INVESTIGATING THE ROLE OF PREDICTIVE REPRESENTATIONS IN IMPLICIT EVENT BOUNDARIES, STATISTICAL LEARNING, AND CATEGORIZATION.

A Dissertation Presented

by

TEJAS SAVALIA

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2024

Psychological and Brain Sciences

INVESTIGATING THE ROLE OF PREDICTIVE REPRESENTATIONS IN IMPLICIT EVENT BOUNDARIES, STATISTICAL LEARNING, AND CATEGORIZATION.

A Dissertation Presented

by

TEJAS SAVALIA

Approved as to style and content by:

Andrew Cohen, Chair

Jeffrey Starns, Member

Youngbin Kwak, Member

Meghan Huber, Member

Maureen Perry-Jenkins, Chair
Psychological and Brain Sciences

ACKNOWLEDGMENTS

Work done during my PhD would not have been possible without the contribution of many people. This dissertation is indeed a baby which required a village to take care of.

I would first like to thank my advisors who saw this dissertation through. None of the work presented here would have been possible without Andrew Cohen and Jeffrey Starns who over countless meetings have helped me hone my projects, and ensured scientific rigor. They also let me work on whatever I wanted, provided the resources I needed, and generally made research fun. I would also like to thank Youngbin Kwak and Meghan Huber who served on my dissertation committee – first for continuing on it as I changed projects and for providing invaluable feedback that significantly improved my dissertation. My first advisors in grad school at UMass; David Huber and Rosie Cowell taught me the fundamental steps of research in my early years as a PhD student really honed my ideas and allowed me to grow as a researcher. Other cognitive faculty at UMass Amherst provided invaluable experience in academic presentation through brown-bag questions and through courses I took over the years.

I would next like to thank my ~~collegues~~ friends in the cognitive program throughout the years: Sean, Anna, Melisa, Mar, John, Jerome, Michael, Yun, Chiungyu, Kuan-Jung, and Sandarsh for being a stable presence in this intellectual journey. You made this program feel like home. I would also like to call out Natasha and Kuan Jung, my co-hort mates who went through the roller coaster with me.

Weekly board games in PBS is one of the most consistent social activity I developed is at UMass and the group of friends have practically served as my family

away from home. Anna, Clara, Tori, Sean, Trina, Fran, Ramiro, and Kuan Jung; your dedication to put up with my, ahem, eccentric, hobby has never failed to lift my spirits and always provided with an outlet I look forward to every week. The last two years of my grad school would not have been as much fun if not for you and I am glad that this tradition is set to continue going forward. Finally, I'd be remiss if I did not mention the 'tiny living room' folks; Ramiro, Sandarsh, and Hyejoo for being my first non-work friends at work.

Several people helped me through grad school before and through the pandemic – Aarohi, Meet, Rik, Sohini, Pracheta, and Princy, I will never forget the absolute random conversations, and trips that provided a venue to turn my mind off work and let me stay sane. Kunal, while you were not at UMass, you have never let go of this friendship with regular check ins, and truly showed me the value of real friendship. I don't know how you do it, but please continue doing it.

Manasa – I can write another dissertation to describe your tireless support. Instead, I will just say that absolutely no part of the dissertation would have been possible without you. You helped me in literally everything I wanted your help. From non-work ups and downs, to serving as an initial ideas sounding board, to helping me with analyses and programming, to feedback on my writing, and through celebrating me, you have seen it all, done it all and been my consistent cheerleader in these six years. I could not have chosen a better human being to do so. Thank you.

Finally, I would not have been here if not for continued, unwavering support from my family. Mummy, pappa, diku, Nishita, and Aaru – thank you for letting me do something regardless of whether it is clear what it is I exactly do. Your faith in me has been the key to make this work happen.

ABSTRACT

INVESTIGATING THE ROLE OF PREDICTIVE REPRESENTATIONS IN IMPLICIT EVENT BOUNDARIES, STATISTICAL LEARNING, AND CATEGORIZATION.

SEPTEMBER 2024

TEJAS SAVALIA

B.E., GUJARAT TECHNOLOGICAL UNIVERSITY

M.S., INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY, HYDERABAD

M.S., UNIVERSITY OF MASSACHUSETTS AMHERST

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Andrew Cohen

We make sense of the world by extracting meaningful information from a continuous sensory stream. Extracting meaningful information involves first segmenting this continuous sensory stream into shorter, processable chunks. These discrete chunks of events represent our recalled experiences and allow us to develop heuristics representing the statistical regularities in our environment.

In this dissertation, I present a predictive context representational account of segmenting the continuous sensory stream into smaller chunks. I demonstrate that maintaining a distributed context representation defined by an expectation of upcoming future events and learned through temporal difference learning naturally leads to

the separation of temporally disjoint events without perceptually explicit markers. I contrast this predictive, error-driven account of context representation with an associative learning account and provide behavioral evidence in support of the predictive representational account.

I then show that such predictive context representations can be used as a common framework to understand higher order cognitive processes of event cognition and categorization. I first assess whether implicitly operationalized event boundaries, where changes in ongoing context that mark boundaries are not perceptually salient, provide the same behavioral properties as explicitly operationalized event boundaries thereby providing evidence for shared representations between the two. Finally, I apply the representational framework to understand the cognitive processes behind implicit category learning. I show that predictive representations can arbitrate category learning via the shared temporal context for items in each category.

Work in this dissertation provides a mechanistic account for statistical learning through widely applicable framework of temporal difference learning. I further demonstrate a use of predictive representations as a common framework to understand higher-order cognitive processes such as event cognition, and categorization.

TABLE OF CONTENTS

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

CHAPTER 1

INTRODUCTION

We experience a constant stream of sensory information from the moment we are born. Our brain parses this information and slowly learns to extract meaning from it. From recognizing the mother's scent as a survival instinct to formulating complex plans to defeat a board game opponent, our brain extracts meaning from our surroundings, considers prior experience and the current state of the world, and makes decisions to interact accordingly. My key question in this dissertation is: How do we learn to extract meaning from our surroundings?

Extracting meaningful information from our experience is beneficial to our functioning and survival. When approached by a cheetah in a forest, we do not wait to evaluate the exact number of spots on its skin before deciding to run. Such abstractions and formation of heuristics allow us to process naturally complex surroundings in quick time and act accordingly.

When extracting meaning from our surroundings, we often (need to) ignore minute details and abstract out towards a coherent thought of our surroundings. The extreme example above aside, we observe such abstractions in almost all day-to-day activities. Imagine someone asking you about the events during the day before reading this manuscript. Perhaps you are reading it on your office computer and you recall a sequence of events starting from waking up, to getting ready, driving or taking public transit to work, getting your morning coffee and breakfast in with an email check before turning your attention to this manuscript. Each event described above combines several sub-events that are abstracted away in such a verbal recall and de-

scription. For example, getting ready involves several steps from brushing your teeth, showering, and wearing your work outfits. Each sub-event can be further thought of as an abstraction from sub-sub-events – brushing your teeth is a combination of putting toothpaste on the brush head, the physical act of brushing, followed by rinsing. While we perform each act continuously in time, our recall (and by extension, representation in memory) of these past events is discrete, segmented, and abstracted. The meaningfully separable chunks of a continuous stream of events help in storage in (and retrieval from) memory.

There is often agreement on what it means for a chunk and for boundaries defining such chunks to be 'meaningful'. In the example above, it is reasonable to argue that brushing teeth, showering, and putting on clothes are three distinct activities. Furthermore, even when the true transitions between these activities are continuous and seamless to a independent, naive observer, the boundaries between these events are perceptually meaningful. What aspect of the environment dictates this agreement about the points at which we segment events and what is special about the properties of the events between those points that lead to distinct representations in memory?

One could argue that these events that occur at different points in time also occur at different spatial locations, thus providing different contexts and hence separate representations in our brain to be considered distinct. Transitioning from one (temporal) event to another can thus be akin to transitioning from one room of the house to another. However, it is almost impossible to decouple temporal events from spatial events assuming a causality from spatial segmentation to temporal segmentation for any spatially experienced distinct event is also a temporally experienced distinct event. Instead, arguing that temporally distinct events lead to a spatially distinct representation can provide a more encompassing explanation of distinct representations for events in distinct spatial *and* temporal contexts. One could similarly argue that the formation of "meaningful" chunks is through perceptual differences between

the events we experience. While lower-level perceptual experiences are indeed often different for different events, the mapping of perceptual differences onto segmented events is arbitrary and not a *sufficient* condition. For example, eating an apple is recalled as eating an apple regardless of whether it has a green leaf added to its top. Perceptual distinctions are not enough to determine whether events are represented distinctly. What then is the key mechanism that leads to events being segmented?

In this dissertation, I argue that the primary reason and mechanism through which we segment events is based on temporal contingencies of various sub-events that encompass an event. Specifically, we recall being in the kitchen as different from being in the bedroom because we have a coherent set of experiences in the kitchen that are distinct from a coherent set of experiences in the bedroom. For an infant forming knowledge of the world, a kitchen while perceptually distinct from a bedroom, is not meaningfully different. With experience and observation of the functions within these spatially (and perceptually) distinct locations, the child slowly develops distinct representations of the two rooms.

This dissertation focuses on understanding temporal contingencies' role in event segmentation, and by extension, general pattern extraction. I argue that even *without* any spatial or perceptual information that may aid us in separating events in memory, we can use temporal coherence to experience separate events and abstract information to aid higher-level cognition. Specifically, I investigate the parsing of a continuous stream of information into discrete chunks in three ways:

- The possible algorithmic representations that naturally lead to such segmentation and the impact of environmental properties in aiding this abstraction.
- The properties of the temporal boundaries when such temporal abstraction occurs naturally and implicitly.

- The role of temporal events separated by underlying transition structures in forming higher-order abstractions such as categories.

1.1 Scope of this dissertation

The human brain is a complex machine – millions of neurons act as computational units and combine in specific ways to form a functioning human being. These neurons come together to implement several levels of function from lower-level automatic perception to higher-level planning and conscious thought. This dissertation does **not** focus on these implementational-level mechanisms of cognition. Rather, it focuses on *algorithmic* computations that neurons may, collectively, implement that lead to us acquiring patterns in the environment around us (Marr & Poggio, 1976). These algorithmic computations are then experimentally tested through behavioral data and computational modeling.

Analyses in this dissertation use several models of cognition in investigating the role of implicit statistical learning. In most cases, the focus of this dissertation is **not** to evaluate the validity of these models. Indeed, most models of cognition are wrong but are useful (Fisher et al., 2019) and I use several such models to evaluate specific aspects of how we acquire patterns. Similarly, this dissertation proposes modifications to the previously known models based on context representations. These modifications are solely to derive predictions and provide possible explanations of findings from these models and are not rigorously tested. Future work should aim to test the updates proposed in this work and therefore provide more holistic explanation of the cognitive processes explored in this manuscript. Nevertheless, models and proposed modifications to these models used in this dissertation play a key role in generating experimentally testable predictions presented in this dissertation which are then used to motivate experimental designs.

Finally, the data collected and used in this dissertation is much more rich than presented. In order to limit the scope to the specific questions of interest, only a subset of analyses involving simpler (often linear) models is presented. Future work will incorporate more complex models on this data to extract fine grained information of the representation driven cognitive processes.

The key contribution of this dissertation is in presenting a representational framework to understand the cognitive processes of pattern recognition and statistical learning. Furthermore, this dissertation provides for a common representational framework, with prior evidence for this representational framework being implemented in the brain (Gershman, 2018) to understand a myriad of cognitive processes at different levels of cognition. In most prior work, cognitive processes of event cognition, learning, memory, and categorization in prior work have largely been studied separately within those sub-fields. However, this dissertation shows that one common representational framework of learning can be used to explain these processes by varying operations on that representation (Cowell et al., 2019).

1.2 Format of this dissertation

In the rest of this dissertation, I present three lines of studies investigating the role of implicit temporal boundaries in cognition. In Chapter 2, I present an algorithmic representational framework that naturally leads to a representation of separable events without the need to rely on explicit properties of the experienced events. I also show that this predictive framework allows for a distinct representation of event boundaries as special events and contrast it with an associative representation. In Chapter 3, I present work comparing the properties of these event boundaries which are operationalized implicitly (i.e. through no perceptually special information) with event boundaries as they have been studied in prior literature which are operationalized explicitly. In Chapter 4, I present work investigating the role of these implicitly

operationalized event boundaries in categorization to serve as a gateway for understanding higher-order cognition in the context of temporal segmentation and pattern acquisition. Finally, in Chapter 5, I present a summary of findings presented in this dissertation and provide an overview of broader implications of this work.

CHAPTER 2

QUALITY OF ENVIRONMENTAL EXPOSURE MODULATES STATISTICAL LEARNING

2.1 Introduction

Imagine you just moved to the United States and are visiting Target for the first time. Perhaps since you just moved in, your first goal is to furnish your apartment. You look around at the entrance, and navigate your way to the furniture section perhaps while taking a few false turns on the way, buy the stuff you need, pay and leave. The next time you visit for, for example, groceries and produce. You visit again for sporting goods and then again for gifts for your friends. A year later, all such subset of visits through the Target store makes you an expert in knowing the specific route to the section you need to visit. What algorithmic mechanisms allow us to build such expertise to create connections in an explored environment even when during no single visit, you explored all possible connections between regions of the store?

We are able to build up a map of the environment without being exposed to the full extent of it in a single iteration simply based on local exposures to it. Such building is often implicit – you just know that in order to go to the furniture area, you need to pass through the gifts even when you did not explicitly explore this connection before. In this chapter I explore two algorithmic mechanisms that can be used for abtracting structural information from local exposures. I compare the Successor Representation (Dayan, 1993), and the Temporal Context Model (Howard et al., 2005) to generate quantifiable predictions from predictive and associative representations of context,

respectively. I then experimentally test these predictions and show that humans rely on predictive representations to learn the global environmental structure from local exposure.

The effect of local exposure in acquiring structural knowledge of the environment have been explored in several areas of cognitive psychology through artificial grammar and language learning (Aslin & Newport, 2012; Dehaene et al., 2015; Knowlton et al., 1992; Romberg & Saffran, 2010), visual statistical learning (Brady & Oliva, 2008; Fiser & Aslin, 2002; Turk-Browne et al., 2008), or motor sequence learning (Baldwin et al., 2008; Cleeremans & McClelland, 1991; Kahn et al., 2018; Nissen & Bullemer, 1987). In recent work, (implicit) acquisition of higher order knowledge of the environment from lower order exposure is studied through structured graph based transitions between stimuli (Kahn et al., 2018; Karuza, 2022; Karuza et al., 2017; Lynn & Bassett, 2020; Lynn, Papadopoulos, et al., 2020; A. C. Schapiro et al., 2013). For example, A. C. Schapiro et al., 2013 asked participants to study a stream of stimuli based on a connected modular graph **such as the one in** in Figure 2.1. Each stimulus **in their experiment** was associated with a node of the graph (blue circles) and the stream was generated through a random walk where each subsequent stimulus was randomly chosen from the connected neighbors of the current node. Participants were then asked to hit a key wherever it felt like a 'natural break'. Participants often parsed the edges that connect two modules as 'natural breaks' even when their local exposure does not distinguish between the cross-module **e.g. edge between nodes 0 and 14 in Figure 2.1** and within-module edges **e.g. edge between nodes 7 and 8 in Figure 2.1**.

More commonly, global-scale structure acquisition **(i.e. when participants are said to have acquired (implicit or explicit) knowledge of the graph in Figure 2.1)** has been **measured through response times**. Earlier work in serial reaction time tasks shows that breaking an implicitly learned motor sequence leads to slower

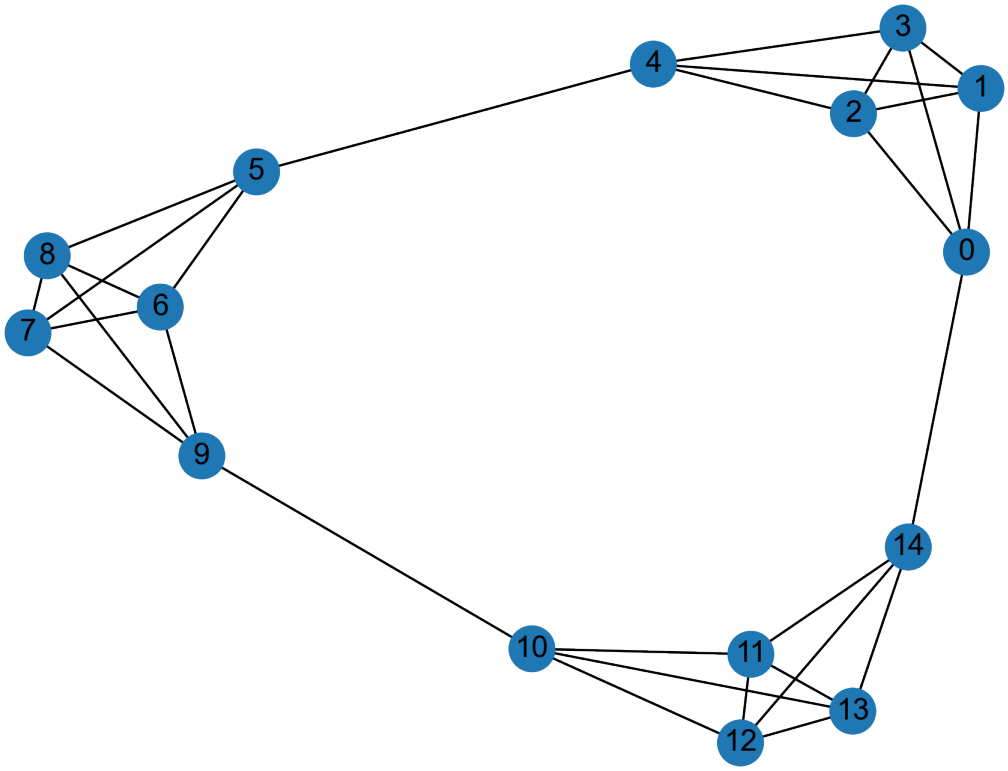
reaction times (Cleeremans & McClelland, 1991; Nissen & Bullemer, 1987). The slowed reaction times when crossing the between-module edges have also been shown in recent work on statistical learning in modular graph structures (Kahn et al., 2018; Karuza, 2022; Karuza et al., 2017, 2019; Lynn & Bassett, 2020; Lynn, Papadopoulos, et al., 2020).

This slowdown across module edges appears to be mediated by the nature of the walk experienced across the community structure where random and Eulerian walk (a walk where each edge of the graph is visited exactly once before repeats) experiences continue to show this slowdown whereas a Hamiltonian walk (a walk were each node of the graph is visited exactly once before repeats) experience does not (Karuza et al., 2017). Thus it appears that the kind of experience through the graph alters the knowledge of underlying statistical patterns. Similarly, the topographical structure of a graph in motor skill learning tasks also appears to alter structural knowledge (Lynn & Bassett, 2020; Lynn, Kahn, et al., 2020; Lynn, Papadopoulos, et al., 2020) where modular graphs like in Figure 2.1 produce the largest dip in reaction times when responding to boundary items.

While these effects are not unique to the modular graph in Figure 2.1 and graphs of various topological variations produce similar effects (Karuza et al., 2019), for the remainder of this work, we focus on using the same modular graph used in A. C. Schapiro et al., 2013. The graph not only provides both the desired modular structure that can translate to pattern extraction, it also provides useful symmetry in the number of nodes in each cluster, the degree of the graph **(equal number of connections at each node), and in allowing for tractability of stimuli associated with each node given constraints on working memory and learning capacity**.

Why do we slow down at boundary nodes that lead to the adjacent module even when the local probability of that particular transition is the same as any other

Figure 2.1. Modular graph structure used in A. C. Schapiro et al., 2013. Locally, each node is connected to four nodes with each edge equally probable. However, globally, the graph structure consists of three sub-modules interconnected through 'boundary nodes'



transitions? Understanding this particular property of human behavior may provide deeper insights into the kind of representations that lead to global-scale structure acquisition. Stimuli in tasks typically used in such statistical learning paradigms are either not meaningless or randomly assigned to each node – the only difference between the boundary node and other non-boundary nodes is in context of the global structure of the graph. The event boundary literature (where boundaries are typically operationalized through explicit changes in context) suggests that boundaries alter the predictability of future events and this predictability leads to event segmentation (Clewett et al., 2019; Zacks & Swallow, 2007). Thus, in implicitly operationalized

boundaries such as in serial reaction time tasks, the slowdown at the boundary node may imply a similarly increased uncertainty at boundary nodes leading to slowed responses. Prior work aimed at understanding human representation of graph structures indeed points to an increased 'cross-entropy' between a learner's estimate of the transition probability and the true transition probability of the environment (Lynn & Bassett, 2020; Lynn, Kahn, et al., 2020; Lynn, Papadopoulos, et al., 2020).

In particular, Lynn, Papadopoulos, et al., 2020 show that algorithms of contextual representations such as the Successor Representation (SR) model in Reinforcement Learning (Dayan, 1993; Gershman, 2018; Momennejad et al., 2017) or the associative learning based Temporal Context Model (TCM) can naturally lead to an increased cross-entropy for cross-cluster transitions relative to within-cluster transitions in modular graphs. In the current work, by using the framework of entropy as a proxy for an estimate reaction times in a modular graph we aim to 1) Experimentally test the predictions of these two models when exposure through the modular graph structure of is partial and 2) Identify which of the two models of representation best explain the observed data.

2.1.1 Representations of Temporal Context

Successor Representation

The Successor Representation (SR) model of reinforcement learning has been used as a model to understand the generalization of reinforcement learning behavior in large action spaces (Dayan, 1993). In recent work, the SR model has also been shown to be a reliable model for explaining human decision-making behavior in multi-step environments. The model's mechanism of model-free, trial and error learning of transition probabilities and model-based learning of rewards accurately predicts that humans are worse at adapting to changes in the transition probability of a learned environment than to changes in the end-point rewards (Momennejad et al., 2017).

