*(5895), 1502-1507.

Logothetis, N. K., & Sheinberg, D. L. (1996). Visual object recognition. *Annual review of neuroscience*, *19*(1), 577-621.

Loiotile, R. E., & Courtney, S. M. (2015). A signal detection theory analysis of behavioral pattern separation paradigms. *Learning & Memory*, *22*(8), 364-369.

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

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature neuroscience, 17*(3), 347-356.

Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior*, *9*(5), 596-606.

Mitroff, S. R., & Alvarez, G. A. (2007). Space and time, not surface features, guide object persistence. *Psychonomic Bulletin and Review, 14*, 1199-1204.

Miyamoto, K., Adachi, Y., Osada, T., Watanabe, T., Kimura, H. M., Setsuie, R., & Miyashita, Y. (2014). Dissociable Memory Traces within the Macaque Medial Temporal Lobe Predict Subsequent Recognition Performance. *The Journal of Neuroscience*, *34*(5), 1988-1997.

Moore, C. M., Stephens, T., & Hein, E. (2010). Features, as well as space and time, guide object persistence. *Psychonomic Bulletin and Review, 17*, 731-736.

Neunuebel, J. P., & Knierim, J. J. (2014). CA3 Retrieves Coherent Representations from Degraded Input: Direct Evidence for CA3 Pattern Completion and Dentate Gyrus Pattern Separation. *Neuron*, *81*(2), 416-427.

Newman, G. E., Keil, F. C., Kuhlmeier, V. A., & Wynn, K. (2010). Early understandings of the link between agents and order. *Proceedings of the National Academy of Sciences*, *107*(40), 17140-17145.

Noles, N. S., Scholl, B. J., & Mitroff, S. R. (2005). The persistence of object file representations. *Perception & Psychophysics*, *67*(2), 324-334.

Olson, I. R., & Jiang, Y. (2002). Is visual short-term memory object based? Rejection of the "strong-object" hypothesis. *Perception & psychophysics*,*64*(7), 1055-1067.

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

Pinto, N., Cox, D. D., & DiCarlo, J. J. (2008). Why is real-world visual object recognition hard? *PLoS Comput Biol*, *4*(1), e27.

Poth, C. H., Herwig, A., & Schneider, W. X. (2015). Breaking object correspondence across saccadic eye movements deteriorates object recognition. *Frontiers in systems neuroscience*, *9*, 176.

Poth, C. H., & Schneider, W. X. (2016). Breaking object correspondence across saccades impairs object recognition: The role of color and luminance. .*Journal of Vision*, *16*(11), 1-1.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of experimental psychology: human learning and memory, 2*(5), 509-522.

Ranganath, C., Cohen, M. X., & Brozinsky, C. J. (2005). Working memory maintenance contributes to long-term memory formation: neural and behavioral evidence. *Journal of Cognitive Neuroscience*, *17*(7), 994-1010.

Ratcliff, R., Sheu, C. F., & Gronlund, S. D. (1992). Testing global memory models using ROC curves. *Psychological review*, *99*(3), 518-535.

Reagh, Z. M., & Yassa, M. A. (2014a). Object and spatial mnemonic interference differentially engage lateral and medial entorhinal cortex in humans. *Proceedings of the National Academy of Sciences, 111*(40), E4264-E4273.

Reagh, Z. M. & Yassa, M. A. (2014b). Repetition strengthens target recognition but impairs similar lure discrimination: evidence for trace competition. *Learning & Memory, 21*(7), 342-346.

Rentz, D. M., Parra Rodriguez, M. A., Amariglio, R., Stern, Y., Sperling, R., & Ferris, S. (2013). Promising developments in neuropsychological approaches for the detection of preclinical Alzheimer's disease: a selective review. *Alzheimer's research & therapy*, *5*(6), 58.

Richard, A. M., Luck, S. J., & Hollingworth, A. (2008). Establishing object correspondence across eye movements: Flexible use of spatiotemporal and surface feature information. *Cognition, 109*, 6688.

Rust, N. C., & Stocker, A. A. (2010). Ambiguity and invariance: two fundamental challenges for visual processing. *Current opinion in neurobiology, 20*(3), 382-388.