There has been further evidence of SR being represented in the Hippocampal cells which represent space (Gershman, 2018; Stachenfeld et al., 2017).

Briefly, the SR model represents each state in the actionable space as a vector of predictive representations. For an environment of N discrete states, the SR matrix M of size (NXN) maintains expected future visits to a given state from each state. Specifically, element $M_{i,j}$ of the matrix represents the expected future visits to state j from state i. This transition matrix is learned over time based on the temporal difference error learning rule (Sutton, 2018). For example, consider at a given point in time, t, an agent maintaining the SR matrix is in state i. The agent now moves to state j out of the possible N states. The i^{th} row of the SR matrix is updated as follows:

$$\hat{M}_{i,j} \leftarrow \hat{M}_{i,j} + \alpha[\delta(s_{t+1}, j) + \gamma * \hat{M}_{s_{t+1},j} - \hat{M}_{s_t,j}] \tag{2.1}$$

where $\delta(.,.)$ equates to 1 if both arguments are equal otherwise it equates to 0. Thus, the matrix increases the probability of visiting a state j from state i if state j is visited in the current experience and it decreases the probability of visiting all other states from state i. Parameter α is a learning rate parameter that determines how much of the previous estimate of visiting state j from i is factored into the current update. Parameter γ is a future discount parameter that dictates how much in the future the agent sees – specifically, a higher value of γ indicates future visitations to state j are weighed high in the current update.

For example, let's assume that the entire world (from the perspective of a participant) constitutes the 15 items they will see in the study. Now, as they start their experience of a random walk on the graph in Figure 2.1, they will initially have no information about the transition probability structure. Thus, they will assign an equal probability across all possible transitions (i.e., $\frac{1}{15}$ between each pair of nodes/stimuli).

Let's say they experienced a transition from node 1 to node 2. The prediction error for observed transition will be positive (example below):

$$\begin{aligned} \Delta\hat{M}_{1,2} &= 1 + \gamma * \hat{M}_{s_{2,2}} - \hat{M}_{1,2} \\ &= 1 + \gamma * \frac{1}{15} - \frac{1}{15} > 0 \end{aligned} \tag{2.2}$$

Similarly the prediction error for all transitions that were *not* observed from node 1 will be negative (example below):

$$\Delta\hat{M}_{1,3} = 0 + \gamma * \hat{M}_{s_{2,3}} - \hat{M}_{1,3} = \gamma * \frac{1}{15} - \frac{1}{15} < 0 \tag{2.3}$$

Note that in the above equations, the model also accounts for a future transitions $M(2,2), \& M(2,3)$ while updating its expectation of transition $M(1,2)$, and $M(1,3)$ respectively. This way, any current transition impacts the expected trainstitions to the future nodes as well. **By allowing for transitions expected in the future to weigh in on current updates, this learned matrix allows for inferring for transitive properties in the graph structure.**

Temporal Context Model

The Temporal Context Model (TCM) was devised to explain the primacy and recency effects in human recall and recognition memory (Howard et al., 2005). The TCM model assumes that the items or stimuli shown to a participant during a study phase through a sequential exposure maintain a temporal context **as a vector of activity of all stimuli in the experiment**. As new items get encoded, existing context from the previous items allows the new items to be bound to the previously seen items thereby sharing the temporal context. Briefly, the TCM can be formalized as in Gershman et al., 2012:

$$
\begin{aligned}
t_n &= \rho * t_{n-1} + f_n \\
\hat{M}_{i,j} &\leftarrow \hat{M}_{i,j} + \alpha f_{n+1} t_{n,i}
\end{aligned}
\tag{2.4}
$$

where t_n is said to be a 'context' vector for item n. The context drift parameter ρ determines the proportion of the previous elements's context that gets incorporated in the current context. f_n is a one-hot encoded vector for item n – element of the N item-vector f_n is 1 for the item it represents and 0 otherwise. The learning rate parameter α determines what proportion of the currently experienced state binds with the existing context.

For a similar experience as described in the example for the SR model, the the TCM context vector t_n is first updated based on the prior, existing context t_{n-1}. For the experienced transition from node 1 to 2 and assuming the current transition to node 1 came from node 0:

$$
\begin{aligned}
t_1 &= \rho * t_{n-1} + 1 \\
\hat{M}_{1,2} &= \hat{M}_{1,2} + \alpha * 1 * t_{1,1}
\end{aligned}
\tag{2.5}
$$

Note that for a transition that is *not* experienced, the cell $\hat{M}_{i,j}$ does not get updated as f_{n+1} would be 0.

The key difference between the two models of temporal context is two fold: (1) SR Relies on error-based learning whereas TCM relies on hebbian, associative learning and (2) Through the future discount parameter γ, SR also learns the predictability observing states in the near future based on the locally experienced transitions. This future discount parameter in SR thus allows the model to represent transitive associations as well (Gershman et al., 2012).

2.1.2 Model Simulations

The differences between the models stated above lead to differing representations of learned temporal structure when the models are exposed through the random walk

in Figure 2.1. **To preview, representations associated with boundary nodes carry more information in the SR than the TCM.**

Simulating the models described above can thus lead to an estimate of expected behavior from both these models. Differences in these expected behaviors thus allow us to generate experimentally testable hypotheses. Figure 2.2 shows the context matrix representation after the models have been simulated for a random walk through the graph structure in Figure 2.1 as a result of a random walk after 1000 trials for both models. The matrices of Figure 2.2 represent the model predictions of context representations. Each row corresponds to a context representation of that node (of the 15 total nodes). The 'activity' levels in each cell indicated by the heatmap represent the relative proportion of nodes that are active when a particular node is visited.

Parameters used in the simulations shown in Figure 2.2 were determined by a Representational Similarity Analyses style procedure (Kriegeskorte et al., 2008). Specifically, for a combination of parameters over a valid range, a distance matrix was computed **to represent** an euclidean distance between each pairs of rows of the generated context matrix (SR or TCM). Parameters that maximize the correlation between this generated distance matrix and the true distance matrix (derived from transition matrix representing equal probability transitions across the connected edges in the modular graph of Figure 2.1) were chosen via a grid search.

By using a behavioral measure of response times, previous work has shown that participants can acquire the global structure of the graph for a random walk. Particularly, as participants acquire knowledge of the underlying structure, their reaction times are slower in responses following a cross-cluster transition relative to a within-cluster transition (Kahn et al., 2018; Lynn, Kahn, et al., 2020). To model this observed difference in reaction times and link them to the apparent differences shown in Figure 2.2, we apply principles of information theory. Specifically, we assume that response time for each stimulus is a function of the uncertainty in its surrounding

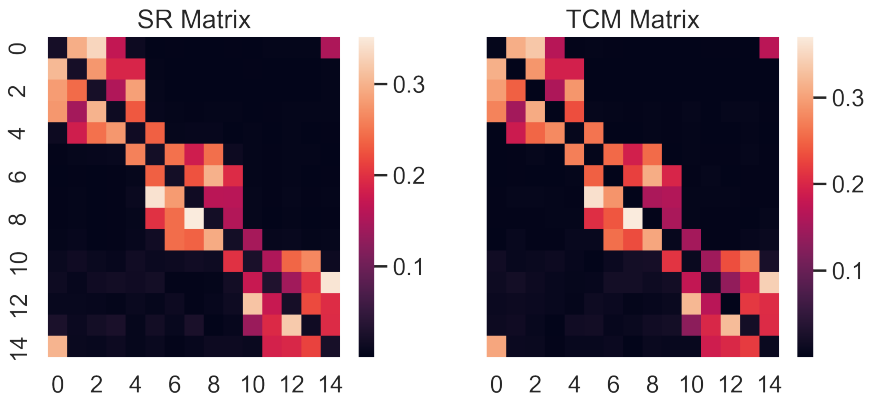


Figure 2.2. Successor Representation and Temporal Context Model representations of context following a random walk through the modular graph structure.

context (Fitts & Peterson, 1964). Measures of information entropy have previously been used to explain RT differences between cluster transitions while traversing similar graph structures – where higher entropy, which implies more uncertainty or more information available for a participant to process leads to higher response times (Lynn & Bassett, 2020; Lynn, Kahn, et al., 2020; Lynn, Papadopoulos, et al., 2020). Formally,

$$RT(node) \cong Entropy(node) = \sum_{s' \in S} \hat{M}(s, s') * log(\hat{M}(s, s')) \tag{2.6}$$

where $M(s, s')$ is the context representation at node s. For SR, this expression evaluates to the expected future visits to state s' from state s whereas for TCM this expression evaluates to the extent to which s' is activated as a result of s.

As noted previously, a common indicator of participants having acquired the global structural knowledge is a slowdown in responses when the ongoing stimulus stream crosses a cluster (relative to transitions within a cluster) of the modular graph. Context representations can be used to model the cross cluster-transitions by computing a 'surprisal' effect. For simulations, the surprisal effect is computed as the Jenson-Shannon distance between the context representations of two nodes. Formally,

$$RT(s \to s') \cong JS(s, s') = \sqrt[2]{\frac{D(M(s,.)||p) + D(M(s',.)||p)}{2}} \tag{2.7}$$

where $M(s,.)$ is the context representation of node vector s, p is the point-wise mean of nodes s and s' and $D(M||p)$ is the Kullback-Leibler divergence between probability distributions M and p. Jenson-Shannon distance thus scales with differences between the two context representations. Intuitively, since the representations of nodes 1 and 2 are similar (as shown in Figure 2.2), the Jenson-Shannon distance between these two nodes will be smaller than nodes 1 and 6. A direct measure or surprisal derived from the context matrix was also considered (See Figure A.1 in the appendix for details)

The formalization of observed response time differences due to surprisal (and node entropy) allows us to simulate expected reaction time distributions for novel walk types. Specifically, to understand the mechanisms behind acquiring the global modular graph pattern following a limited exposure (for example, infering the target store's organization via limited exploration over multiple visits), each model was simulated for random walk with lengths of 0, 3, 6, and 999. A random walk length of 0 translates to a completely random selection of one of the 15 nodes of the modular graph on each trial. Walk length of 3 and 6 translate to a random walk visiting 3 and 6 edges (4 and 7 nodes) respectively before being reset to a random node (similar to visiting the Target store in short bursts to purchase relevant items and checking out without visiting the entire store). Finally, a walk length of 999 translates to visiting 999 edges (with repetition) through their connections on the modular graph. Parameters of the simulations in Figure 2.3 are determined through the same best-fitting RSA procedure described above.

The acquisition of the global structure can be modeled using surprisal as has been done in previous research (Lynn & Bassett, 2020; Lynn, Kahn, et al., 2020; Lynn, Papadopoulos, et al., 2020). To investigate the differences between models for

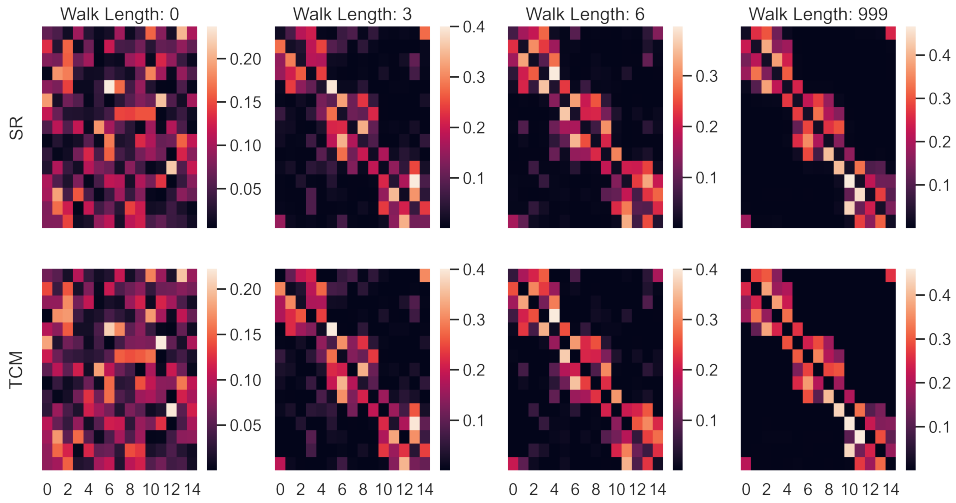


Figure 2.3. Example model predictions of context representations for SR and TCM models across different walk lengths for one specific set of parameters. Both models seemingly predict that the modular structure of the original graph is increasingly recovered with longer walk lengths.

various walk lengths and relate to measurable response time differences, for a subset of parameters in the valid range of 0 to 1, each model was simulated to produce a context matrix. Jensen-Shannon distance was computed between each pairs of nodes and averaged over cross-cluster pair and within cluster pairs. Simulation results below show the transition Jensen-Shannon distances over 100 simulations of the model for each parameter combination. For SR, 'param_a' is the learning rate parameter α and 'param_b' is the discount parameter γ. For TCM, 'param_a' is the learning rate parameter α and 'param_b' is the context drift parameter ρ.

In Figure 2.4 The Y-axis represents the difference in surprisal between boundary nodes and non-boundary nodes as measured by Jensen-Shannon distance. The X-axis represents the parameter α and different hues represent parameters γ for the SR model and ρ for the TCM model. Top row are predictions for the SR model and bottom row for the TCM. Columns represent predictions for different walk lengths..

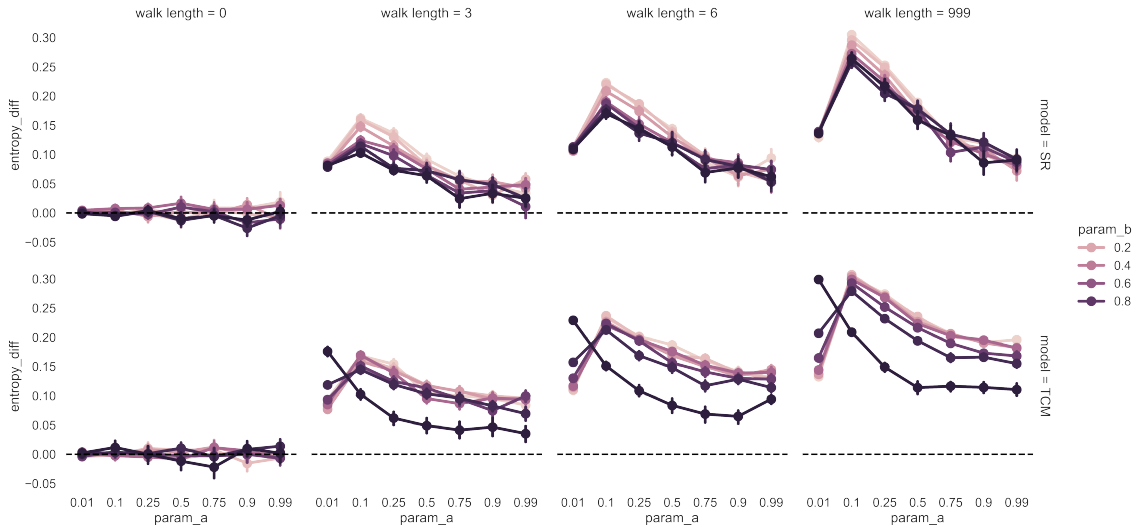


Figure 2.4. Simulated model predictions for differences in surprisal comparing across cluster transitions to within cluster transitions across walk lengths for a range of possible parameter values. Both models predict that cross cluster surprisal effect will increase with walk length leading to an increased reaction time.

The figure thus shows that both context models predict an increased surprisal in cross-cluster transitions relative to within-cluster transitions as walk length through the modular graph gets longer. As walk length increases, context associated with each node increasingly represents neighboring nodes. Since neighbors of the boundary nodes are largely within the cluster of that boundary node, representations of boundary nodes becomes increasingly similar to that of the non-boundary nodes within the same cluster, and thus increasingly different than boundary nodes in the neighbouring cluster. Thus, crossing a cluster (from a boundary node to another) leads to an increased surprisal of having encountered a node that is representationally dissimilar to the previous node. On the other hand, non-boundary nodes within the same cluster get increasingly closer in their representations with other non-boundary nodes in that cluster. Thus transitions between non-boundary nodes within a cluster does not increase surprisal.

The two context models, however, differ in their predictions in the role of a boundary node. The Y-axis of Figure 2.5 shows the difference in entropy between boundary nodes over a range of parameters (X-axis and hue) for both SR and TCM models (rows) across the 4 walk lengths (columns). SR predicts an increased entropy in its representation of the boundary nodes with walk length relative to the non-boundary nodes for some values of the α and γ parameters. On the other hand the TCM does not predict such increased in boundary vs non-boundary entropy differences.

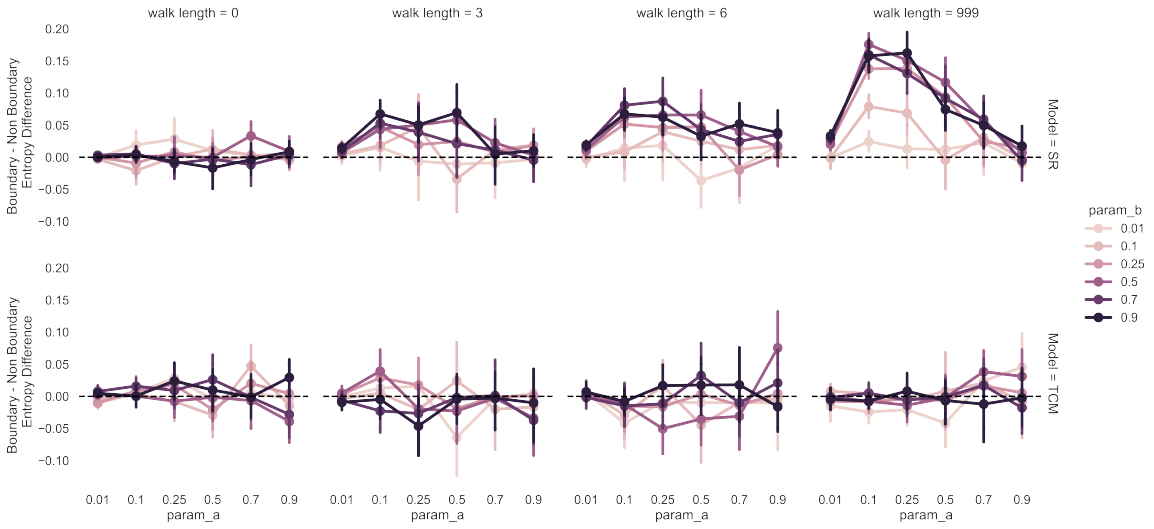


Figure 2.5. Simulated model predictions of differences between the SR and the TCM after different walk lengths for a range of possible parameter values. SR predicts that entropy of boundary nodes will scale with walk lengths whereas TCM does not.

The predictive nature of SR (as modeled by the future discount, γ parameter) allows for a representation of nodes in the neighboring cluster to impact entropy on the boundary node of the current cluster that leads to that neighboring cluster. This effect is unique on boundary nodes of a cluster as non-boundary nodes of the second cluster are closer to the immediate neighbor of the current cluster (i.e. the boundary node that serves as an entry point to the second cluster). Since TCM is associative (as opposed to predictive), only nodes that are 'active' in representation impact the representation of the just experienced node thereby. This mechanism thus reduces

the impact of the non-boundary nodes in neighboring cluster. Rescaled heatmap in figure 2.6 presents this effect.

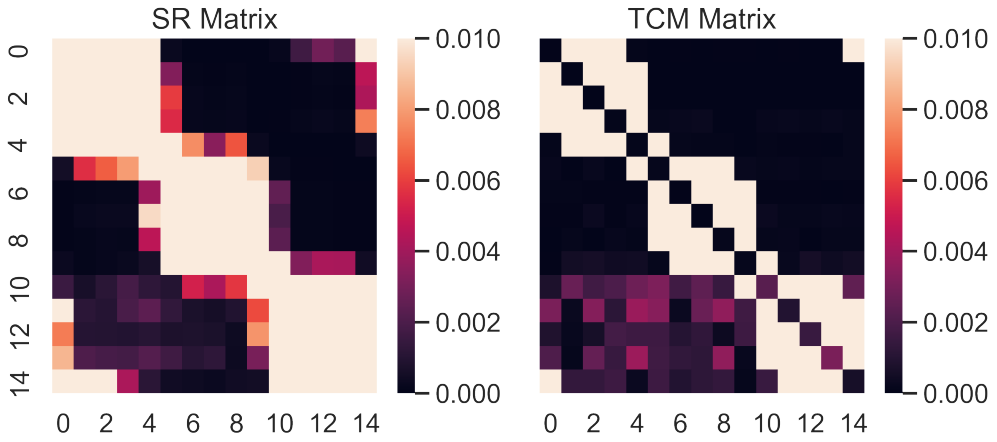


Figure 2.6. Rescaled SR and TCM matrices depict differences between context representations of the two models. Boundaries in SR incorporate more information than those in TCM.

The SR-based predictive context representation in particular shows that boundary nodes carry more information than non-boundary nodes whereas the associative context representation does not produce this effect. [1]

Thus, both SR and TCM models would predict slow down in cross-cluster transitions relative to within cluster transitions, and that this slow down will increase with walk length. However, predictive context representations through SR are unique in predicting the scaled slow down at boundary nodes with random walk length, *independent* of where the boundary node has been visited from but as a result of the boundary nodes' inherent role in serving as a gateway between clusters. While lack of a scaled slow down to boundary nodes does not invalidate the SR model (because

[1]The activity in the lower third of both matrices is due to recency; while these are interesting patterns, and seem to indicate that SR can account for the recency effects in memory which was the primary motivation behind introduction of the TCM (Gershman et al., 2012; Howard et al., 2005). Investigating recency effects in this implicit statistical learning context is out of scope for this dissertation.

some values of the parameters allows SR to not scale the slowed reactions with walk length), the presence of such a slow down provides evidence for predictive representations in such statistical learning tasks. The study presented next, thus tests this prediction.

2.2 Experiment 1: Testing Context Representations for implicit event boundaries

2.2.1 Methods

Participants

125 undergraduate students at the University of Massachusetts Amherst participated in this study for course credit. Data from 12 participants who did not complete the study was disregarded from further analyses. All study protocols were approved by the university institutional review board. Participant sample size was not predetermined via a statistical procedure but was a rough equivalent of previous studies (Kahn et al., 2018; A. C. Schapiro et al., 2013). [2]

Design and Procedure

The general experimental procedure was similar to the one used by Kahn et al. (2018). Participants were randomly assigned into one of four between-subject groups. All procedures for participants in all groups remained the same except for experimentally defined walk-lengths. Participants sat in an isolated room with an LCD computer screen operated by Windows 7. The experiment was designed using Psychopy (Peirce, 2007).

As shown in Figure 2.7, at the beginning of the study, participants were instructed to place their right hand on the computer keyboard such that their fingers aligned on

[2] All inferences in this work are made in form of the probabilities of an effect as estimated via Bayesian analyses.

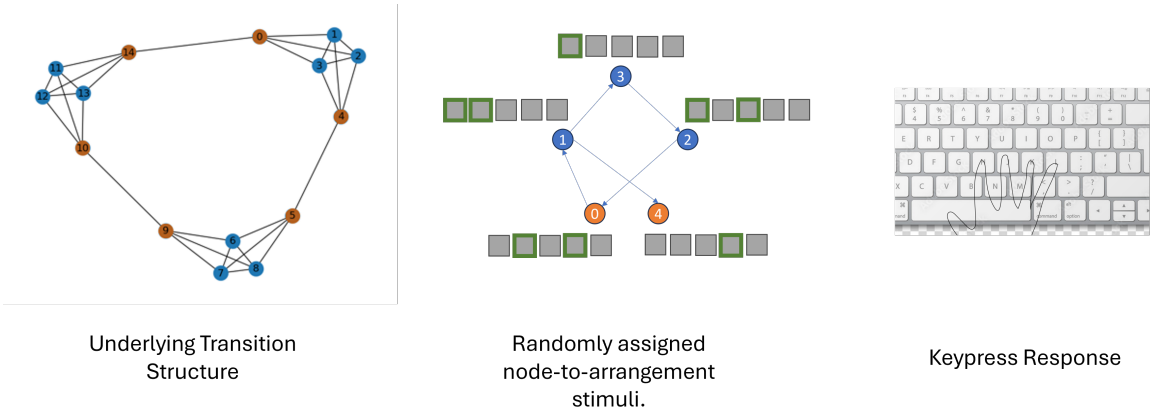


Figure 2.7. Task design for experiment 1. *Left panel* modular graph used to generate random walks. *Middle panel* Each node is randomly assigned to a combination of one or two highlighted boxes. *Right panel* Participants place their hands on the keyboard as shown and are instructed to press keys that correspond to highlighted boxes.

the appropriate keys. On each trial, participants were presented with five grey boxes. One or two of the five grey boxes were highlighted using green borders. Participants were instructed to hit the combination of keys (simultaneously in cases where two keys were required to be hit) corresponding to the highlighted boxes as fast as possible without making any errors. A trial did not end until participants hit the correct combination of keys. Participants were informed of their incorrect key presses and a trial where participants did not hit the correct combination of keys on the first try was marked as an inaccurate trial. The experiment lasted for 1400 trials, or 1 hour, whichever came first. To prevent fatigue, participants were given self-paced breaks after every 200 trials. Data from participants who did not complete all 1400 trials was discarded and not used for further analyses. At the end of the study, participants were debriefed about the research question of the study.

Each of the 15 possible key combinations (where one or two of the grey boxes are highlighted) were randomly assigned to a node of the graph in 2.1. Trial sequences were generated based on walk lengths. For all walk lengths, the first trial was selected at random. For walk lengths of 1399 (29 participants), the subsequent trials followed

a random walk through the graph structure along the edges with edges connected. For walk lengths of 3 and 6, trials proceeded on a similar random walk for 3 (29 participants) and 6 (29 participants) edges respectively (thus visiting 4 and 7 nodes) before resetting to any of the fifteen nodes. Finally for random walk of length 0 (29 participants), each node of the 15 was picked with equal probability on each trial.

Data Processing and Preliminary analyses

All inaccurate trials (around 9.2%) were removed from further analyses. Furthermore, accurate trials with response times beyond 3 standard deviations of the global response times (around 1.1%) were removed from further analyses as well. In all, around 10% of the total trials were discarded.

2.2.2 Results

Table 2.1 shows the descriptive statistics (means, medians, and standard deviations) of response time for each node and transition types aggregated over participants in each condition. The 'Transition Type' column refers to the transition experienced immediately prior to a particular node where 'within cluster' transition is the transition to a boundary or a non-boundary node from that same cluster and 'cross cluster' transition is a transition to a boundary node from a boundary node of a neighbouring cluster. 'Node Type' refers to the role of the current node in the graph of the left panel in Figure 2.7.

As expected, response times decreased with practice for all nodes in all conditions (Tables A.4, A.5, and A.3 for statistical comparisons). The median response times separated by transition type and node types are shown in Figures 2.8 and 2.9 respectively. Note that responses to Boundary node in Figure 2.9 are a combination of within- and cross-cluster transitions.

Response time graphs show the overall pattern **expected from model simulations above** – cross cluster transitions are slower in longer walk lengths than within

Table 2.1. Response times descriptive statistics (in seconds) for experiment 1.

transition type	node type	walk length	rt mean	rt std	rt median
cross cluster	Boundary	0	0.954	0.565	0.786
		3	0.963	0.585	0.774
		6	0.973	0.594	0.785
		1399	0.990	0.596	0.802
within cluster	Boundary	0	0.996	0.565	0.822
		3	0.980	0.598	0.787
		6	0.943	0.572	0.769
		1399	0.953	0.600	0.767
	Non Boundary	0	0.963	0.565	0.790
		3	0.936	0.545	0.772
		6	0.985	0.616	0.790
		1399	0.932	0.592	0.747

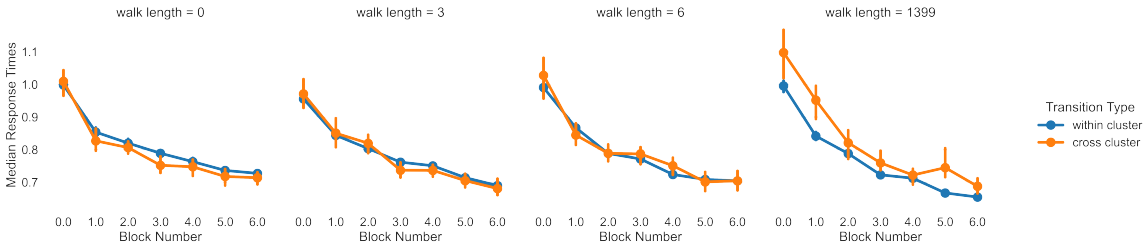


Figure 2.8. Median response times at nodes following a transition for each walk length separated by the type of transition.

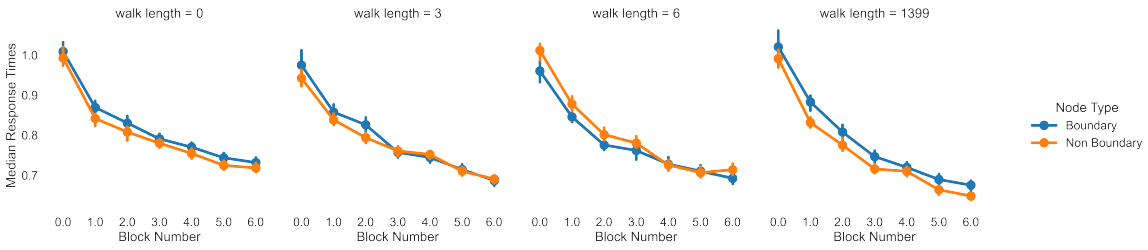


Figure 2.9. Median response times for each walk length separated by node types.

cluster transitions. Similarly, responses to boundary nodes are slower at longer walk lengths than responses to non-boundary nodes.

Modeling

The key question of interest is whether response times to boundary nodes slow down further *after* accounting for slowdowns due to transitions. However, the effects of node type (comparing to non-boundary nodes) wll necessarily include effects of transition type since boundary nodes are accessed through both within cluster and across cluster transitions. Therefore, in order to isolate the effects of node type, response time differences between boundary and non-boundary nodes for within-cluster transitions were compared; thereby removing the effects of cross cluster transitions. Since transitions for walk length of 0 were random, this condition was also removed from further analyses. Similarly, all reset transitions in walk lengths of 3 and 6 (where the random walk ended and the subsequent node was picked between any of the 15 possible nodes) which were *not* a part of the random walk were discarded from further analyses to isolate the effect of the transition structure during a random walk (since nodes following resets do not follow the transition structure).

Each block in the experiment consisted of 200 trials. Thus, by the end of the first block, participants in longer walk length conditions may have already experienced the graph structure sufficiently enough to acquire knowledge of the graph, thereby slowing down at the boundary nodes. While characterizing the entire 1400 trial learning curve is ideal to measure the differences in reaction times of the critical transitions (non-boundary to non-boundary node compared with non-boundary to boundary node transitions), such linear model fit leads to the differences between reaction times at boundary and non-boundary nodes to map onto different parameters of the (see Appendix for a linear model of the entire learning curve) across different walk length making across walk length comparisons difficult. Thus, to assess acquired patterns,

the first two blocks of the data are compared using the following Bayesian model where standardized log response times were fit as a skewed normal distribution, separately for each walk length.

$$\begin{aligned} node\ transition\ type : block &\sim \mathcal{N}(0, 0.5) \\ transition\ experience &\sim \mathcal{N}(0, 0.2) \\ lag &\sim \mathcal{N}(0, 0.2) \\ \mu = node\ transition\ type : block + lag + transition\ experience & \\ \sigma &\sim Exponential(1) \\ skewness &\sim \mathcal{N}(0, 3) \\ log(RT) &\sim Skew\mathcal{N}(mu, sigma, skewness) \end{aligned} \tag{2.8}$$

Where *node transition type* is either non-boundary to non-boundary or non-boundary to boundary; block is 0 or 1, lag is the number of trials before which the current key combination was seen and transition experience is the number of times a particular transition leading into the current node was previously experienced. Figure 2.10 shows the Bayesian estimates of differences between walk lengths. The X-axis of the posterior histogram represents an estimate ofthe differences between response times at boundary nodes and non-boundary nodes in the second block (Block 1) after controlling for responses in the first block (Block 0) separeted by walk lengths of 3, 6, and 1399.

In particular, participants in walk length of 1399 experienced the largest gains in response times of non-boundary to non-boundary transitions relative to those of non-boundary to boundary transitions from the first block to the second. As expected from the Successor Representation (SR) model, these gains were reduced for walk length of 6 and further so for walk length of 3 implying varying levels of structure acquisition depending on walk length. This pattern, which is uniquely expected in

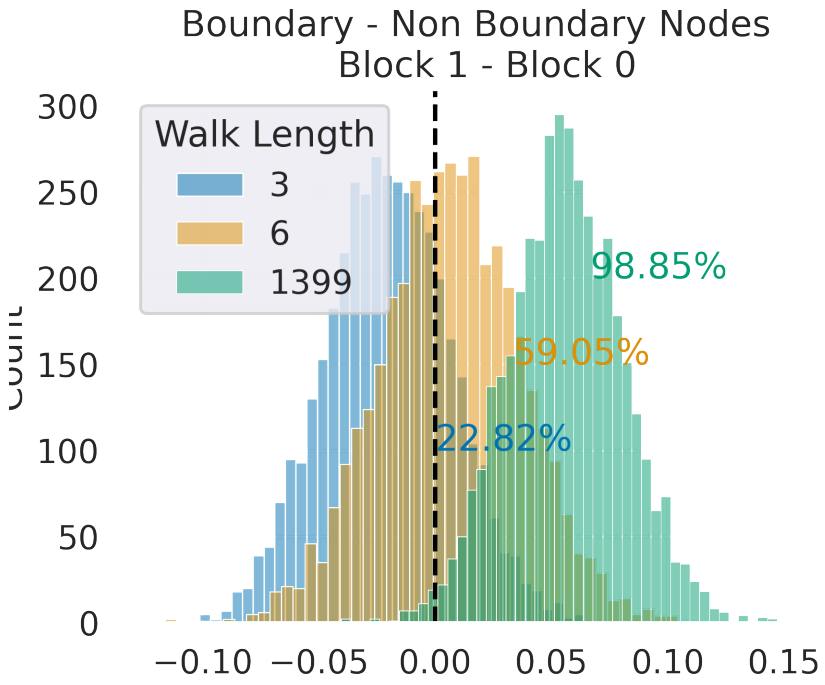


Figure 2.10. Estimated differences in response times to boundary and non-boundary nodes when they are transitioned into from the same cluster (i.e. another non-boundary node). When accounting for the response times in the first block, as walk length increases, response times in the second block are increasingly slower to boundary nodes than non-boundary nodes.

the SR model and not the TCM, thus provides support for predictive representations driving the formation of implicit event boundaries.

2.3 Discussion

The primary aim of this work was to characterize the creation of implicitly operationalized event boundaries as a function of context representations. In particular, two models of context representations were contrasted: the associative TCM model and the predictive SR model. Both models express an increase in reaction time when crossing boundary nodes into a new cluster in the three-module graph structure (Figure 2.1) as available context at boundary nodes across clusters drastically differs with each boundary node strongly representing events within its own cluster. However, the

SR model expresses the importance of boundary node as carrying additional information (measured by information theoretic entropy). The SR model predicts that as the 'quality' of exposure (here operationalized by length of random walk) increases, the apparent importance of boundary nodes increases as well.

To test this qualitative prediction of the SR (and thereby compare it with the TCM representation), a serial reaction time task was conducted with participants experiencing the modular graph at 4 different lengths of a random walk. As predicted by the SR, response times at boundary nodes slowed down the most for the longest random walk, and less so for shorter random walks.

The experimental findings in this chapter thus provide support for maintaining a predictive representation of our environment and that associative, Hebbian mechanisms, are not enough to explain the observed data. This error-driven predictive representation, which does not rely on explicit rewards, naturally leads to learning the statistical regularities in the environment and is thus crucial in informing our understanding of statistical learning and pattern acquisition.

While SR is not unique in its expression of increased boundary information for a modular graph used in this work, other formulations (where increased boundary information is a result of erronous estimation of the transition probability) in prior work have been linked to closely follow the SR model (Lynn & Bassett, 2020; Lynn, Kahn, et al., 2020; Lynn, Papadopoulos, et al., 2020). Furthermore, this SR formulation allows us to use a neuro-psychologically plausible model in understanding pattern extraction and statistical learning (Gershman, 2018; Momennejad et al., 2017; Stachenfeld et al., 2017). Finally, such predictive representations set stage for use of the SR (or SR-like) models over associative models to understand the broader work in event cognition and event boundaries (Rouhani et al., 2020).

In the current work parameters of the SR (or the TCM) model are not directly estimated as SR and TCM do not provide a direct measure of reaction times. While the

assumption of reaction times scaling with increased available information (entropy) is logical, this assumption needs further testing. Future tests of such predictive representation should incorporate parameter estimation and hence also check the validity of the relationship between reaction time and information entropy. Similarly, a simpler model comparing two blocks was used to make inference in this task. More complex models (such as the exponential or the multi-rate state space models (McDougle et al., 2015; Savalia et al., 2024; M. A. Smith et al., 2006)) should aim to characterize the entire learning curve to understand when participants start to acquire (and use) existing patterns.

An SR based predictive representation, on its own is likely not sufficient to explain all patterns in the data. In some cases, participants may become explicitly aware of an existing structure, leading to a model-based reinforcement learning intervention on reaction times (Momennejad et al., 2017) or constraints on working memory may require further considerations in how transition probabilities are learned (McDougle & Collins, 2021). Future work should account for these possibilities in understanding the cognitive processes that underlie statistical learning. Finally, findings presented in this work are limited to a single graph structure. Prior work has found that graphs of different topologies produce similar effects in cross-cluster slowdowns (Karuza et al., 2019) and future work should examine the modeling and experimental differences for a range of graph structures to assess whether findings in this work are dependent on the specific topological structure used.

2.4 Conclusion

The findings in this chapter provide evidence in favor of using predictive representations (as opposed to associative representations) to account for event boundaries that are operationalized implicitly. Findings in the current chapter do not distinguish between specific algorithms that lead to predictive representations; future work

could contrast potential differences in these algorithms. It however remains unclear whether 'boundary' nodes are truly so in context of event cognition – the experiment presented in this chapter (and similar past literature) does not test whether stimuli at boundary nodes follow similar properties as stimuli at boundaries when events are operationalized through explicit context change. In the next chapter, I explore how such implicitly operationalized boundaries share properties with boundaries that are operationalized explicitly.

CHAPTER 3

COMPARING IMPLICIT EVENT BOUNDARIES WITH EXPLICIT EVENT BOUNDARIES

3.1 Introduction

We receive a continuous stream of sensory information in our daily lives. In order to make sense of it, we often parse it into meaningful chunks for storage, retrieval and comprehension. For example, we may recall our drive to work as a series of discrete events; got into the car, got coffee, picked up a colleague, hit traffic on a particular street, parked, and walked over to the office. What aspects of the incoming stream help us organize continuous temporal information in such discrete chunks? Temporal chunking in cognitive psychology has been studied under several domains from event boundaries (Baldwin et al., 2008; Clewett et al., 2019; DuBrow & Davachi, 2013; Rouhani et al., 2018, 2020; Zacks & Swallow, 2007), language learning, (Knowlton et al., 1992; Romberg & Saffran, 2010), categorization (Gabay et al., 2015; Unger & Sloutsky, 2022), and motor sequencing (Bera et al., 2021; Ostlund et al., 2009; Savalia et al., 2016; Tremblay et al., 2010). Chunking a repeated sequence of experiences is crucial to abstracting patterns in the environment and formation of habits for quick and efficient interactions with the environment (Botvinick, 2012; Dezfouli & Balleine, 2012; Dezfouli et al., 2014; Dolan & Dayan, 2013; Gershman & Niv, 2010; K. S. Smith & Graybiel, 2016).

Models of temporal event segmentation suggest that the points which lead to temporal segmentation seem to be unique in their properties in both segmenting the continuous stream of information and integration of information across the temporal

event. These 'event boundaries' are, for example, shown to be remembered better (Heusser et al., 2018; Radvansky & Zacks, 2017; Rouhani et al., 2018; Swallow et al., 2009; Zacks, 2020), serve as points of retrieval (Michelmann et al., 2023) and replay to promote long term memory (Hahamy et al., 2023; Sols et al., 2017) and easy parsing, help integrate memory across time (Clewett et al., 2019), and separates across boundary events while collapsing within boundary events (Brunec et al., 2018; Clewett et al., 2019; Ezzyat & Davachi, 2014; Lositsky et al., 2016).

In most prior studies, event boundaries have been primarily studied using explicit context shifts. For example, when stream of stimuli are surrounded by colored border, event boundaries are operationalized by first showing the stimuli surrounded by a color and abruptly changing that color (Heusser et al., 2018). In another study, event boundaries were operationalized via explicit context changes by changing the associated stimulus (Ezzyat & Davachi, 2014). In this study, a pair of images were presented on each trial one image of the pair, the 'scene' image remained constant for a short sequence of trials whereas the other ('object' or 'face') changed on each trial. Participants were asked to make judgments about the object/face image (Ezzyat & Davachi, 2014). In these and other prior studies on event boundaries, context changes had been operationalized as either perceptual or semantic shift in ongoing set of events by having participants watch clips (Swallow et al., 2009). In more recent work, context change has also been operationalized as changes in ongoing reward contingencies associated with each stimulus (Rouhani et al., 2020).

Consistent findings across most studies in explicitly operationalized event boundaries (events, or stimuli in an experimental paradigm which signal a shift in the ongoing context) show that event boundaries are often remembered better (Baldassano et al., 2017; Clewett et al., 2019; Ezzyat & Davachi, 2014; Heusser et al., 2018; Radvansky & Zacks, 2017; Rouhani et al., 2020; Swallow et al., 2009). Furthermore, events that are separated by boundaries events appear to be perceptually farther

whereas events within boundaries appear to be perceptually closer (Brunec et al., 2018; Clewett et al., 2019; Ezzyat & Davachi, 2014; Lositsky et al., 2016).

Recent work has shown that event boundaries can also be formed *without* explicit changes in context. After being exposed to a stream of stimuli such that the ordering is controlled by a modular graph shown in figure 2.1, participants seem to recognize across-cluster transitions as 'natural breaks' more often than within-cluster transitions (A. C. Schapiro et al., 2013). This finding has been linked to statistical learning of temporal graph structures and the effect of event boundaries is often measured by slowed reaction times when responding to a stimulus after experiencing a transition across-clusters than within-clusters (Kahn et al., 2018; Karuza, 2022; Karuza et al., 2019; Lynn & Bassett, 2020; Lynn, Kahn, et al., 2020; Lynn, Papadopoulos, et al., 2020).