Scholl, B. J. (2001). Objects and attention: The state of the art. *Cognition*, *80*(1), 1-46.

Scholl, B. J., & Flombaum, J. I. (2010). Object persistence. In B. Goldstein (Ed.), *Encyclopedia of Perception, Volume 2* (pp. 653-657). Thousand Oaks, CA: Sage Publications.

Scholl, B. J., & Pylyshyn, Z. W. (1999). Tracking multiple items through occlusion: Clues to visual objecthood. *Cognitive Psychology*, *38*(2), 259-290.

Schurgin, M. W., & Flombaum, J. I. (2015). Visual long-term memory has weaker fidelity than working memory. *Visual Cognition*, *23*(7), 859-862.

Schurgin, M. W. & Flombaum, J. I. (2017). Exploiting Core Knowledge for Visual Object Recognition. *Journal of Experimental Psychology: General.*

Schurgin, M. W., Reagh, Z. M., Yassa, M. A., & Flombaum, J. I. (2013). Spatiotemporal continuity alters long-term memory representation of objects. *Visual Cognition*, *21*(6), 715-718.

Shallice, T., & Warrington, E. K. (1970). Independent functioning of verbal memory stores: A neuropsychological study. *The Quarterly journal of experimental psychology*, *22*(2), 261-273.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. J*ournal of verbal Learning and verbal Behavior, 6*(1), 156-163.

Shohamy, D., & Turk-Browne, N. B. (2013). Mechanisms for widespread hippocampal involvement in cognition. *Journal of Experimental Psychology: General*, *142*(4), 1159.

Shotton, J., Winn, J., Rother, C., & Criminisi, A. (2006, May). Textonboost: Joint appearance, shape and context modeling for multi-class object recognition and segmentation. In *European conference on computer vision* (pp. 1-15). Springer Berlin Heidelberg.

Singh, I., Oliva, A., & Howard, M. (2017). Visual memories are stored along a compressed timeline. *bioRxiv*, 101295.

Smolen, P., Zhang, Y. & Byrne, J. H. (2016). The right time to learn: mechanisms and optimization of spaced learning. *Nature Reviews Neuroscience, 17*, 77-88.

Spelke, E. S. (1990). Principles of object perception. *Cognitive science, 14*(1), 29-56.

Spelke, E. S., Kestenbaum, R., Simons, D. J., & Wein, D. (1995). Spatiotemporal continuity, smoothness of motion and object identity in infancy. *British Journal of Developmental Psychology, 13*(2), 113-142.

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental science*, *10*(1), 89-96.

Sperling, G. (1963). A Model for Visual Memory Tasks 1. *Human factors*, *5*(1), 19-31.

Squire, L. R. (2004). Memory systems of the brain: a brief history and current perspective. *Neurobiology of learning and memory, 82*(3), 171-177.

Stahl, A. E., & Feigenson, L. (2015). Observing the unexpected enhances infants' learning and exploration. *Science, 348*(6230), 91-94.

Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science,19*(2), 73-74.

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

Strickland, B., & Scholl, B. J. (2015). Visual perception involves event-type representations: The case of containment versus occlusion. *Journal of Experimental Psychology: General*, *144*(3), 570.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*(6582), 520-522.

Van Marle, K., & Scholl, B. J. (2003). Attentive tracking of objects versus substances. *Psychological Science*, *14*(5), 498-504.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(1), 92.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 32*(6), 1436–51.

Wallis, G., & Bülthoff, H. H. (2001). Effects of temporal association on recognition memory. *Proceedings of the National Academy of Sciences, 98*(8), 4800-4804.

Wallis, G. (2002). The role of object motion in forging long-term representations of objects. *Visual Cognition, 9*(1-2), 233-247.

Warrington, E. K., & Shallice, T. (1969). The selective impairment of auditory verbal short-term memory. *Brain, 92*(4), 885-896.

Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General, 131*(1), 48-64.

Wickens, T. D. (2002). *Elementary Signal Detection Theory*. Oxford, UK: Oxford University Press.