Crucially, in these studies, there are no systematic perceptual differences between the stimuli associated at the boundary nodes in Figure 2.1 and those associated at non-boundary nodes. Rather, boundaries are operationalized as a function of the temporal properties in which these stimuli are presented. These implicitly operationalized boundaries have been labelled as 'event boundaries', suggesting that implicit boundaries share representational properties with explicitly operationalized boundaries typically studied in the event boundary literature. However, past studies on implicit event boundaries do not assess the memory representations of these boundaries using the same tests used in explicitly operationalized boundary paradigms, instead relying on observations from reaction times (or the rate at which boundaries are detected as 'natural breaks'). To claim that implicit event boundaries are, in fact, event boundaries, such that they share mental representations with explicit boundaries, these implicit boundaries must be tested on the same tests that have been used in the explicit event boundary literature. Testing whether the two types of event boundaries share mental representations will thus directly test whether there

exist common mechanisms underlying boundary formation and provide a unifying framework to study event cognition.

In this work, I present two tests on implicitly operationalized boundaries to assess whether they elicit the same behavioral properties as the explicitly operationalized boundaries. In particular, I use the paradigm and graph structure previously used in A. C. Schapiro et al., 2013 to test whether participants recall boundary items better (or worse) than non-boundary items. I then use a two module graph structure in Figure 3.8 to test whether items across the two clusters appear farther than items within a cluster (similar to findings in explicitly operationalized boundary paradigms).

3.2 Experiment 2a: Boundary Memory

Modeling Boundary memory benefits

Event segmentation theory suggests that the segmentation of the continuous sensory experience occurs automatically and through prediction errors (Swallow et al., 2009; Zacks & Swallow, 2007; Zacks et al., 2007). According to the event segmentation theory, we maintain an ongoing 'context' which is predictive of upcoming events. Event boundaries are created when this prediction breaks. More recent work has shown that prediction errors are not necessary for creation of event boundaries; a change in uncertainty of the upcoming events can also produce event boundaries (Shin & DuBrow, 2021). Prediction errors particularly lose their value in learning new information when the explored environment is uncertain (Behrens et al., 2007). Nevertheless, under environments with high regularities, prediction errors remain the key mechanisms driving boundary formation.

As reviewed above, prediction errors need not be explicitly operationalized for an event boundary to be learned. Prediction errors which imply shifts in ongoing context, similar to implicitly operationalized event boundaries, can also be implicit. In chapter 2, I showed that context models can be used to estimate representations of implicitly

operationalized event boundaries. Particularly, predictive representations such as the SR provide a natural representation of event boundaries which form bottlenecks in transitioning between clusters in modular graphs such as one in Figure 2.1. In the current work, I propose that the same predictive context-representation framework using the Successor Representation model of Reinforcement Learning (Dayan, 1993; Gershman et al., 2012; Momennejad, 2020; Momennejad et al., 2017; Russek et al., 2017) can be used to model differences in memory representations.

To simulate a recognition memory task, I employ a simplified version of the exemplar-based Generalized Context Model (Nosofsky, 1986, 2011; Nosofsky et al., 2011). The GCM falls under a class of global matching exemplar models where each studied item is stored as an image or an exemplar in memory. At test, the presented test item is matched with memory representations of stored exemplars by computing the psychological similarity between them. It is assumed that if the similarity, summed over all similarities of the test items with exemplars in memory, has a higher value, the participant has a higher chance of recognizing that item and the 'old' response is chosen in the old/new recognition test. Similarly, a 'new' response is chosen with a higher probability when the summed similarity of the test item is low.

The GCM model for recognition memory can be formalized with the following equations from Nosofsky et al., 2011:

$$\begin{aligned} d_{ij} &= [\sum_{k=1}^{K} w_k (x_{ik} - x_{jk})^2]^{1/2} \\ s_{ij} &= \exp^{-c_j d_{ij}} \\ a_{ij} &= m_j s_{ij} \end{aligned} \tag{3.1}$$

where d_{ij} is the psychological distance between test item i and Exemplar j, and w_k is the weight a participant may place on the k^{th} dimensions (and $k \in K$). The distance metric is thus computed as an euclidean distance between exemplars in memory and the test item weighted by where each feature is allowed to have a different weight to

reflect differentially important features. sij is the similarity between Test item i and Exemplar j which decreases exponentially with psychological distance. c_j is a scaling factor determining how much the similarity falls off for a unit of distance for each exemplar. a_{ij} is the activation of exemplar j when compared with test item i and is scaled by the memory strength m_j of the Exemplar j.

To demonstrate the potential role of temporal structure, a few simplifying assumptions are made to the recognition memory model. Specifically, in simulations presented below, it is assumed that each feature dimension of the studied (and test) items is weighted equally. This assumption is likely not valid for most realistic stimuli, however, the stimuli used in standard implicit boundary experiments (and those that will be used in the current work) are not meaningful and are randomly assigned to a node in the modular graph (Figure 2.1). Thus, any effect of the feature weights should be similar for boundary and non-boundary nodes. **Furthermore, in typical global matching models, recognition is said to be supported by context reinstatement at test.(Cox & Criss, 2020; Hicks & Starns, 2006; Osth & Dennis, 2020; Polyn et al., 2009). For the purposes of simulations, unlike these global matching models, the recognition model used here assumes no meaningful differences in context reinstatement at test.**. To simulate the differences between boundary and non-boundary nodes in memory, it is assumed that the memory strength of an item associated with each node is proportional to the entropy in its successor representation of that node. Formally,

$$\begin{aligned} \hat{M}_{i,j} &\leftarrow \hat{M}_{i,j} + \alpha[\delta(s_{t+1}, j) + \gamma * \hat{M}_{s_{t+1},j} - \hat{M}_{s_t,j}] \\ Entropy(s) &= \sum_{s' \in S} \hat{M}(s, s') * log(\hat{M}(s, s')) \\ m_s &\sim f(Entropy(s)) \end{aligned} \tag{3.2}$$

where $M_{i,j}$ of the matrix represents the expected future visits to state j from state i. $\delta(.,.)$ equates to 1 if both arguments are equal otherwise it equates to 0. Thus, the

matrix increases the probability of visiting a state j from state i if state j is visited in the current experience and it decreases the probability of visiting all other states from state i. Parameter α is a learning rate parameter that determines how much of the previous estimate of visiting state j from i is factored into the current update. Parameter γ is a future discount parameter that dictates how much in the future the agent sees – specifically, a higher value of γ indicates future visitations to state j are weighed high in the current update. $f(.)$ is a monotonic function.

While evidence for relating memory strength to context based entropy is scarce, past work has shown that entropy (as a measure of uncertainty) has been a helpful factor in motivated learning and is a contributing factor in hippocampal activation (Davis et al., 2012). Furthermore, the slow down associated with increased entropy as demonstrated in previous statistical learning tasks (Lynn & Bassett, 2020; Lynn, Kahn, et al., 2020; Lynn, Papadopoulos, et al., 2020) implies that participants at the least spend more time on such high-entropy boundary nodes, thereby allowing for a better chance of remembering these nodes better.

Given these assumptions, simulating recognition memory on the final SR representation provides an expected comparison of recognition memory accuracy for old boundary, old non-boundary and new items. Figure 3.1 shows the what this modeling approach yields. For the purposes of this simulation, 'stimuli' were assumed to be ten-dimensional and drawn from a uniform distribution bound between 0 and 1. Entropy was computed from the Successor Representation matrix derived by a 'structured' random walk of length 1000 (i.e. randomly choosing one of the connected node following the presentation of any given node for 1000 consecutive steps through the modular graph in Figure 3.2) and by an 'unstructured' random walk of length 1 (i.e. randomly choosing one of the fifteen nodes for 1000 steps and ignoring the graph structure). Parameters to learn the SR model were derived using the same

best fitting RSA procedure described in Chapter 2 [1]. This entropy was then translated to memory strength in the equation above assuming $f(x) = e^x$. 'Old' response probabilities were computed using Luce's choice rule (Luce, 1977) and choices were derived from a binomial distribution. This procedure was repeated for 100 hypothetical experiments with 15 randomly generated 'old' and 'new' items across a range of possible values of the scaling parameter c. As Figure 3.1 on average, implicitly operationalized boundaries stimuli associated with the nodes that lead into and out of a cluster of the graph in Figure 3.2 are expected to be remembered better than non-boundaries (although only for higher values of c). This benefit is expected to be apparent for participants who are exposed to the temporal structure (right panel, walk length of 1000) relative to participants who are not (left panel, walk length of 1).

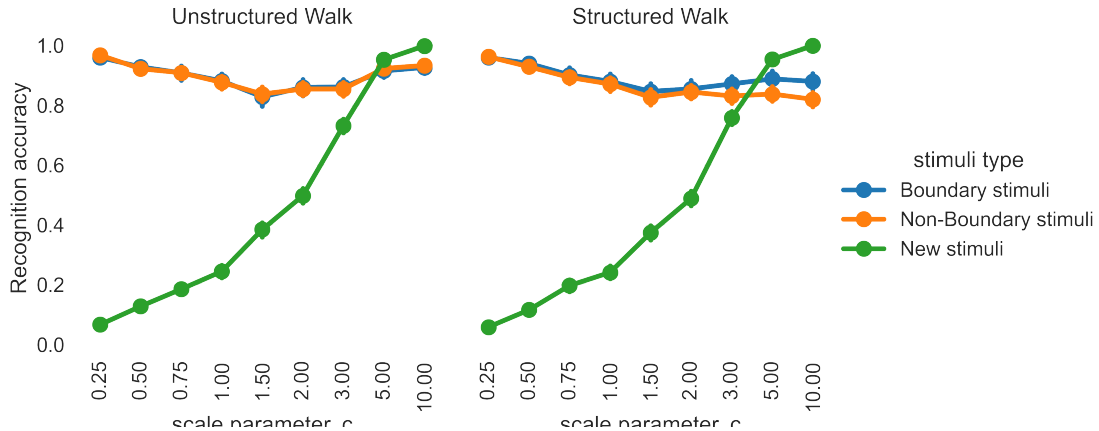


Figure 3.1. Simulated recognition memory test performances for walk lengths of 1 and walk lengths of 1000 on modular graph in Figure 2.1. On average, recognition memory performance is expected to be better for boundary items than non-boundary items.

[1]Note that since this procedure aims to minimize the distance between the learned matrix and the true transition matrix, it essentially aims to minimize the entropy at boundary nodes relative to non-boundary nodes, since objective entropies at boundary and non-boundary nodes are equivalent. These simulations thus provide a loose lower-bound of an effect.

3.2.1 Methods

Participants

63 undergraduate students at the University of Massachusetts Amherst participated in this study. Participants were at least 18 years of age and were compensated via course credit. All procedures were approved by the University Institutional Review Board. Data from 6 participants who did not complete the study were discarded from further analyses. Participant sample size was not pre-determined via a statistical procedure but was a rough equivalent of previous studies (Heusser et al., 2018). All statistical inference in this article is done as probabilities of effects measured through posterior parameter estimates in Bayesian models.

Stimuli

Randomly polygon shapes were used as stimuli for this experiment. Each polygon consisted of 6 vertices. 2 vertices were randomly placed around the center of the screen with their X-Y coordinates drawn from univariate normal distributions with a standard deviation of 0.1 inches. Coordinates of the remaining four vertices were drawn from univariate uniform distributions between 0.1 and 0.3 inches. 300 polygons were generated and randomly chosen 60 (15 'old' and 45 'new') were used for each participant.

Design and Procedures

Participants were randomly assigned to either a structured exposure or an unstructured exposure group. The overall experimental procedures were the same across both groups.

15 polygons were chosen randomly (from the set of 300) for each participant as 'old' stimuli. At the beginning of the experiment, participants were asked to carefully study these polygons and to remember their orientation. During 750 exposure trials, separated into 3 blocks of 250, participants were presented these polygons one at a

time and asked to judge whether the presented polygon was in its original orientation or rotated. Participants were provided feedback on their accuracy on each rotation judgment response and an on-screen score was maintained to motivate accurate responses. Polygons were surrounded by a (purple, orange, or dark green) colored border to use for a source memory test.

Each polygon with its border was associated with a node in the modular graph in figure 3.2. Order of polygons during exposure was determined by a participant's group. For 27 participants in the structured exposure group, the order of exposure was determined by a random walk through the modular graph (Figure 3.2) where each subsequent node was determined based on a random choice of the connected node. Exposure order for 30 participants in the unstructured group was determined by a random selection among the 15 items on each trial.

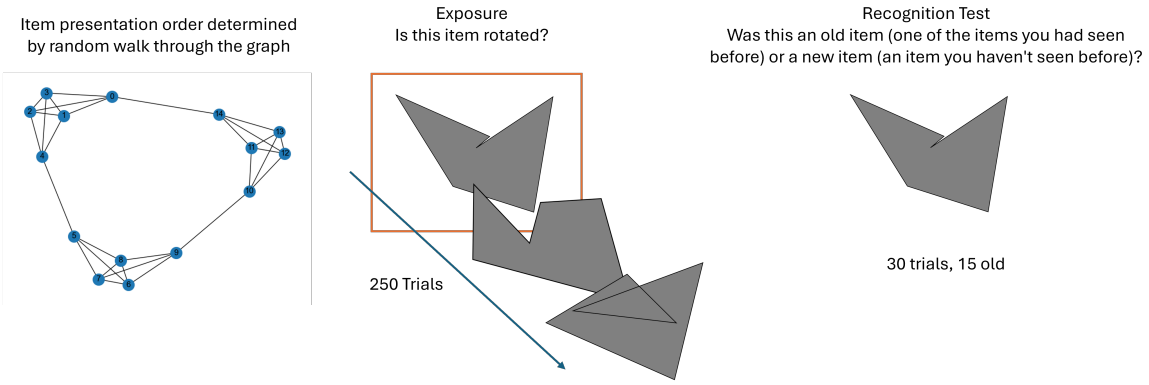


Figure 3.2. Design schematic for Experiment 2a. Three alternating blocks of exposure and recognition test were presented. A Stroop task was conducted prior to the final recognition test.

After each block of the exposure phase, participants went through a recognition memory test (Figure 3.2). They were shown the 15 'old' items from the exposure phase (in their original orientation) in addition to 15 new random polygon items (chosen from the set of 300) and were asked to determine whether each of these items was old or new. Order of presentation of old and new items was randomized

during the recognition memory test. Of the three recognition tests, the final test was conducted after a short Stroop task to washout any effects of short term memory.

After the final recognition memory test, participants went through a source memory test. Each of the 15 studied items were shown without the colored borders that surrounded them during exposure. Participants were asked to choose which of the three colored borders, provided as on-screen options, surrounded any particular item. This source memory task was added to provide an additional signal for memory in case of ceiling effects of recognition memory tasks. However, no analyses have been done on this source memory task.

3.2.2 Results

As expected, accuracies for old stimuli increase with increased exposure (Figure 3.3, Table B.2 for a statistical test of increased accuracy and decreased response times over blocks) whereas the response times decreased with more experience with the stimuli across blocks (Figure3.4, Table B.3 for a statistical test). See Table B.1 provides the means and standard deviations for response times and accuracies during the exposure and recognition phases across each block for all stimuli types for both conditions. Interestingly, overall accuracy of participants in the unstructured exposure condition is higher, than those in the structured exposure condition across all stimulus types. However, this effect apperas to be due to participant variability (See Table B.4 for statistical results of a hierarchical model accounting for a varying effect of participants.).

To assess differences between stimulus types (boundary vs non-boundary) on recognition memory, a signal detection (SDT) model was first fit separately for items associated with boundary and non-boundary nodes. The model for boundary items describes the probability of responding 'old' to old boundary stimuli relative to new stimuli. Similarly, the model for non-boundary items describes the probability of

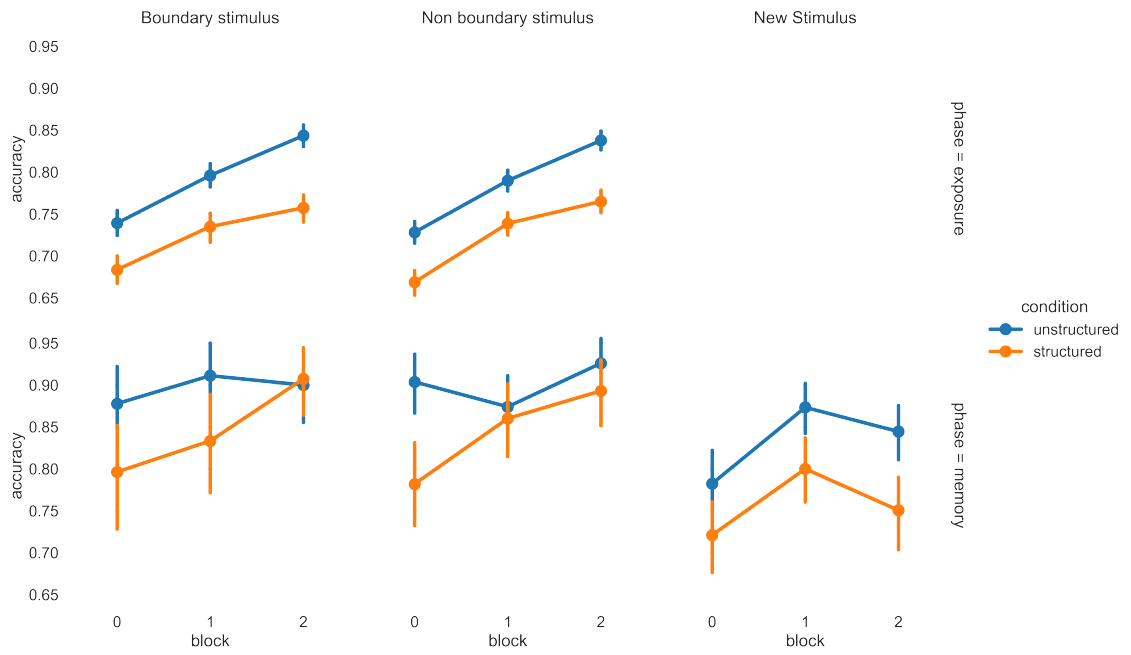


Figure 3.3. Mean accuracies for both participant groups (structured and unstructured) across blocks, for different stimulus types and phases of the experiment

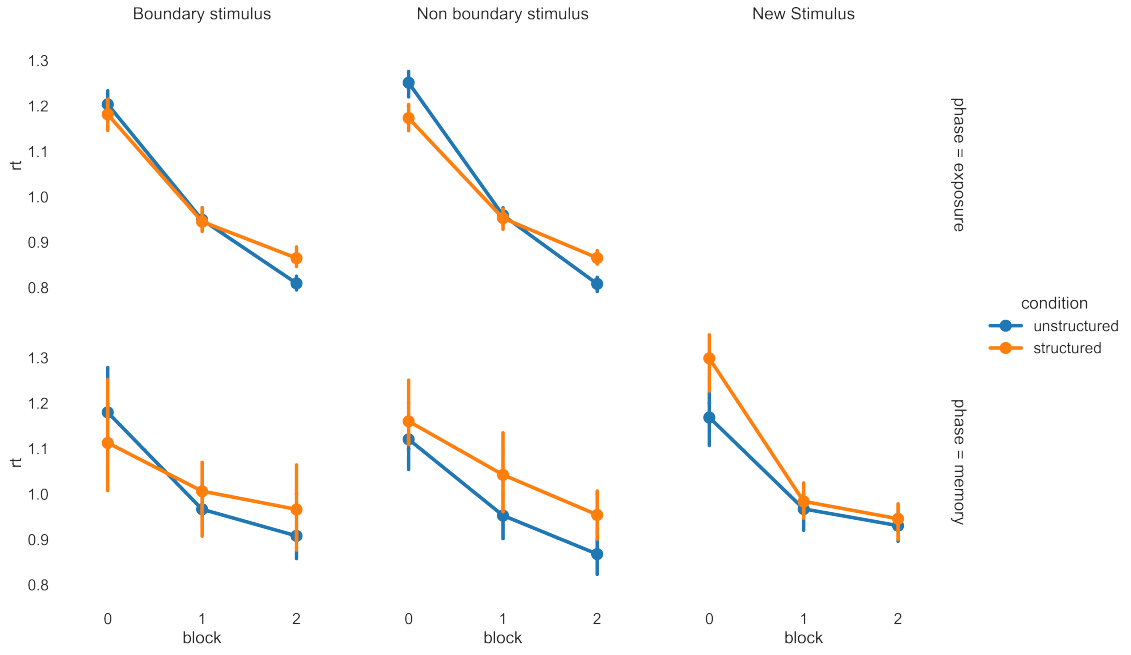


Figure 3.4. Median response times for both participant groups (structured and unstructured) across blocks, for different stimulus types and phases of the experiment

responding 'old' to old non-boundary items relative to new stimuli. A participant intercept term allows for each participant to have their own decision criterion. Finally, accuracy at exposure was included as a factor in the SDT model to account for differences in encoding accuracy. Exposure accuracy factor for old items was computed by averaging the rotation judgment accuracy for each of the old items in the block immediately before the recognition memory block. For new items, this factor was computed by averaging the rotation judgment accuracy for all exposure items in the exposure block before that recognition phase. The SDT model can be described as:

$$\begin{aligned} accuracy\ exposure &\sim \mathcal{N}(0, 13.87) \\ true\ old|condition &\sim Normal(mu : 0.0, Half\mathcal{N}(sigma : 5.5)) \\ participant\ criterion &\sim \mathcal{N}(0, Half\mathcal{N}(sigma : 11.5)) \\ \mu = (true\ old|condition) &+ accuracy\ exposure + participant\ criterion \\ p(resp\ old) &\sim Bernoulli(\mu) \end{aligned} \tag{3.3}$$

d', the coefficient for $true\ old|condition$ in the linear model above, measures the distance between distributions of old and new items. Parameter estimates of d' for structured relative to unstructured for boundary and non-boundary nodes for the final recognition block are shown in figure 3.5. Parameter statistics are reported in appendix tables B.5 and B.6

The SDT modeling implies that while there are no differences in recognition memory for boundary nodes based on exposure (60% of posterior samples of the difference between structured an unstructured conditions below 0), non-boundary nodes seem to become less recognizable under structured exposure condition (97% of the posterior samples of the difference between structured and unstructured conditions are below 0).

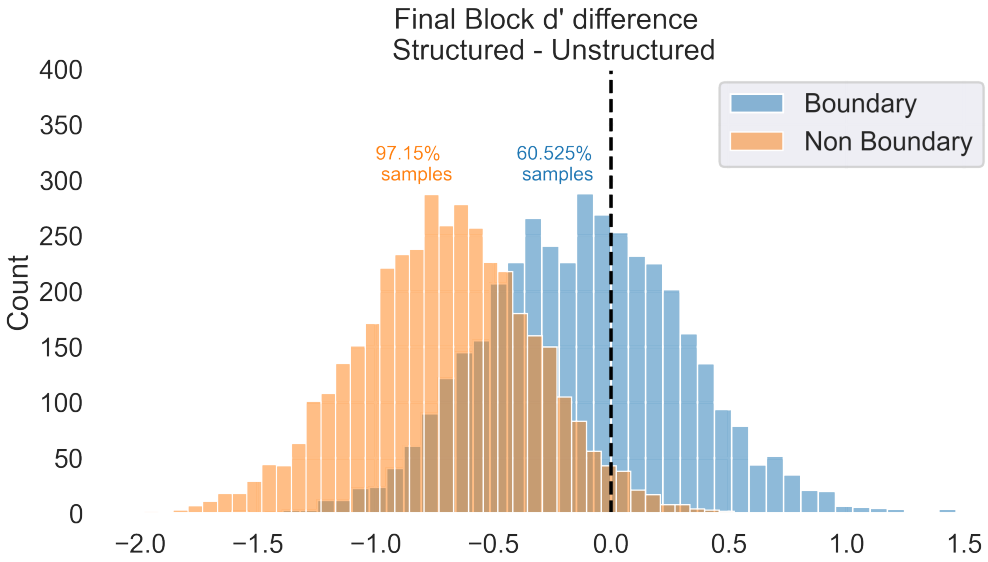


Figure 3.5. Differences in d' for models fit to separately boundary and non-boundary nodes for both structured and unstructured exposure conditions

Diffusion Modeling

The SDT model used above fails to account for ceiling effects – accuracy for old items is near perfect or could have reached an asymptote. The SDT model was also fit separately to derive d' for boundary and non-boundary items, thus losing shared variability within participants.

To be able to account for ceiling effects in recognition accuracy, we can use additional information available in the form of response times during the recognition memory task. For example, for participants equally accurate in recognizing boundary and non-boundary participants, being able to recognize boundary items faster may provide additional evidence for better memorability of these items relative to slower recognized non-boundary items. Figure 3.4 shows median response times across three blocks of recognition test.

To understand recognition memory differences between the boundary and non-boundary items in context of response accuracy and response time distributions, we use the Drift Diffusion Model (DDM, Figure 3.6). The DDM, which falls under a

class of sequential sampling models, has been a widely successful model in modeling two-choice tasks in recognition memory (Ratcliff & Starns, 2009; Ratcliff et al., 2004, 2022; Starns, 2014; Starns & Ratcliff, 2014).

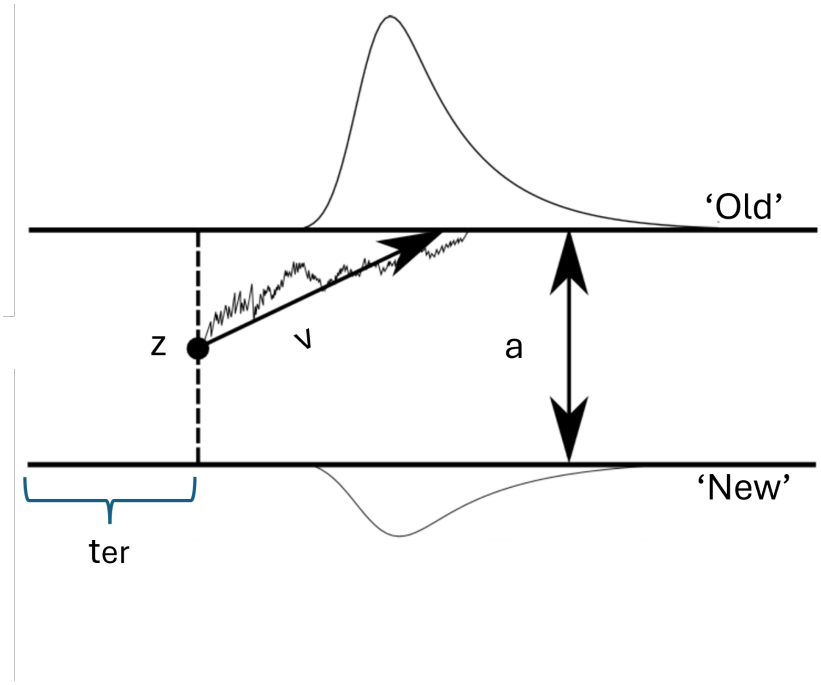


Figure 3.6. The Drift Diffusion Model of Choice Response Times for Old/New recognition memory tasks.

Briefly, the DDM assumes that at choice time, evidence from two presented options accumulates sequentially over time towards one of the two boundaries. For a previously studied item presented at test, the evidence from the item accumulates slowly towards the 'old' boundary whereas for a non-studied item, the evidence accumulates towards the 'new' boundary. The rate of evidence accumulation is controlled by the drift rate parameter, v. Participants may be biased towards making an old or a new response at test; this bias is measured by the starting point parameter, z. The boundary separation between the two responses is modeled by a parameter a. Finally, the observed response consists of cognitive processes not affiliated with decision making (such as time it takes to visually process the test item, time for the

motor systems to click the relevant key) which are modeled by a non-decision time parameter t_{er}. [2]

Prior work has shown that memory strength of previously studied items impacts the drift rate towards old/new responses (Ratcliff et al., 2004, 2022). A higher drift rate parameter implies a stronger match to memory which leads to a quicker accumulation of evidence towards the 'old' response boundary. Similarly, a stronger mis-match to memory (as measured by the higher drift rate parameter) allows for a quicker accumulation of evidence towards the 'new' response boundary (Ratcliff et al., 2004, 2022).

The DDM thus allows us to circumvent ceiling effects by modeling response time distributions (as faster accurate trials may reflect better memory than slower accurate trials) and estimate whether boundary items are indeed remembered better than non-boundary items in the structured exposure or whether the effect is driven by worse-remembered non-boundary items (as faster accurate non-boundary stimuli recall may reflect better memory than slower accurate non-boundary stimuli).

For the recognition task in the current study, the DDM was parameterized as follows:

$$\begin{aligned} v &\sim 0 + nodetype : condition : block + accuracy\ exposure \\ z &\sim 0 + block \\ a &\sim 0 + condition : block \\ t_{er} &= 0.25 \end{aligned} \tag{3.4}$$

The fixed value of the non decision time parameter t was derived by first fitting the DDM over a range of possible t values (i.e. a grid search) and picking value with

[2]Note that this version of the DDM is a simplified model. More complex DDMs account for trial-to-trial variability in each parameters as well.

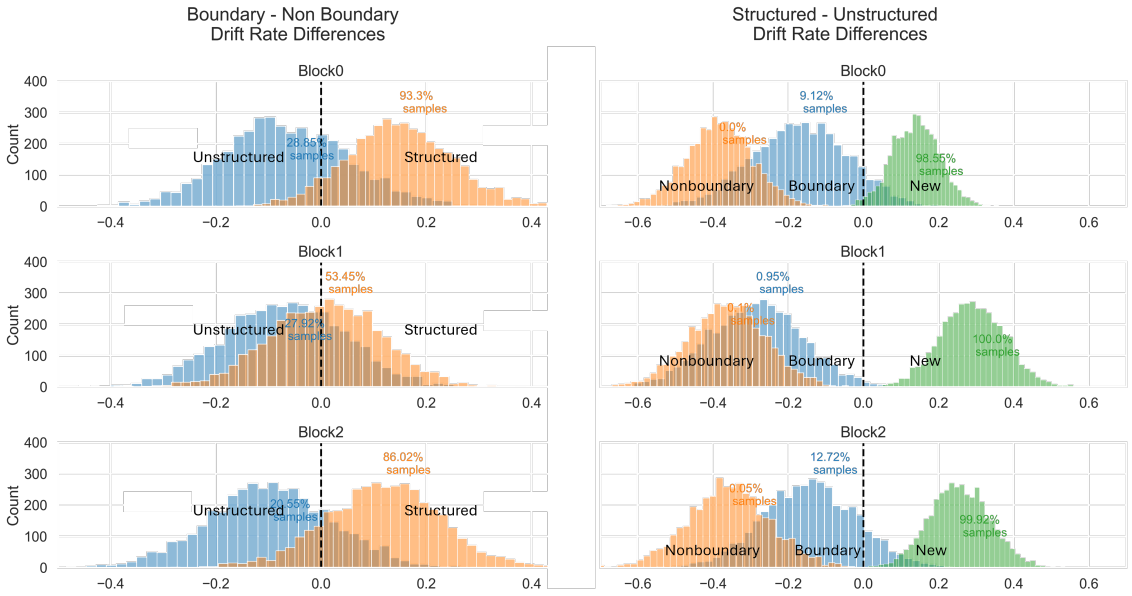


Figure 3.7. Drift rate differences. *Left Panel.* Differences between boundary and non-boundary nodes for structured and unstructured exposure conditions. *Right Panel* Differences between drift rates between structured and unstructured conditions for each type of recognition memory stimulus. Figure text over each difference distribution depicts the proportion of posterior samples above 0.

the best fitting model in that range. DDM models were fit using the HSSM package (Fengler et al., 2022). [3]

DDM Results The modeling framework above allows incorporating response times during recognition memory tasks. In particular, if an item is quickly and accurately recognized as old or new, in addition to good accuracy, such recognition would result into faster reaction times. This effect is captured by the drift rate parameter of the DDM.

Figure 3.7 shows the differences between drift rate parameters. The modeling framework introduced earlier used SR to derive an estimate of entropy for each stimulus depending on its boundary or non-boundary role in the modular graph of Figure

[3]Unfortunately the non decision time parameter is too difficult to fit – this is a known problem with the HSSM package

3.2. Node entropy further was assumed to enhance memory strength of the stimulus associated with that node such that, on average, stimuli that are associated with boundary nodes are expected to be remembered better than those associated with non-boundary nodes for participants who are exposed to a structured random walk. No such difference is expected for participants exposed to an unstructured walk. We thus expect that the drift rate, as proxy for memory strength, would be higher for boundary nodes for structured random walk participants than for non-boundary nodes. This difference in boundary vs non-boundary item drift rates should be relatively higher for participants exposed to the structured condition than for the participants in our control condition; those exposed to the unstructured walk.

The left panel in Figure 3.7 shows that, as expected, boundary nodes have a higher drift rate than non-boundary nodes in the structured conditions relative to the unstructured conditions. This difference is more apparent in the first and the final block with 93.3% and 86.02% of posterior samples of the differences in drift rate being above 0 for structured condition whereas 28.85% and 20.55% of samples above 0 for unstructured condition. This difference disappears for the middle block (53.45% samples for the structured condition and 27.92% for unstructured) likely due to recognition test presented immediately after exposure and at this point participants have been exposed to the old stimuli for 500 trials. The difference likely reappears in the final block due to the Stroop distractor task administered before recognition test.

The right panel shows the same effect within conditions. New items appear to have better drift rates than old items in the structured condition than the unstructured conditions across all blocks. While drift rates for 'old' items were generally lower for the structured condition (relative to the unstructured condition), they were higher for the boundary nodes than non-boundary nodes indicating that even when stimuli at boundary nodes are remembered worse in structured condition than those at boundary nodes in the unstructured condition, they are still remembered better

than the stimuli at non-boundary nodes in the structured condition. See Table B.7 for full parameter statistics of the DDM.

Interim Conclusion

Findings in the first experiment in the current work show that similar to explicitly operationalized boundaries, implicitly operationalized boundaries are remembered better than non-boundaries. Furthermore, these improved boundary memory effects can be supported by an SR-derived entropy formulation. Stimuli associated with boundary nodes carry more information about the graph structure than those at non-boundary nodes. This additional information leads to slower reaction times (Chapter 2) and improved memory for those stimuli.

3.3 Experiment 2b: Boundary Distance Effects

Modeling Boundary distance effects

Another replicated finding in explicitly operationalized event boundary literature is an apparent increased separation of events across boundaries (Brunec et al., 2018; DuBrow & Davachi, 2013; Ezzyat & Davachi, 2011; Heusser et al., 2018; Horner et al., 2016). This separation of events across boundaries helps shape narratives in long term memory (Clewett et al., 2019).

It is unknown, however, whether the increased temporal separation across event boundaries generalizes to boundaries operationalized implicitly as well. Context models such as SR provide a mechanism to directly estimate perceived distance between events. Specifically, each cell in the SR matrix $M(s, s')$ indicates the future expected visit probabilities from state s to state s'. Under the assumption that states closer to the current state are visited more often in a random walk than state farther, the probability $M(s, s')$ provides a direct estimate of probability of seeing node s from

node s' and an hence an indirect estimate of the perceived distance from node s to s'.

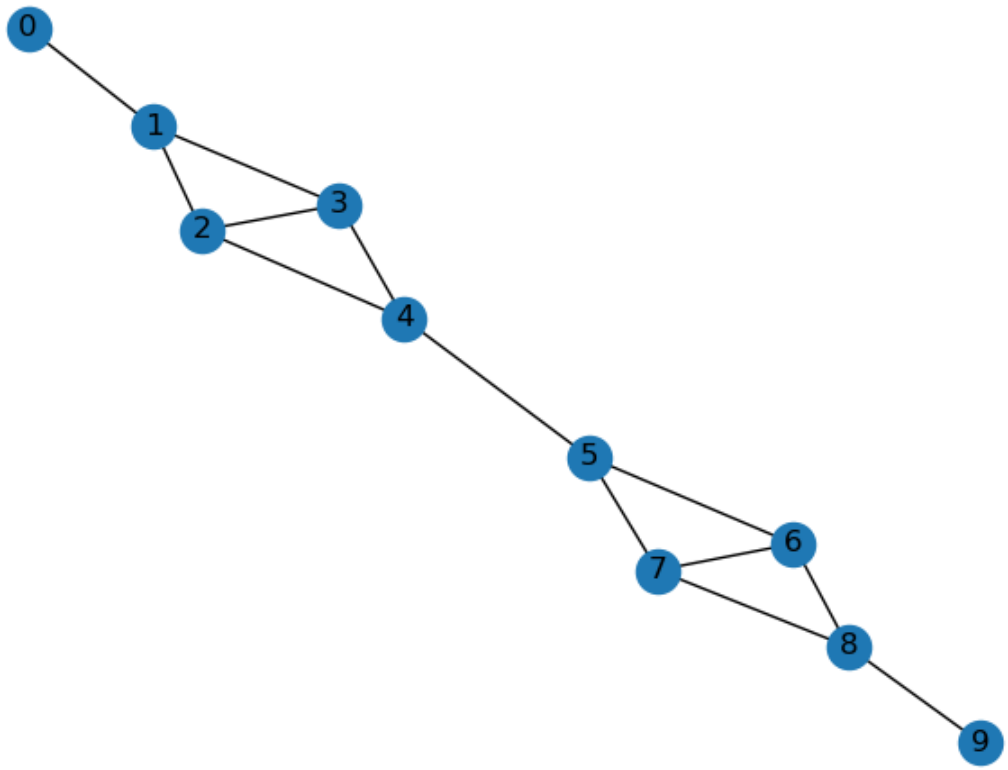


Figure 3.8. Two module graph used for distance judgments.

To test whether cross boundary events are perceived to be farther from each other relative to within-boundary events graph in Figure 3.8 is used. This modular graph provides some desirable properties for a distance judgment task such as fewer stimuli to remember, and remote nodes along with non-remote, non-boundary nodes with same number of connections. The key transitions of interest that are compared in this graph are transitions at equal distances (distance defined by the number of connections needed to be traversed to reach a node from another node). Below I specify example transitions at 3 different distances however, for simulations and the experiment, all symmetrical transitions at those distances were tested.

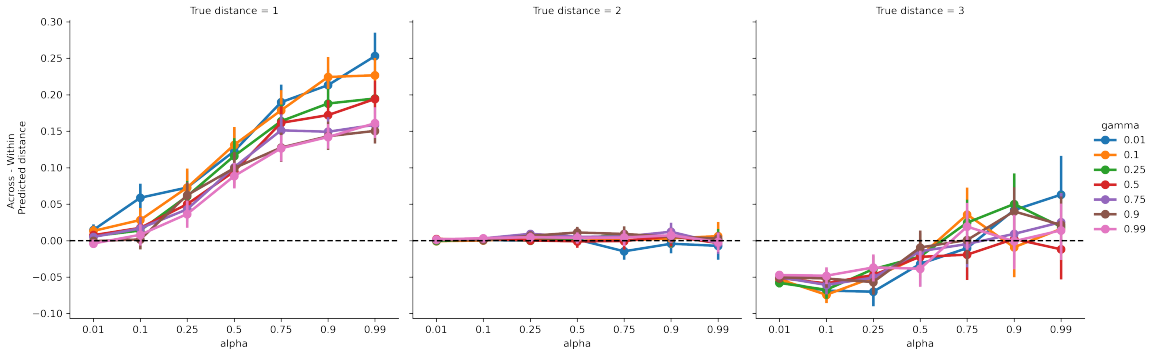


Figure 3.9. SR predictions of distances across boundaries relative to distances within boundaries for nodes at true distance of 1, 2, and 3 and different parameter combinations.

- $0 \leftrightarrow 4$ vs $1 \leftrightarrow 5$ at distance 3.
- $1 \leftrightarrow 4$ vs $2 \leftrightarrow 5$ at distance 2.
- $2 \leftrightarrow 4$ vs $5 \leftrightarrow 4$ at distance 1.

SR estimate of distances is shown in Figure 3.9. Simulations predict that while boundary nodes themselves get perceptually farther with increased discount rate, neither of the node-pairs at distance 2 or 3 that involve boundary nodes reliably show an increased cross-cluster distance. For remote nodes (which are involved in nodes at distance 3), some parameter combinations may lead to cross cluster transitions being perceived as closer.

This SR framework predicts that neighboring boundary nodes themselves appear farther from each other relative to nodes within a cluster from that cluster's boundary node. In fact, the SR also makes a stronger prediction that for no combination of parameters, the neighboring boundary nodes should be perceived closer to each other than a boundary node and its non-boundary neighbor. This prediction of the model is in line with past findings in explicit event boundary literature where events across boundaries are perceived to be farther from each other than event within the boundaries (Ezzyat & Davachi, 2011; Heusser et al., 2018).

Model predictions at other distances are mixed and dependent on parameters that allow learning of the SR. Nevertheless, the SR predicts that stimuli associated with cross cluster nodes at distance 2 should be neither perceived as closer nor farther relative to stimuli associated with within cluster nodes at distance 2. On the other hand, cross cluster nodes at distance 3 may be perceived closer or farther from each other relative to within-cluster nodes at distance 3 depending on the parameter of the SR model. These predictions are in contrast with findings in explicitly operationalized event boundary literature where stimuli presented across boundary events are perceived to be farther from each other. (Ezzyat & Davachi, 2011; Heusser et al., 2018).

The next experiment thus tests 1) whether the typical finding of increased perceived separation between cross cluster nodes (relative to within cluster nodes) is replicated in implicitly operationalized event boundary paradigms and 2) Whether SR continues to be a reasonable framework to understand the representations of implicit event boundaries. An increased cross-cluster distance observed for nodes at distances of 2 or a decreased cross-cluster distance for nodes at true distance of 1 will serve as evidence *against* the current formulation of the SR model's role in estimating temporal distances.

3.3.1 Methods

Participants

48 undergraduate students at the University of Massachusetts Amherst participated in this study. Participants were at least 18 years of age and were compensated via course credit. All study procedures were approved by the University Institutional Review Board. Data from 3 participants who did not complete the study were discarded from further analyses. No apriori statistical power analyses was computed to

estimate sample sizes. Sample sizes were determined by rough equivalence from prior studies (DuBrow & Davachi, 2013; Heusser et al., 2018).

Stimuli

The same polygon stimuli used in the previous experiment were used for this experiment as well. For each participant, 12 random polygon stimuli wer chosen from the set of 300.

Design and Procedures

Participants were randomly assigned to either a structured exposure or an unstructured exposure group. The overall experimental procedures were the same across both groups. The experiment consisted of 2 phases, an 'exposure' phase where participants made judgments about the orientation of the polygons and a 'distance judgment' phase where participants went through a choice task making judgments about relative distance of the stimuli they saw during the exposure phase.

Participants were first introduced to 10 randomly generated polygons and informed that these polygons are in their canonical orientation. Participants were asked to study these carefully and remember their orientations as they will make judgments about the orientation of these polygons in the coming phase. During the exposure phase, participants were shown one polygon at a time from the set of 10 they were introduced to. On each trial, the polygon was either rotated by 90 degrees or shown in its canonical orientation. Participants were asked to judge whether the polygon is rotated or not.