Wilcox, T., & Baillargeon, R. (1998a). Object individuation in infancy: The use of featural information in reasoning about occlusion events. *Cognitive Psychology, 37*, 97-155.

Wilcox, T., & Baillargeon, R. (1998b). Object individuation in young infants: Further evidence with an event monitoring task. *Developmental Science, 1*, 127-142.

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

Wimmer, G. E., & Shohamy, D. (2012). Preference by association: how memory mechanisms in the hippocampus bias decisions. *Science*, *338*(6104), 270-273.

Wittmann, B. C., Schott, B. H., Guderian, S., Frey, J. U., Heinze, H. J., & Düzel, E. (2005). Reward-related FMRI activation of dopaminergic midbrain is associated with enhanced hippocampus-dependent long-term memory formation. *Neuron, 45*(3), 459-467.

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological review*, *114*(1), 152.

Xu, Y. (2002). Limitations of object-based feature encoding in visual short-term memory. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(2), 458.

Xu, F., & Carey, S. (1996). Infants' metaphysics: The case of numerical identity. *Cognitive psychology, 30*(2), 111-153.

Xu, F., Carey, S., & Quint, N. (2004). The emergence of kind-based object individuation in infancy. *Cognitive Psychology, 49*(2), 155-190.

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

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of memory and language*, *46*(3), 441-517.

Yonelinas, A. P., Aly, M., Wang, W. C., & Koen, J. D. (2010). Recollection and familiarity: Examining controversial assumptions and new directions. *Hippocampus*, *20*(11), 1178-1194.

Yi, D. J., Turk-Browne, N. B., Flombaum, J. I., Kim, M. S., Scholl, B. J., & Chun, M. M. (2008). Spatiotemporal object continuity in human ventral visual cortex. *Proceedings of the National Academy of Sciences, 105*(26), 8840-8845.

Zhang, J., Marszałek, M., Lazebnik, S., & Schmid, C. (2007). Local features and kernels for classification of texture and object categories: A comprehensive study. *International journal of computer vision*, *73*(2), 213-238.

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

Mark Schurgin

maschurgin@jhu.edu

Johns Hopkins University
3400 N. Charles Street
Baltimore, MD 21218

EDUCATION

Johns Hopkins University, Baltimore, MD Ph.D. 2017
Department of Psychological and Brain Sciences
Concentration: Cognitive Psychology

Johns Hopkins University, Baltimore, MD M.A. 2014
Department of Psychological and Brain Sciences
Concentration: Cognitive Psychology

Vassar College, Poughkeepsie, NY B.A. 2010
Major: Psychology
Minor: Political Theory
Research Abroad: Qingdao University Summer Program in China 2007

RESEARCH EXPERIENCE

Lab Manager 2010 - 2012
Advisor: Steven Franconeri (Visual Attention & Cognition Lab, Northwestern University)
Advisor: Joan Chiao (Social Affective & Cultural Neuroscience Lab, Northwestern University)

- Coordinated and managed the research activities of over 20 undergraduate research assistants
- Operated SIEMENS 3T MRI scanner, including implementation of physiological equipment
- Designed and implemented ERP and eye-tracking experiments using Bio-Semi EEG equipment and SR-Research Eyelink II eye-tracking equipment
- Designed and implemented experiments using Psychtoolbox, Experiment Builder, & DirectRT
- Analyzed neuroimaging data using SPM

Research Assistant
Advisor: Steven Franconeri (Visual Attention & Cognition Lab, Northwestern University)

- Full time employee Summer 2009
- Head researcher for project analyzing cross-cultural eye-tracking data
- Performed statistical analysis of eye-tracking data using Excel and SPSS Summer 2008
- Used MATLAB to create research images and videos

RESEARCH INTERESTS

General: Perception, Memory, Cognition

Specific: Visual working memory; long-term memory; object recognition; spatiotemporal continuity; spatial relations and distortions; perception and action; visual strategies and human performance