The order of trials in exposure phase was determined based on a participant's group. 27 participants in the structured exposure condition were exposed to a stimulus stream generated by a random walk through the graph in figure 3.8. 18 participants in the unstructured exposure condition were shown a stimulus stream generated by randomly choosing any of the 10 polygons to be shown on any trial. The exposure

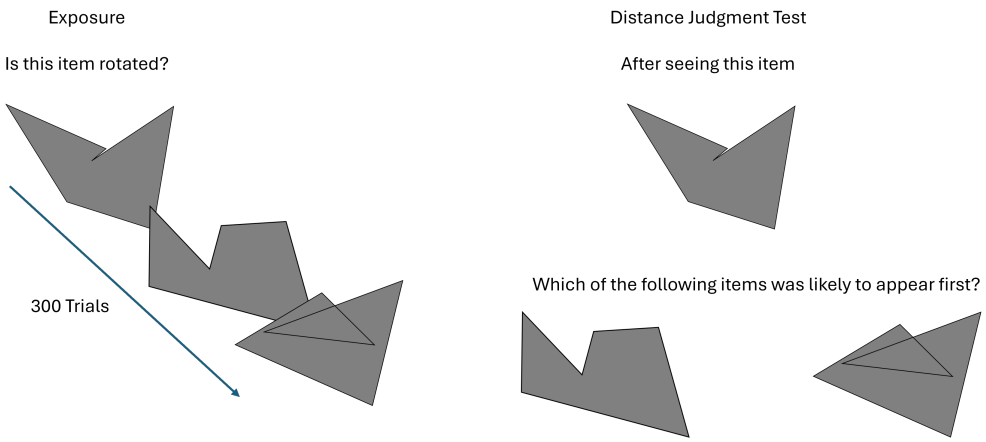


Figure 3.10. Design schematic for experiment 2b. After exposure through the graph structured (based on a random walk through the connected nodes or a random selection between all 15 nodes), participants went through a distance judgment phase

phase lasted for 300 trials or 30 minutes, whichever came first. Participants were provided with an opportunity to take a self-paced break after 150 trials.

During the test phase, participants were shown a triplet of polygons (see Figure 3.10). For each top polygon, participants were asked which of the bottom polygons was likely to appear first after seeing the polygon at the top in the stream they had experienced during exposure. The distance judgment phase lasted for 20 trials where 10 trials consisted of the critical pairs (examples listed in the previous section) and 10 filler trials were based on randomly generated (non repeated) triplets. Responses to these filler trials were not analyzed.

3.3.2 Results

Figure 3.11 provides an overview of the proportion of choices participants made to indicate a within cluster item is closer to the top item than a between cluster item. Descriptive statistics in Table 3.1. For all conditions, response probabilities were largely at chance indicating that within-cluster option did not appear closer to most participants relative to the across-cluster option regardless of the true distance.

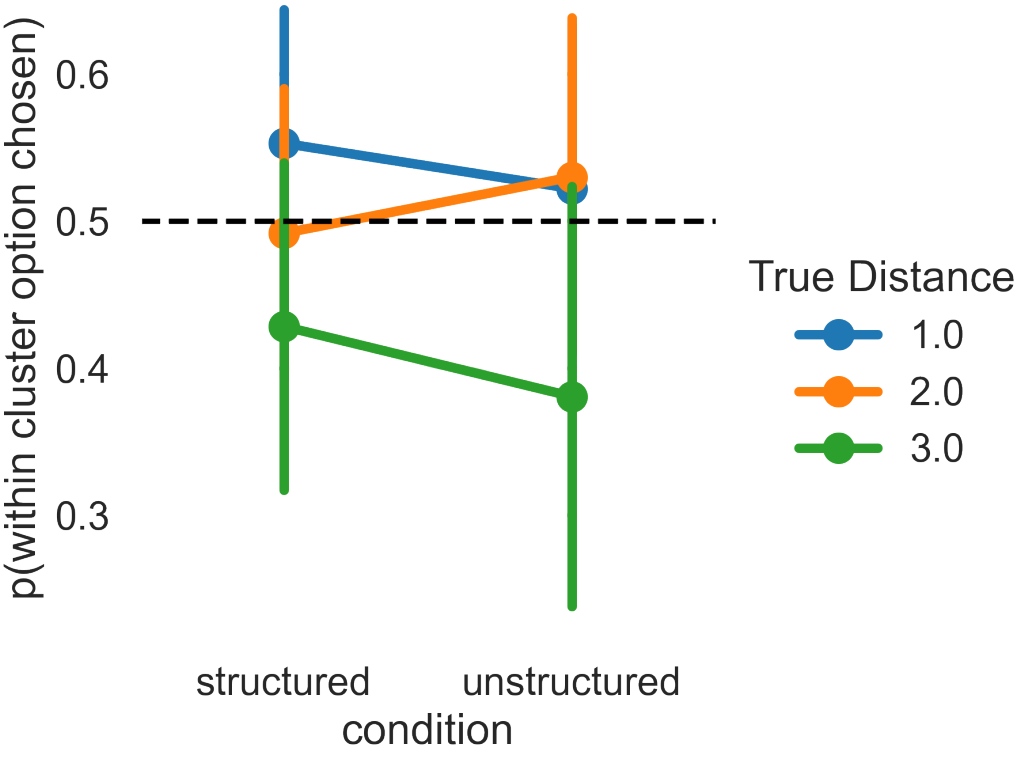


Figure 3.11. Proportion of trials where the within cluster option was chosen when the distance between and within clusters were equal (ranging from a distance of 1, 2, and 3 connections).

condition	true distance	within chosen mean	std
structured	1.0	0.565	0.497
	2.0	0.505	0.501
	3.0	0.450	0.500
unstructured	1.0	0.528	0.501
	2.0	0.533	0.501
	3.0	0.521	0.503

Table 3.1. Proportions of trials where within cluster option at the same distance as the between cluster option was chosen.

To assess this statistically, a hierarchical Bayesian Model was fit the within cluster choice probability as follows:

$$p(\textit{within option chosen})\ 0 + \textit{true distance} : \textit{condition} + (1|\textit{participant}) \tag{3.5}$$

Figure 3.12 provides a bayesian estimate of the difference in proportion of within cluster option chosen relative to the between cluster option when participants are exposed to a structured, random walk presentation order compared to when they were exposed to unstructured presentation order. For all distances, there is no apparent difference between these proportions in either direction. The within cluster option was chosen more often in the structured exposure relative to unstructured exposure with 65.53% probability for true distance of 1, 31.92% probability for true distance of true and 66.38% probability for a true distance of 3. Since 95% HDIs for all True distances include 0, the structured exposure did not lead to higher selection of within cluster stimuli as being closer.

3.4 Discussion

The primary goal of this chapter was to test whether findings in classical explicitly operationalized event boundary literature replicate when boundaries are operational-

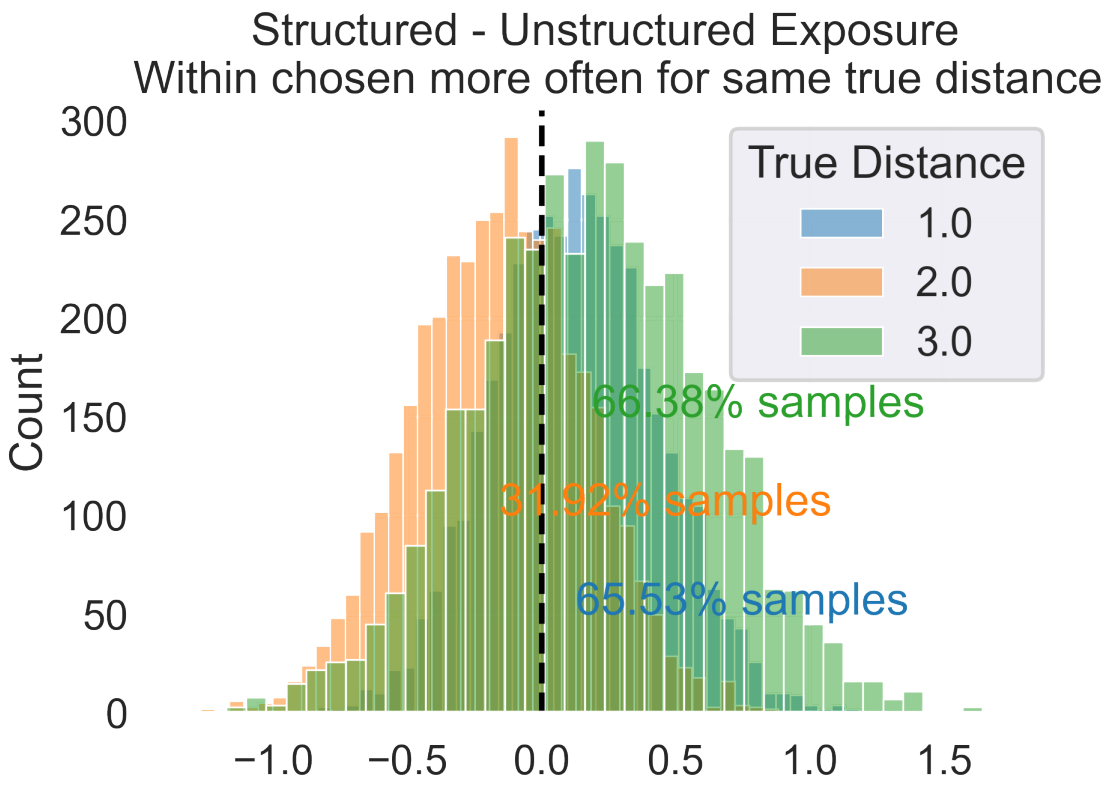


Figure 3.12. Posterior estimates of the differences between structured and unstructured exposure condition for the proportion of times when the option within cluster was chosen more often as being closer than the option across the cluster at the same shortest distance.

ized implicitly, through temporal statistics. Shared properties between these two boundaries imply shared representations in memory thereby providing a framework for future research to study shared algorithmic processes that lead to formation of event boundaries.

In Experiment 2a, I test whether similar to findings in the explicitly operationalized event boundary literature (where boundary events are perceptually different from the ongoing stream of events), participants remember implicitly operationalized boundar events (where boundary events are not perceptually different but serve as a gateway to a different cluster defined by temporal co-occurences). I propose an SR based framework to model potential differences in memory between implicit boundary and non-boundary events. The SR based framework predicts that since boundary nodes carry higher information about the structure of the environment they will be remembered better than the non-boundary nodes. Findings of Experiment 2a support this prediction. Specifically, implicit boundary nodes have a higher drift rate than non-boundary nodes.

DDM is one (highly successful) model of memory that incorporates response time distributions along with choices. Several other sequential sampling models can be used in this formulation to provide evidence of better memory strength for boundary items. Most notably, a direct extension of the GCM model for recognition memory is the Exemplar Based Random Walk (EBRW) model (Nosofsky et al., 2011) which assumes that evidence towards old/new recognition choice options is accumulated made in proportion to background activation of all items in memory may be extended to incorporate node entropies in its evidence accumulation process. For example, higher entropy (typical for stimuli at boundary nodes in structured random walks relative to non-boundary nodes) may lead to increased activation of boundary stimuli relative to background activation thereby leading to better recognition. Further work

is needed to understand how extensions to other choice reaction time models can be extended to incorporate effects of temporal order.

The effect of structured exposure on identifying new stimuli is intriguing (Figure 3.7). Higher drift rates for new items in the structured condition relative to the unstructured condition may imply support for event integration at boundary nodes (Zacks & Swallow, 2007). Structure may allow extraction of higher order knowledge (segmented at boundary stimuli) and hence makes it easier to identify events that do not belong to that structure. Such SR framework may help future investigation should aim to diagnose the underlying cognitive processes that lead to recognition of new stimuli.

Finally, model simulations in Experiment 2a have followed from simplifying assumptions about mental processes at test. In particular, the assumption of a stable context for all test items warrants a mention. In most context-based recognition memory models, and especially global matching models, items at test are often said to reinstate the context in which they were studied (Cox & Criss, 2020; Hicks & Starns, 2006; Murdock, 1997; Osth & Dennis, 2020; Polyn et al., 2009). This assumption is relaxed for recognition memory simulations for Experiment 2b, thereby implicitly assuming that context reinstatement does not play a role in the boundary memory effect. While the comparison between unstructured and structured exposure across participants and between boundary and non-boundary stimuli within participants provides support for an internal validity of this assumption, future research should aim to account for context reinstatements at test items for similar implicit event boundary models. Indeed, event cognition literature suggests explicit event boundaries serve as points of replay (Hahamy et al., 2023) and points at which memory is scanned (Michelmann et al., 2023). It is likely that recall of

boundary events reinstates surrounding context and future work should thus investigate the role of SR-based context reinstatement on recall performance..

Similarly, the model simulations also assume that the effect of boundary entropy is on the memory strength parameter. As stated above, it is possible that boundaries are not necessarily remembered better but reinstatement of context, which the boundaries represent better, is what leads to improved recall. Future work should also investigate the validity of the assumption of memory strength being impacted by boundary node. Regardless of why boundary nodes are remembered better, the fact that they are as shown in the current work, is an important step in deriving common representations between implicit and explicit event boundaries.

In the Experiment 2b, I aim to test 1) Whether similar to explicitly operationalized boundaries, implicit boundaries also temporally stretch events in memory, and 2) Whether the SR framework continues to be a reliable model of representation of implicit boundaries. The SR model allows us to simultaneously test both these hypothesis for a graph structure with two modules and remote nodes (Figure 3.8). Specifically, for across-boundary nodes at distance 2 or 3, the model predicts no reliable increase in selection of within boundary nodes as being closer. On the other hand, within boundary nodes should be selected more often as being closer than across boundary nodes for nodes at distance of 1. Furthermore, the model predicts that for nodes at distances 1 or 2, across boundary nodes should *not* be chosen as being closer; thereby providing an avenue for the SR model to be falsified. The model further predicts that for some parameter values, cross cluster nodes may be chosen as closer than within cluster remote nodes, a prediction that suggests that implicitly operationalized event boundaries may not follow similar behavioral properties as explicitly operationalized boundaries.

Findings in Experiment 3b suggest that implicitly operationalized event boundaries may not share properties with explicitly operationalized event boundaries – participants do not reliably select the within cluster stimuli to be closer relative to across cluster stimuli at the same distance. Furthermore, the lack of this within-across cluster effect for nodes at distance 1 indicates lack of support for the SR model framework that's used for predictions. Nevertheless, since for distances 1 and 2, across cluster stimuli were *not* chosen more often than within cluster stimuli, the SR model was not falsified.

The lack of support for model predictions in experimental findings provides an additional avenue for further testing of this SR framework. While SR allows for small effects at distance 1 for lower learning rates and discount parameters, the lack of the observed effect may simply reflect a harder-to-detect small effect. The model can be further falsified through experimental designs that allow for varying idiosynchratic learning and discount rate parameters. For example, providing participants information about an existing temporal structure may lead to an increased learning rate or discount parameters. Lack of increased selection of within cluster option as being closer in such a case would provide a stronger evidence against the SR framework. Finally, future work should also incorporate other context models such as the TCM (Howard et al., 2005; Polyn et al., 2009) to directly compare them with the current SR framework (Gershman et al., 2012) their predictions on such distance judgment tasks.

Finally, both experiments used specific graph structures to investigate possible shared properties of implicit and explicit event boundaries. While the SR model makes a theoretical prediction of similar findings across topological variations, future work should also attempt to identify if altering the properties of various graph topologies can lead to different experimental observations.

3.5 Conclusion

In this chapter, I aimed to assess whether implicitly operationalized boundaries lead to the same behavioral properties as explicitly operationalized boundaries. Implicitly operationalized boundaries are indeed remembered better. However, evidence from the distance judgment is mixed. Future experimental paradigms should investigate whether more regular graphs (without remote nodes, for example) lead to the increased cross boundary distance.

CHAPTER 4

CATEGORY LEARNING THROUGH TEMPORAL ABSTRACTION

4.1 Introduction

We naturally categorize items we encounter daily for ease of storage, processing, and decision-making. For example we know instinctively that regardless of shape and form, all lamps form a 'lamp' category based on its function. Depending on the complexity of rules that determine categories, some categorizations are easier than others (Nosofsky et al., 1994; Shepard et al., 1961). Category variability can modulate how often exemplars are classified into that categories (A. L. Cohen et al., 2001). The study of categorization in cognition has largely focused on explicit category learning where feedback or instructions are provided to participants along with category labels (Nosofsky et al., 1994; Shepard et al., 1961). However, in most daily encounters, a majority of categorization experience is unsupervised and automatic. In this work, I aim to assess the psychological processes underlying implicit categorization – done without knowledge of an underlying category structure but through temporal contingencies. In particular, I focus on how the (implicit) temporal order in which category exemplars are learned impacts attention towards features which define a category.

The order of presentation items in category learning tasks has been shown to be an important factor in how category diagnostic features are learned. In particular, when items are presented in a blocked categorical design, participants seem to learn the similarities between the same category items. On the other hand, when items are presented as an interleaved design, participants seem to focus more on learning the features that differentiate the underlying categories (Carvalho & Goldstone,

2017). As a result of order-dependent differing focus on category diagnostic features, interleaved presentations seem to benefit general category learning. In most prior category-learning tasks assessing order of presentation effects, participants are explicitly asked to learn the underlying categories and given explicit feedback. There appear to be clear differences when participants focus on learning categories based on how exemplars of these categories are presented (Carvalho & Goldstone, 2014, 2017; Kornell & Bjork, 2008; Kornell et al., 2010; Vlach et al., 2008; Whitehead et al., 2021). In this article, we investigate the effects of order of presentation when category learning is implicit.

One primary focus on category learning through order of presentation is comparing blocked or interleaved exemplar presentations. For example, Kornell and Bjork, 2008 showed participants paintings made by two different painters. The order of presentation during exposure was modulated to either be blocked (paintings of one artist shown together followed by the second artist) or interleaved (paintings made by both artists were mixed). When presented with new paintings, and asked which of the two studied artists made them, participants who were exposed to the interleaved format were found to be more accurate at guessing the creator. Category learning also improved for interleaved presentation compared to blocked presentation when tested on items where relevant category features were visually occluded (Whitehead et al., 2021). When three-year-old children are tested on the generalization of category-specific features, they appear to benefit from the spaced study of exemplars as compared to a blocked (Vlach et al., 2008). By modulating the similarity of presented items, interleaved presentation was found to be better than blocked presentation design on generalization performance particularly when learned exemplars were more visual (Carvalho & Goldstone, 2014; Kornell & Bjork, 2008).

Interleaved presentation has been theorized to improve in category induction because of context-based variability during encoding (Glenberg, 1979). Particularly, for

each presented item, an observer will store both the item-specific features along with the context in which the item is encoded. During interleaved presentation, a category diagnostic feature gets encoded under different contexts. Thus, that diagnostic feature will be recalled when tested on novel category items within that context.

Two theories have been proposed to explain this discriminability-based advantage of interleaving. According to the attention attenuation account, when categories are blocked, participants may think that they have learned the relevant category features after viewing a few items and stop paying attention to additional exemplars of the (Kornell et al., 2010). On the other hand, according to the discrimination account, the interleaved presentation allows participants to directly compare the differences between exemplars of different categories that are presented close to each other thereby highlighting these differences (Kornell & Bjork, 2008). In a direct test Wahlheim et al., 2011 found that when participants were shown pairs of exemplars, each belonging to a different category, the interleaving benefit was magnified compared to when they were presented as single items. The authors posit that showing pairs of exemplars would enable participants to carefully study and infer distinctions between category features and hence improve categorization performance .Furthermore, the authors find evidence against the attention attenuation theory by observing that classification performance did not differ as a function of the position in which the exemplar was presented in a stream.

This benefit of interleaved presentation is shown to be modulated by the 'level' at which categorization occurs. For example, when Mack and Palmeri, 2015 modulated exposure time to individual exemplars along with order of exposure, they found that interleaved presentation was no longer beneficial under short exposure conditions particularly when participants were asked to make a more abstract, 'super-ordinate level categorization. On the other hand, when exemplars were presented in a blocked format, a lower, 'basic' level categorization was hindered. Thus, category knowl-

edge through order of presentation can be modulated by the level of categorization participants are asked to produce.

It is clear that order of presentation of categories matters during explicit category learning. However, the effect of such order of presentation has not been investigated when category learning is implicit. Indeed recent work shows that participants do acquire category knowledge that when presented implicitly instead of being explicitly asked to learn categories. Unger and Sloutsky, 2022 found that assessed on category knowledge, participants appeared to learn category structures without being explicitly instructed to do so. This category knowledge was modulated by the strength of association of the category diagnostic features. Unger et al., 2023 later found that when presented with implicit feature-based categories during a cover task, participants were sensitive towards category diagnostic knowledge.

More evidence for incidental category learning comes from auditory cognition. Gabay et al., 2015 found that participants were sensitive to audio categories learned implicitly as measured by increased reaction times when audio-category-to-response mapping was altered. Incidental category knowledge is further modulated by the sampling category distributions from which exemplars are drawn. Roark and Holt, 2018 show that probabilistic sampling of exemplars leads to weaker category learning compared to deterministic sampling. Incidental category learning can be further enhanced by task-relevant and disrupted by task-irrelevant feature-to-category mappings (Roark et al., 2022). Incidental learning may also be disadvantaged compared to supervised intentional learning when categories are non-linearly separable (Love, 2002).

More incidental category learning has been studied under the 'unsupervised' category learning. Participants could infer rules based on correlating features without being explicitly asked to categorize during exposure (Billman & Knutson, 1996). Category-related items were recognized better when presented close to each other

then when category-unrelated items were (Medin & Bettger, 1994), with people being able to sort stimuli by a single dimension (Medin et al., 1987). Models of categorization have attempted to explore mechanisms of such unsupervised category learning. SUSTAIN (Love et al., 2004) seems to explain several of these unsupervised category learning phenomena using a distributed representation whereas ALCOVE (Kruschke, 2020) further provides for error based diagnostic feature attention learning in GCM (Nosofsky, 1986, 2011).

While implicit category learning appears to be consistent and dependent on several aspects of the underlying categories, unlike explicit category learning, it is unclear whether implicit category learning enjoys the same advantage when category exemplars are presented in an interleaved vs. a blocked design. Most implicit categorization tasks involve manipulation of features as opposed to manipulation of the temporal order of exposure.