PRESENTATIONS AND PUBLICATIONS

Publications

- **Schurgin, M. W.** & Flombaum, J. I. (2017). Exploiting Core Knowledge for Visual Object Recognition. *Journal of Experimental Psychology: General, 146*(3), 362-375.
- **Schurgin, M. W.** & Flombaum, J. I. (2015). Visual long-term memory has weaker fidelity than working memory. *Visual Cognition*, *23*(7), 859-862.
- **Schurgin, M. W.** Nelson, J., Iida, S., Ohira, H., Chiao, J. Y., & Franconeri, S. L. (2014). Eye movements during emotional recognition in faces. *Journal of Vision, 14*(13):14, 1-16.
- **Schurgin, M. W.** & Flombaum, J. I. (2014). How undistorted spatial memories can produce distorted responses. *Attention, Perception & Psychophysics*, *76*(5), 1371-1380.
- **Schurgin, M. W.,** Reagh, Z. M., Yassa, M. A. & Flombaum, J. I. (2013). Spatiotemporal Continuity Alters Long-Term Memory Representation of Objects. *Visual Cognition, 21*(6), 715-718.

Manuscripts Under Review

- Petre, B., Tetreault, P., Mathur, V. A., **Schurgin, M. W.,** Chiao, J. Y., Huang, L., and Apkarian, A. V. (under review). A central recurrent mechanism underlies nonlinear edge detection in acute pain.

Manuscripts In Preparation

- **Schurgin, M. W.** (in prep). What is Visual Memory?
- **Schurgin, M. W.** & Flombaum, J. I. (in prep). Visual Working Memory Tolerance Supports Future Long-Term Memory Discrimination.
- **Schurgin, M. W.,** Cunningham, C.A. & Flombaum, J. I. (in prep). Visual Search Abilities Dynamically Manage Performance Under Noisy Conditions.
- Mathur, V. A., **Schurgin, M. W.**, Saeed, S. W., Reber, P. J., Apkarian, A. V., Paice, J. A., Richeson, J., Chiao, J. Y. (in prep). Race modulates neural sensitivity to own and other's pain.

Conference Presentations

- Flombaum, J. I. & **Schurgin, M. W.** (2016). Exploiting Core Knowledge for Visual Object Recognition. Talk presented at the 57th annual meeting of the *Psychonomic Society*, Boston, MA.
- **Schurgin, M. W.** & Flombaum, J. I. (2016). First Impressions in Visual Long-Term Memory. Poster presented at the 24th annual meeting on *Object Perception, Attention, and Memory (OPA*M), Boston, MA.
- **Schurgin, M. W.** & Flombaum, J. I. (2016). Visual Working Memory Has Greater Tolerance Than Visual Long-Term Memory. Poster presented at the 16th annual meeting of the *Vision Sciences Society*, St. Pete Beach, FL.
- **Schurgin, M. W.** & Flombaum, J. I. (2015). Visual Long-Term Memory Has Weaker Fidelity Than Working Memory. Talk presented at the 23rd annual meeting on *Object Perception, Attention, and Memory (OPA*M), Chicago, IL.
- **Schurgin, M. W.** & Flombaum, J. I. (2015). Invariant Object Recognition Enhanced by Object Persistence. Poster presented at the 15th annual meeting of the *Vision Sciences Society*, St. Pete Beach, FL.
- **Schurgin, M. W.,** Reagh, Z. M., Yassa, M. A. & Flombaum, J. I. (2014). Building Tolerant Long-Term Memories Through (Object) Persistence. Poster presented at the 14th annual meeting of the *Vision Sciences Society*, St. Pete Beach, FL.

- **Schurgin, M. W.,** Reagh, Z. M., Yassa, M. A. & Flombaum, J. I. (2013). Spatiotemporal Continuity Alters Long-Term Memory Representation of Objects. Talk presented at the 21st annual meeting on *Object Perception, Attention, and Memory* (OPAM), Toronto, Canada.
- **Schurgin, M. W**. & Flombaum, J. I. (2013). Interactions between perception, fixation, and attention determine the endpoint of an action. Poster presented at the 13th annual meeting of the *Vision Sciences Society*, Naples, FL.
- **Schurgin, M. W.,** Levinthal, B. R., List, A., Sherman, A., Suzuki, S., Grabowecky, M. and Franconeri, S. L. (May 2011). Infinite X: Illusions of perpetual increases in magnitude. Demonstration presented at the 11th Annual Meeting of the *Vision Sciences Society*, Naples, FL.