4.1.1 Simulating temporal advantage

Prior work in implicit event boundaries has shown that stimuli shown closer to each other in time develop similar representations in memory and in particular the hippocampus and the medial temporal lobe (Bonner & Epstein, 2021; A. C. Schapiro et al., 2013; Turk-Browne, 2019). Context representations through, for example, SR, can provide algorithmic accounts for such increased similarity between co-occurring events. For example, consider events occurring as the graph in Figure 4.1

All nodes in both graphs are connected to all other nodes. However, connectivity in the graph on the left is determined by weighted edges – darker edges are 4 times as likely to be traversed as lighter edges. On the other hand, all edges (including within-node edges) are equivalent in the graph on the right. Since SR represents expected transition probabilities between each pairs of nodes, a (weighted) random

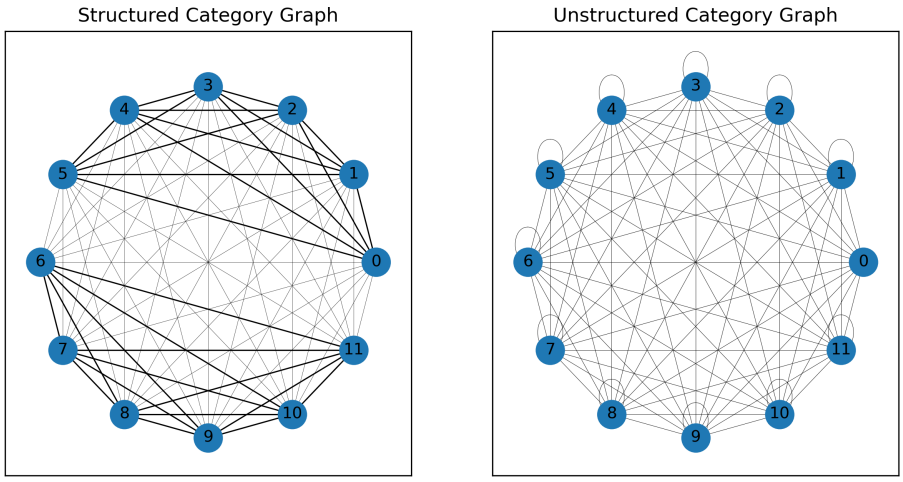


Figure 4.1. Graph structures used in categorization experiments. Edge thickness indicates transition probabilities between nodes.

walk through such a graph structure produces distinctive representations of context (Figure 4.2).

The SR representations thus provides a natural way of representing two temporally defined categories. The key question I ask in this chapter is whether temporal proximity can lead to an increased realization of an inherent visual category structure. That is, in tasks where participants are are unaware of categorization tests or of category diagnosticity of some visual features, does temporal proximity of same-category items lead to a realization of category-diagnosticity of features?

To simulate temporal advantage, I follow the exemplar matching principles from GCM (Nosofsky, 1986, 2011; Nosofsky et al., 1994; Rouder & Ratcliff, 2004). However, unlike the standard GCM, which is used to model categorization learned through explicit feedback of category membership, the tasks presented later in this chapter will not provide any information or an explicit learning signal regarding the true category membership. Additionally, the categorization tests in experiments of this chapter will *not* compare new exemplars with stored category exemplars in memory (especially

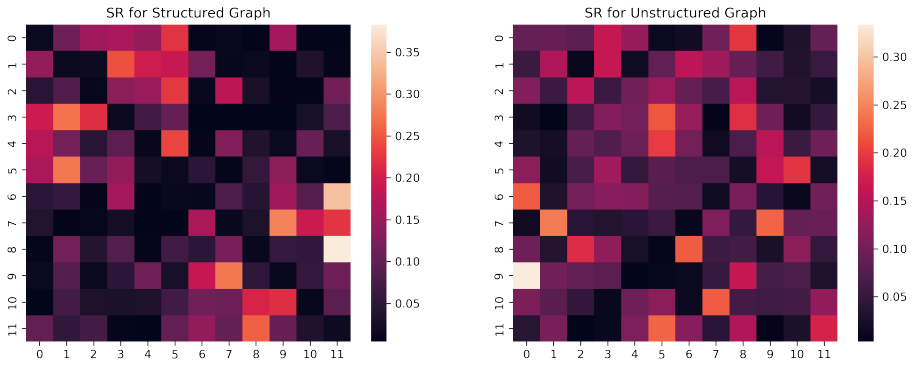


Figure 4.2. SR representations of graph structures used for categorization.

since no stored exemplars will have an explicit category label). Instead, participants will be asked to compare *studied* exemplars with two new exemplars which maintain most features of the studied exemplars while varying either on a subset of category diagnostic features or category non-diagnostic features. Participants will not be informed about which features define a category. Rather, features that stay consistent in the high-probability, within-cluster transitions (Figure 4.1) will be labeled as category diagnostic by the experimenter whereas features that change frequently in these high probability transitions will be labeled as category non-diagnostic. The goal is to thus investigate whether temporal co-occurrences cause participants to notice features that are category diagnostic. Finally, tasks in these experiments use binary valued features to allow for better control of the number of possible feature values and feature dimensions. Few modifications to the formulation of the GCM were therefore necessary. This modified GCM can be formally described as follows:

$$\begin{aligned} d_{ij} &= \sum_m w_m x_{im} \oplus x_{jm} \\ s_{ij} &= exp(-d_{ij}) \\ p(i|c) &= \frac{s_{ic}}{s_{ic} + s_{jc}} \end{aligned} \tag{4.1}$$

where d_{ij} is the bitwise XOR (i.e. an XOR for each binary features across two stimuli i and j) distance between two binary feature vectors representing items i and j. w_m is the weight associated with each of the features. s_{ij} is the similarity between items i and j. $p(i|c)$ represents the probability of selecting option i between options i and j given similarities of item c with items i and j. Notably, in this formulation of the GCM, the weights of dimensions are only relevant upon mismatch between those features (an assumption in line with other recognition memory and categorization theories such as the Diagnostic Feature Detection Theory (Wixted & Mickes, 2014)).

To incorporate the role of SR-driven context representation, I further assume that the attention weights w_m towards each feature are modulated by SR activations. Specifically, when a stimulus transition is experienced, features that stay constant are facilitated with an increased importance. The magnitude of this increase is assumed to be proportional to the SR representation of the two items. Formally,

$$w_m = \sum_{i,j}^{i_m=j_m} M(i,j) \tag{4.2}$$

Where $M(i,j)$ is the SR activation of item j in item i. Feature attentions weights, thus develop over time as items co-occur and the SR matrix M consolidates to represent two clusters seen in figure 4.2. This formulation allows to simulate an upper bound of category membership (for a given number of trials and set of parameters). Figure 4.3 shows this upper bound for the structured graph where transitions are differentially weighted to create temporal clusters relative to the unstructured graph where all transitions are equally likely.

In the experiment presented next, I aim to test this model prediction. In particular, participants who are explored to structured graph in binary featured stimuli will consolidate their experience and categorize items based on their (temporally defined) category diagnostic features. Here category diagnostic features are defined by consistency in the values of a feature among aliens that are presented closer to each other.

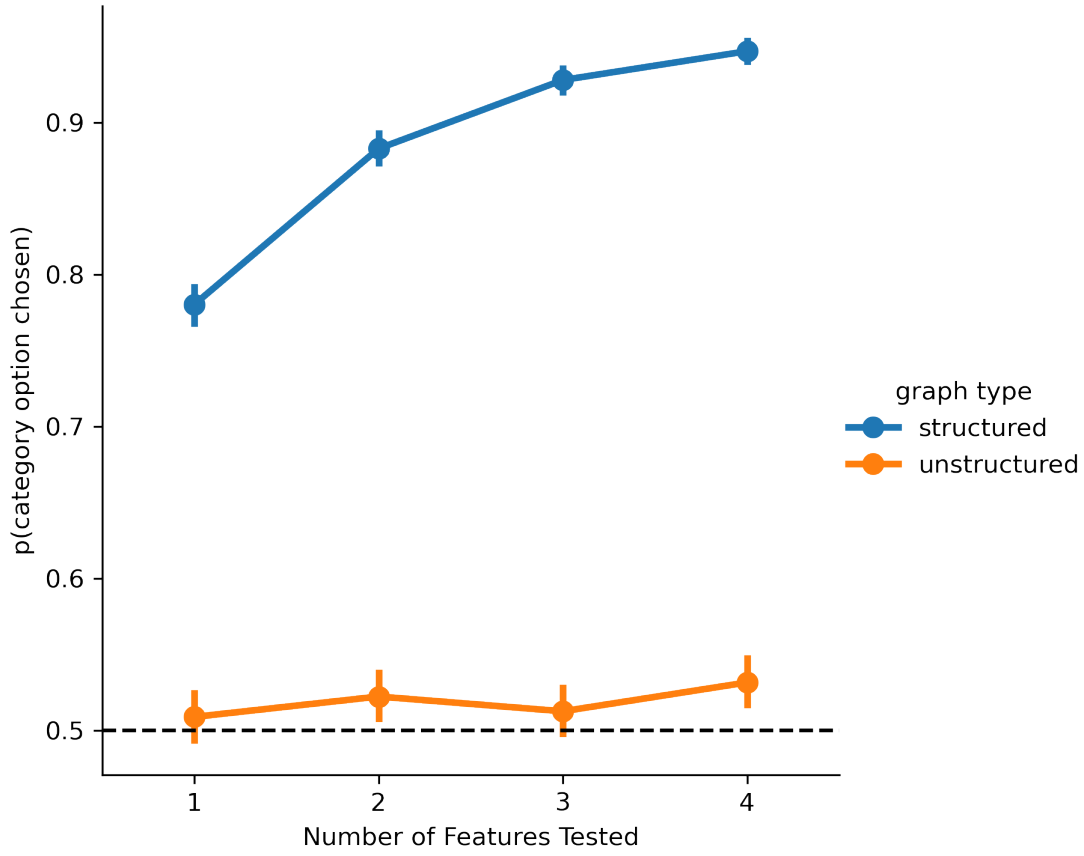


Figure 4.3. Simulated proportions of test trials where the option selected will be consistent on its category diagnostic features assuming feature weights modulated by SR.

That is, if the probability of going from an alien associated with node 0 to an alien associated to node 1 is high, common features between these aliens are labeled as category diagnostic whereas the features that stay consistent during low probability transitions are labeled as category non-diagnostic features.

4.2 Experiment 3a

4.2.1 Methods

Participants

58 participants were recruited online via prolific to participate in the study designed in Psychopy (Peirce, 2007). Participants were paid $5 for their time. All study procedures were approved by the University of Massachusetts Institutional Review Board.

Stimuli creation and assignment

Alien stimuli for used in this study were created using python's Matplotlib library (Hunter, 2007). Each alien consisted of 8 binary valued features, evenly distributed across 4 dimensions (shape, size, color, and orientation). The 8 binary valued features comprised of head color (brown or orange) and torso color (green or purple) for the color dimension; eyes (large or small circles) and arms (large or small lengths) for the size dimension; nose (square or circle) and bellybutton (square or circle) for the shape dimension, and antenna (pointing inwards or outwards) and feet (pointing inwards or outwards) for the orientation dimension.

One randomly chosen feature from each dimension was chosen as a 'category diagnostic' feature whereas the other was deemed as non-diagnostic. Category diagnosticity of a feature was determined based on cluster assignment in figure 4.1. Specifically, all category diagnostic features in the same cluster were assigned the same value (for example, all green torsos). In order to equate the number of features of the non-category diagnostic feature, three stimuli in a cluster were assigned the same value of a non-category diagnostic features whereas the other three were assigned a different value (e.g. three orange heads and three brown heads). The table below shows example feature values for the categorization experiment.

Table 4.1. An example set of stimuli with Antenna Orientation, Eye size, Head Color and Nose shape are category diagnostic based on proximity of occurrence.

Feature Elements Stimulus	Antenna	Arms	Bellybutton	Eyes	Feet	Head	Nose	Torso
1	0	1	0	0	1	0	0	0
2	0	0	1	0	0	0	0	1
3	0	0	0	0	1	0	0	1
4	0	1	1	0	0	0	0	0
5	0	1	0	0	0	0	0	1
6	0	0	1	0	1	0	0	0
7	1	1	0	1	1	1	1	0
8	1	0	1	1	0	1	1	1
9	1	0	0	1	1	1	1	1
10	1	1	1	1	0	1	1	0
11	1	1	0	1	0	1	1	1
12	1	0	1	1	1	1	1	0

This formulation allows to thus test whether temporal proximity of consistent features leads to an increased weights on features that co-occur or whether temporal proximity of the *change* in features (i.e. the non-diagnostic features) leads to increased weights in those features.

Design and Procedure

The experiment consisted of two phases; an exposure phase and a categorization phase. During the exposure phase, participants were shown one of 12 aliens at a time. After a brief period the alien flipped (randomly) left or right. Participants were asked to hit a key to indicate which direction the alien flipped.

The order of stimulus presentation was controlled by a weighted random walk through the graph structure in figure 4.1. 29 participants were presented a stimulus stream following the weighted structured graph (left panel) whereas the other 28 were presented a stimulus stream following the unstructured graph (right panel). For participants experiencing structured weighted random walk, category diagnostic features remained consistent across transitions with 80% probability and changed

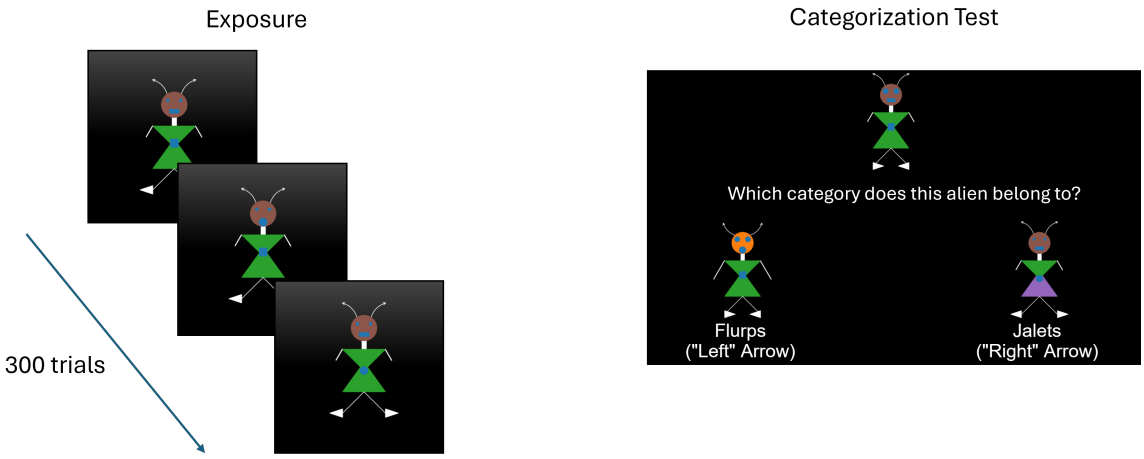


Figure 4.4. Experiment design schematic for experiments 4a and 4b.

with a 20% probability whereas the category non-diagnostic features changed with an 80% probability and remained constant with 20% probability. On the other hand, the participants in unstructured exposure condition experienced an equal probability of change across all features.

The general experiment design is shown in figure 4.4. After 300 trials of exposure, participants were asked to categorize the aliens they had seen at exposure. Specifically, for each of the 12 aliens, two options were provided. The category option matched on all category diagnostic features and all but n category non-diagnostic features (where $n \in [1, 4]$) and the non-category option matched on all category non diagnostic features and all but n category diagnostic features. Participants were asked to select which of the two option does the studied exemplar more relate to in their visual properties. Thus, if a participant selected the category option more often in the structured case, temporal proximity of consistent features increases the weight of shared feature values. On the other hand, if a participant selected the non-category option more often in the structured case, temporal proximity of non-consistent features increases weights of the non-shared feature values.

4.2.2 Results

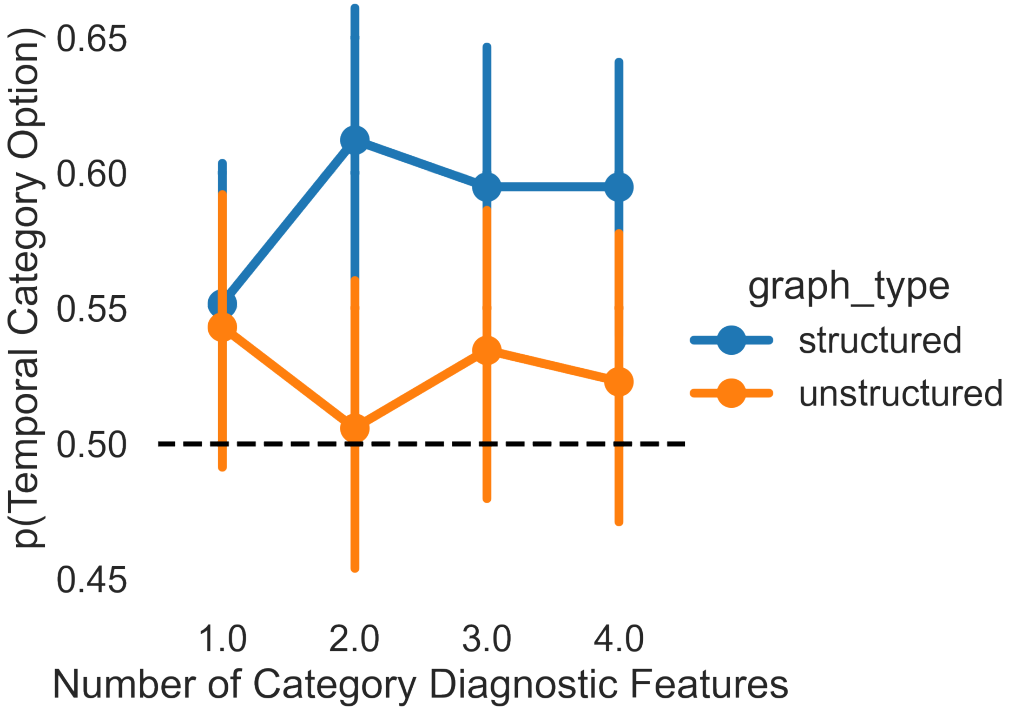


Figure 4.5. Proportion of categorization trials where a category diagnostic feature (one which remained more consistent over time) was used to categorize items.

The following Bayesian model was used to estimate the effect sizes for categorization based on category diagnostic features for structured weighted walk relative to unstructured walk.

$$\begin{aligned} 1|participant &\sim \mathcal{N}(0, Half\mathcal{N}(3.65)) \\ C(num\_feats) : graph\_type &\sim \mathcal{N}(0, 7.5) \\ \mu = 0 + C(num\_feats) : graph\_type + (1|participant) & \\ p(category\ diagnostic\ option) &\sim Bernoulli(\mu) \end{aligned} \tag{4.3}$$

Posterior parameter distributions of choosing the option with common category diagnostic feature in the structured walk relative to unstructured walk are shown in figure 4.6

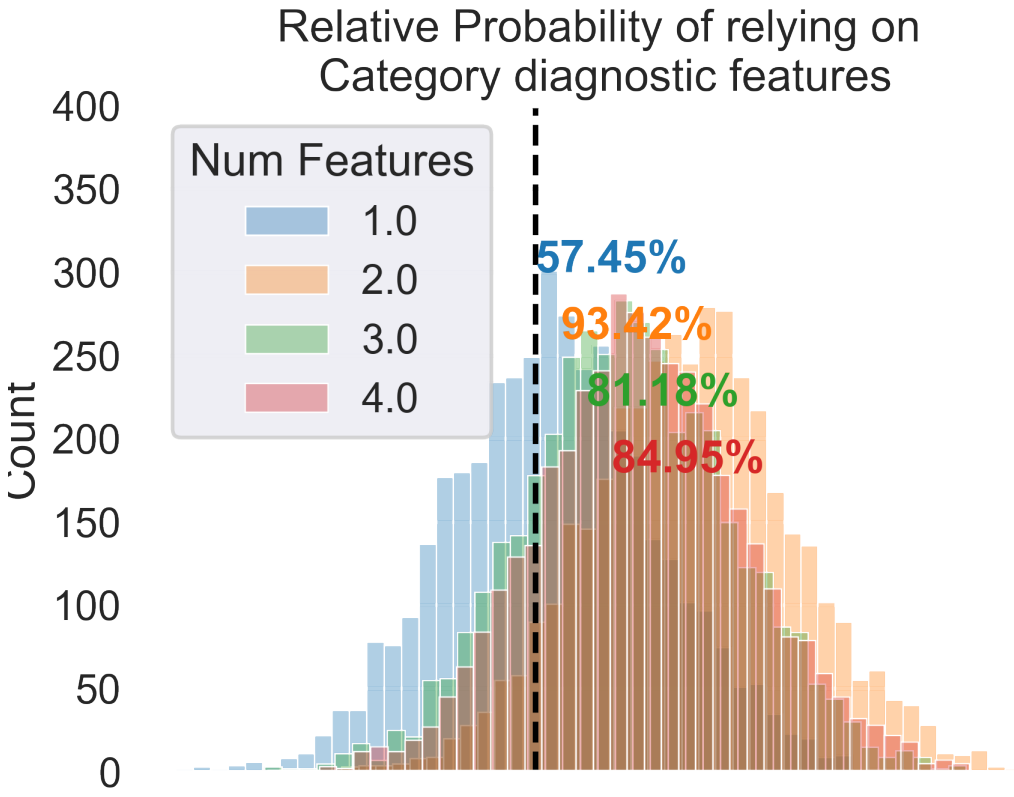


Figure 4.6. Bayesian estimates of proportions of temporally consistent features used as category diagnostic when exposed to structure relative to when not exposed to structure.