HONORS & FUNDING

- Dean's Teaching Fellowship, Johns Hopkins University 2016-2017
- Walter L. Clark Service Award, Johns Hopkins University 2016
- The Johns Hopkins University Graduate Representative Organization (GRO) Travel Award 2016
- Robert S. Waldrop & Dorothy L. Waldrop Graduate Award, Johns Hopkins University 2013-2016
- Best paper award at the 21st annual meeting on Object Perception, Attention, and Memory (OPAM), Toronto, CA 2013

TEACHING EXPERIENCE

Instructor

- The Illusion of Perception (Undergraduate Course), Johns Hopkins University Fall 2016
- Research Methods & Design (Undergraduate Course), Johns Hopkins University Fall 2014

Teaching Assistant

- Advanced Statistical Methods (Graduate Course), Johns Hopkins University Fall 2015
- Intro to Cognitive Psychology (Undergraduate Course), Johns Hopkins University Spring 2014
- Mind, Brain & Experience (Undergraduate Course), Johns Hopkins University Fall 2013
- Human Sexual Orientation (Undergraduate Course), Johns Hopkins University Spring 2013

COMMUNITY ENGAGEMENT

- Creator & Director, Psychological & Brain Sciences High School Engagement Program 2016-present
- Director & Speaker, Brain Awareness Week at Baltimore Polytechnic Institute High School 2013-present
- Graduate Mentor, Women in Science & Engineering (WISE) student outreach program with Garrison Forest School 2015-2016
- Invited Speaker, Patterson High School Baltimore City 2015
- Organizer, Chicago Brain Week outreach talk series 2012

ACADEMIC SERVICE

- Organizer, Psychological & Brain Sciences Event Coordination 2014-present
- Representative, Psychological & Brain Sciences Graduate Steering Committee 2012-2016
- Organizer & Leader, Psychological & Brain Sciences Graduate Recruitment 2013-2015
- Reviewer, Association for Psychological Science Student Research Award 2015
- Organizer, Cultural Psychology Preconference at the Society for Personality & Social Psychology (SPSP) 2012
- Organizer, Social & Affective Neuroscience Society (SANS) Conference 2010

SKILLS

- Programming/Scripting Languages: MATLAB; Python; HTML
- Experiment Presentation Software: Experiment Builder; Psychtoolbox; DirectRT; PsychoPy
- MRI Analysis Packages: SPM2; SPM5; SPM8; Freesurfer
- Data Analysis Software: SPSS; R (rstan)

ADVISING

Undergraduate Researchers

- Faith Shank (Loyola University) 2016-present
- Annapurna Vadaparty (Johns Hopkins University) 2015-present
- Sinan Akosman (Johns Hopkins University) 2015-present
- Cera Hassinan (Johns Hopkins University) 2015-present
- Alana DiSabatino (Johns Hopkins University) 2015-present
- Saman Baban (Johns Hopkins University) 2015-2016
- Victor Kang (Johns Hopkins University) 2015
- Patricia Kingkeo (Johns Hopkins University) 2013-2015
- Erica Lee (Johns Hopkins University) 2013-2014
- Hannah Cowley (Johns Hopkins University) 2014
- Gustavo Beruman (Universidad de Guadalajara) 2013
- Kaan Zaimoglu (Johns Hopkins University) 2012-2013

High School Researchers

- Channing Capacchione (GFS / Hopkins) 2015-2016
- Ruth Tekeste (GFS / Hopkins) 2015-2016
- Jennifer Ren (IMSA / Northwestern) 2011-2012
- Victoria Etherton (IMSA / Northwestern) 2011-2012
- Eva Meyer (IMSA / Northwestern) 2011-2012
- Ruby Morales (ETHS / Northwestern) Summer 2011

PROFESSIONAL MEMBERSHIPS

- Psychonomic Society 2015-present
- Spatial Network 2014-present
- Vision Sciences Society (VSS) 2011-present
- Spatial Intelligence Learning Center (SILC) 2010-2012
- Cognitive Neuroscience Society (CNS) 2011-2012