Results from this experiment thus indicate that participants indeed pick up on features that are consistent between items that are temporally proximal. In the next experiment I aim to extend this finding to assess whether such temporal arrangements can be used as an *intervention* to have participants learn arbitrary features as category diagnostic. Specifically, by using a limited set of features to define a category (as opposed to a pseudo-random selection in experiment 4a). Experiment 4b thus provides (1) An opportunity to replicate findings in experiment 4a, and (2) An

assessment whether temporal proximity can be used as an experimental intervention to have participants learn category diagnostic features.

4.3 Experiment 3b

4.3.1 Methods

Participants

40 participants were recruited over Prolific to complete this 30-minute study. Participants were paid $5 for their time. All study procedures were approved by the University of Massachusetts Institutional Review Board.

Design and Procedure

Two groups of participants experienced two different sets of features as category diagnostic. Both groups of participants experienced a weighted random walk during their exposure phase (similar to the structured condition in experiment 4a). For group A, one feature (out of two) for each of the four dimension was chosen to be category diagnostic. For group B, the other feature was category diagnostic. Specifically, head color, eye size, feet orientation, and nose shape were used as category diagnostic features for participants in group A. Torso color, arm length, antenna orientation, and bellybutton shape were used as category diagnostic features for participants in group B. All other experimental procedures were the same as in experiment 4a (see figure 4.4).

4.3.2 Results

Figure 4.7 shows the proportion of categorization trials where a category diagnostic feature was used to classify the test item. The same model as in experiment 4a was used to estimate effect sizes for both conditions. Figure 4.8 shows the posterior estimates of these effect sizes. Interestingly, features that were category diagnostic for participants exposed to category A were used to categorize test items. On the other

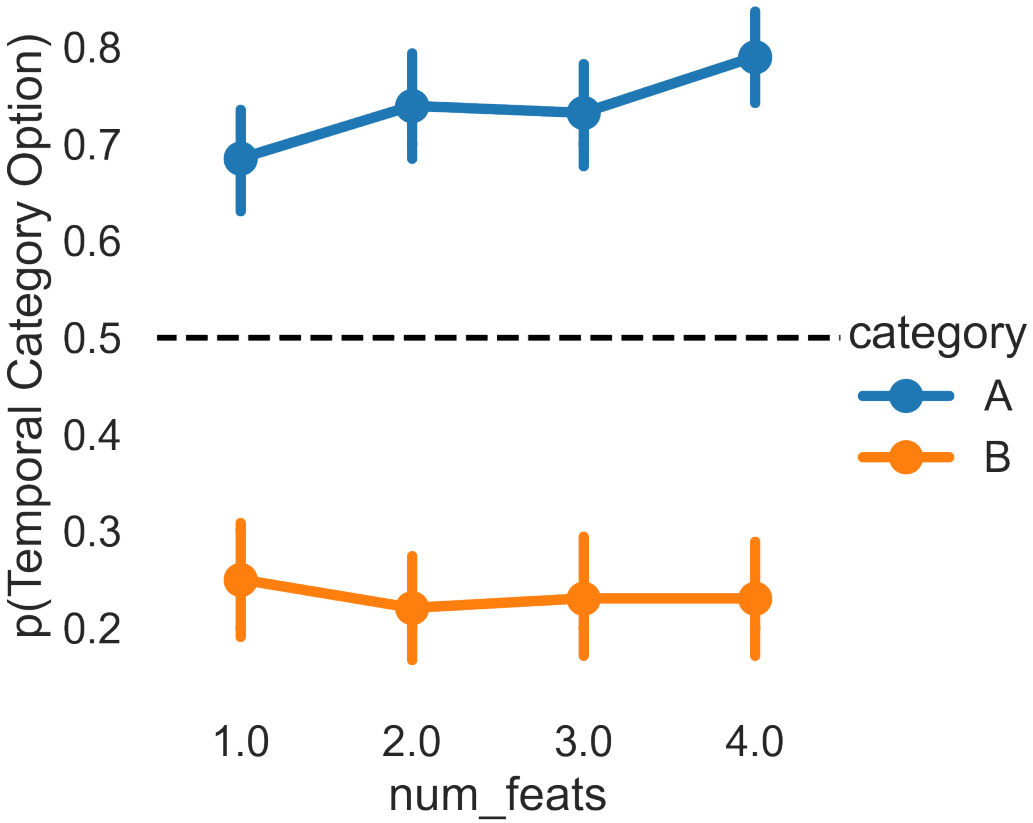


Figure 4.7. Proportion of categorization trials for which category diagnostic features were chosen to determine category membership

hand, features that were category diagnostic for participants exposed to category B were *not* used; instead category non-diagnostic features (i.e. features that remained consistent during low probability transitions) were instead used to group the test item.

4.4 Discussion

The aim of this chapter was to assess if consistency of temporally proximal features leads to category formation. Experiments in this chapter follow a qualitatively similar design relative to past experiments assessing whether category formation is

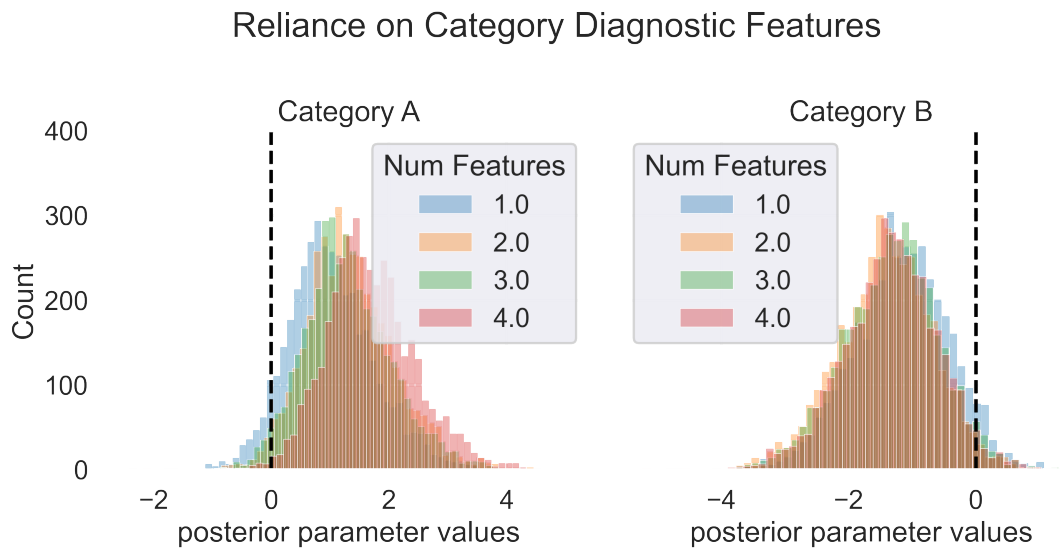


Figure 4.8. Bayesian posterior parameter estimates modeling the proportions of temporally consistent features used as category diagnostic for participants exposed to category A (*left panel*) features as category diagnostic and when participants were exposed to category B features as category diagnostic. (*Right panel*)

better with blocked design (where items from one category are presented serially after items from another category) compared to interleaved design (where items from all categories are presented together). Unlike past work, the present experiments investigate whether similar (or opposite) benefit exists for tasks where category formation is implicit – until test, participants were not informed of any categorical structure in the stimuli.

Past findings in explicit categorization have suggested that category formation is improved with interleaved presentation compared to blocked presentation. However, temporal proximity has been shown to increase representational similarity in the hippocampus (A. C. Schapiro et al., 2013); blocked design may thus lead to realization of consistent patterns based on temporal proximity. Assuming SR as representation for temporal proximity, model simulations show that temporal proximity can indeed lead to categorization.

Experiment 4a provides support for temporal proximity leading to categorization. However, experiment 4b leads to an interesting update to the model of category learning. When temporal proximity was used as an intervention to category learning, participants picked up on consistency of some features (category A) but on the *differences* of the other features (category B) showing an interleave benefit.

Model Update

In the modeling approach shown earlier, two assumptions were made (1) All features were assumed to be equally weighted prior to exposure and (2) Feature weights were modified by consistency temporal proximity – for each pair of stimuli, if they shared a feature, the weight of the shared feature was proportional to SR representation of the stimuli pairs.

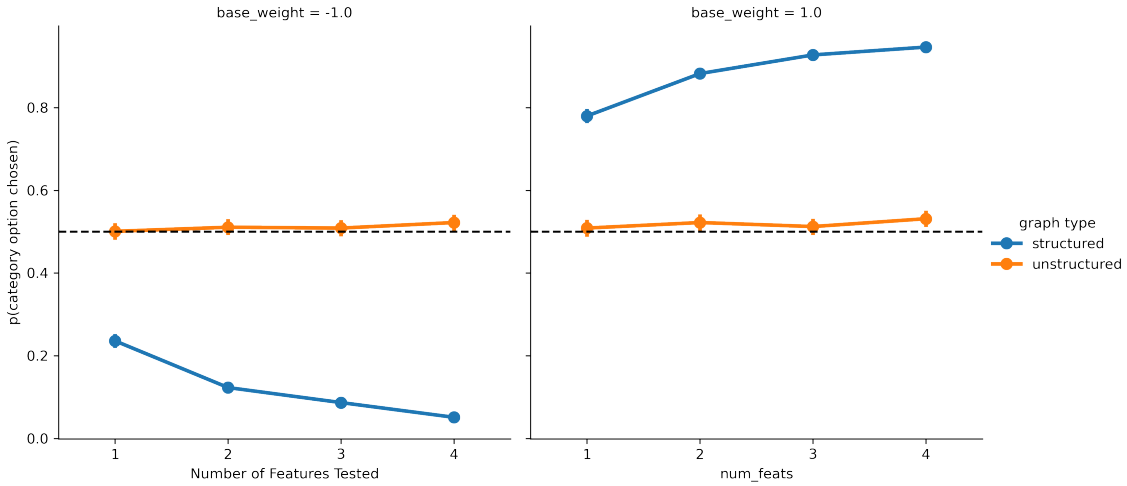
To incorporate results from experiment 4b, equation 4.2, can thus be modified as follows:

$$w_m = w_b * \sum_{i,j}^{i_m == j_m} M(i,j) \tag{4.4}$$

where w_b is the baseline weight for that feature. Updated model simulations are in figure 4.4

The updated model allows for interleaving benefit by increasing the proximally non-consistent feature weights. It is unclear what features lead to an increased attention weight on differences (as would be the case for interleaved categories) or an increased attention weight on consistencies (for blocked categories). Perhaps, features that are more notable lead to an interleaving benefit as more frequent changes in feature values of more notable features become apparent. Norming studies in future experiments can address whether temporal attention benefit depends on baseline noticeability.

Figure 4.9. Updated model simulations. Base feature weights determine whether more weight is placed on proximally similar *(right panel)* or proximally distinct features *(left panel)*.



4.5 Conclusion

In this chapter, I demonstrate a use of a context representation framework (specifically predictive representation using SR) to understand the mechanisms behind implicit category learning based on temporal relationships. Findings of these experiments can further model considerations in classical categorization models such as the GCM to incorporate information about temporal presentation order in assigning feature weights. More broadly, given that temporal associations seem to have an effect on how things are visually categorized (albeit, dependent on the specific feature), manipulating temporal associations can be used in more practical settings such as learning natural categories in classrooms. (For example when training students to visually recognize types of geological rocks (Nosofsky et al., 2017, 2018)).

CHAPTER 5

GENERAL DISCUSSION AND CONCLUSION

We experience a stream of continuous information daily. For ease of reference and recall, we break this stream down and store it in meaningful chunks. Event segmentation is a cognitive process that provides mechanisms to break down this temporal stream of information and integrate it across multiple chunks. Prior work has focused on understanding the event cognition process. In the spirit of decomposing cognitive processes into representations and operations on those representations (Cowell et al., 2019), in this dissertation I explore how this process can be assessed through shared representations of temporal structure in memory and operations on these representations.

In the second chapter, I investigate how representations of ongoing context can produce behavioral patterns in response times. To that end, I contrast two models of context – the Successor Representation (SR) as a predictive model that maintains context as an expectation of the future and follows error-driven learning with the Temporal Context Model (TCM) as an associative model that maintains representation as current activity and follows Hebbian learning. I show that the two models produce qualitatively different predictions when exposed to varying amounts of limited information about the environment. I then test these predictions in a serial reaction time task and show that data are more consistent with predictions from the SR model.

Findings in this chapter provide a theoretical framework with which event boundaries and statistical learning could be studied. Prior work has suggested that slow-

down at boundary nodes may be due to a closely related algorithmic process where errors in realization of the temporal structure (Lynn, Kahn, et al., 2020). In addition to the differences in walk lengths explored in this chapter, the predictive SR framework further allows to make testable predictions about information available at different stages of the learning process. SR further provides an implementation level mechanisms via dopaminergic projections to the Hippocampus may allow for error-driven learning of statistical regularities(Gershman, 2018; Stachenfeld et al., 2017). Future work should aim to directly investigate the differences in predictions between the two algorithmic accounts (erroneous learning vs predictive representation).

One major assumption in chapter 2 warrants some discussion. I assume that higher information contained in a node's representation (derived via SR or TCM) translates to slower reaction times. While there is prior support for slower responses relating higher information in similar tasks (Lynn, Papadopoulos, et al., 2020), this measure of response time is indirect. It is possible that the reaction time increases at a given node are a result of the source node of the incoming transition rather than the node itself – previous research have found that as node degree increases, reaction times increase (Lynn, Papadopoulos, et al., 2020). Nevertheless, in the comparisons tested for experiment 1, source nodes are always non-boundary nodes. Thus, an observed effect must be due to the type of node at which the responses are made. In this work, I made a direct comparison between two models of context representations, the TCM and SR. Future work should also consider other models of context representations and assess how these models of associative memory differ in such statistical learning tasks (Estes, 1955; Mensink & Raaijmakers, 1988; Murdock, 1997). Finally, in the experiments used in this chapter, no distinction has been made between what aspect of the participant's experience is associated with a node in the graph in Figure 2.1. Future work should distinguish whether the effects of nodes predicted by models are due to effects on, for example, the visual stimulus associated with those nodes, the

motor response associated with that node (which is in turn associated with the visual stimulus), or a combination of both.

Prior research and findings in event cognition are often limited to tasks where event boundaries are defined by explicit context changes (e.g. change of scene in a movie). Recently, event boundaries have also been shown to be formed without such explicit context changes but through an underlying temporal structure. However, these implicit event boundary tasks are not typically tested using the same tests used for explicit event boundaries. In Chapter 3, across two experiments I assess whether implicitly operationalized boundaries share behavioral properties with explicit event boundaries by (1) Testing whether they are remembered better than non-boundaries and (2) Testing whether events across boundaries are perceived farther than those within. Results from the experiment recognition memory experiment provide support for shared representations between implicit and explicit boundaries. Results from the distance judgment experiment do not provide sufficient evidence for such shared representations. However, combined results from both experiments provide support for SR-based context representations for implicit event boundaries.

Chapter 3 provides an important indication on shared representations between explicitly operationalized event boundaries and implicitly operationalized event boundaries and whether implicit boundaries are indeed event boundaries. While they have been labeled as boundaries in prior work (2013), work in this chapter provides a more direct test of whether implicit boundaries share behavioral properties of explicit boundaries. The SR modeling framework further provides a representational account for implicit boundaries. Future work should test whether explicit boundaries can be similarly represented through a representational framework such as the SR. While some modeling work has used the associative Context Maintenance and Retrieval Model (Rouhani et al., 2020) to understand explicit event boundaries, findings in Chapter 2 suggest that a predictive model may be more appropriate.

Chapter 3 is limited to two tests that are used in explicit boundary paradigms. Future work should aim at testing whether implicit boundaries share other properties of explicit boundaries such as to serve as access points in memory (Michelmann et al., 2023), points in replay (Hahamy et al., 2023; Sols et al., 2017), and points of memory integration (Griffiths & Fuentemilla, 2020). Systematic assessment of shared properties between implicit and explicit event boundaries will provide further insights into general structure of temporal cognition and pattern recognition.

Finally in Chapter 4, I show that predictive representation of temporal events (such as SR) can be further used to understand the cognitive process of category learning. Prior findings have suggested that there is a differential benefit to presenting different category exemplars in an interleaved fashion vs blocked fashion in category learning. The SR model provides a representational account for this temporal order of presentation effects by modulating attention towards category diagnostic features. While the precise nature of this modulation operation further depends on the specific visual features that define categories, implicit visual category learning tasks in this chapter show, as predicted by SR, that the temporal order of presentation of category exemplars indeed matters in how categories are learned.

Work in Chapter 4 thus provides a framework for understanding higher order cognition and representation. While categorization and category learning in cognitive psychology is well studied, current work provides a deeper algorithmic insight into how we learn categories in a natural, unsupervised manner. The experimental paradigms used can be further adapted to address other higher order cognitive functions of learning as well. For example, a common problem often addressed through Hierarchical Reinforcement Learning (Botvinick, 2012), navigation through space can be thought of as achieving a set of sub-goals where all experiences within a sub-goal can be thought of a category.

While the framing of studies presented in this dissertation is via event boundaries or statistical learning, the representational framework does not necessarily distinguish between these two processes. Chapters 2 and 3 provide an important connection between these two fields of event boundaries and statistical learning (See also Perruchet and Pacton, 2006). Event boundaries, which are implicitly operationalized, are learned over time and provide an important marker to separate statistical regularities in the experienced environment. In turn, extracting of statistical patterns provides a natural break between two statistical regularities that are sufficiently different from each other thereby leading to event boundaries. The representational framework used here provides a way to algorithmically combine the findings in these two fields of cognition.

This dissertation assesses shared representations between implicit and explicit event boundaries. One could also argue that boundaries that naturally became explicit were also originally learned implicitly. For a child growing up, the items in the kitchen are no different than the items in the living room. Nevertheless, over time, the child begins to identify the patterns by interacting (or watching others interact) with these items in different contexts. The shared context between kitchen items slowly consolidates into coherent, abstract knowledge that a kitchen is where one eats. Similarly, the shared context within the living room items slowly leads to the abstract knowledge that the living room is where one hangs out (albeit supported by instructions from parents). This way, talking to a friend in the living room becomes a separate event from talking to the same friend in the kitchen. Parent instruction notwithstanding, this formation of boundaries, which are explicit boundaries for adults, was originally acquired by extracting regularities in the environment through statistical learning.

Statistical learning, a cognitive process through which we acquire patterns in our environment, is widely applicable across multiple domains of cognitive psychology

(See A. Schapiro and Turk-Browne, 2015 for a brief review). Our brain supports learning of such regularities automatically and implicitly. Learning the regularities in our environment allows us to develop useful heuristics to make quick decisions when needed. For example, when we visit a new grocery store, we know to generally expect all medications arranged in a specific section separate from a section of food items. This abstract knowledge of how grocery stores work allows us to perform quicker visual searches through aisle headers depending on what we are looking to shop for. The key question this dissertation seeks to answer is what algorithms our brain may implement in order to acquire abstract knowledge of these patterns.

Overall, this dissertation shows that representing a temporal sequence of events such that each event represents a prediction of what might happen next, provides *one* reasonable explanation of how we may learn statistical regularities from our environment. This dissertation further shows that such a predictive representation then provides a common framework for investigation into various aspects of human cognition such as forming event boundaries, statistical learning, or categorization. Typically, theoretical work in cognitive psychology is aimed at understanding individual cognitive processes that allow us to function. This representational framework opens a way for cognitive scientists to understand the broader role of such implicit, unsupervised error-driven learning. In future work more cognitive processes such as learning and decision-making in complex environments, the role of cognitive flexibility in behavior, the impact of the environmental experience on emotional and other affective internal states, and others can be distilled into such a common predictive representational account, thereby allowing for a more coherent understanding of human brain function.

5.1 Broader Implications

The spatiotemporal hypothesis in event cognition suggests that items that are closer together in both space and time share representations in the hippocampus (Turk-Browne, 2019). The representational framework proposed in this dissertation suggests that this finding can be extended to other aspects of human psychology as well. For example, the cross-race effect studied in social psychology and eyewitness identification (Wilson et al., 2013; Young et al., 2012) where recognition for same race faces than different race faces could be studied through a lens of shared representation in how often exposure through development is higher with other faces of the same race.

Errors that lead to learning of of representations such as SR are often linked to dopaminergic signaling in the midbrain (Gershman, 2018). Future work should test whether increased availability of dopamine during adolescence which has been shown to enhance memory may also allow for a better pattern recognition during that period of development (A. O. Cohen et al., 2022). Impacts on pattern recognition due to varying dopamine availability over development could have further implications for instruction and in teaching complex concepts across different ages. Similarly, future studies should test whether depleted dopamine levels due to Parkinson's or Huntington's would on the other hand deter pattern recognition and statistical learning.

Finally, the focus of studies cognitive psychology (and for this dissertation) has often been limited to participant population in the global north. It is clear that even in meaningless stimuli, the nature of exposure impacts low level cognition of memory and perceptual categorization. Future research must therefore test the generalizability of this representational framework in heterogenous participant population. Recent research in cultural effects on cognition suggests differing processes underlying 'lower' level cognition of numbers and time (Pitt, 2018). Such molding of cognitive processes could also be therefore viewed through the lens of differing representations

with consistent operations on those representations as function of different cultural exposures.

It has been argued that boundaries separating events temporally and spatially share representations. This dissertation further advocates for shared (algorithmic) representations across different processes in cognitive psychology. I show that a common predictive representation framework can be useful in understanding and relating implicit statistical and motor learning (chapter 2), event cognition and memory (chapter 3), and categorization (chapter 4). Using a common representation can provide an important algorithmic constraint in understanding different cognitive processes. Future work in cognitive psychology can therefore focus on testing operations on these common representations that may account for observed behavior.

APPENDIX A

CHAPTER 2

A.1 Model simulations of surprisal

A direct measure of reaction time can be through 'surprisal'; a quantity that measures how much any given node (in the modular graph of Figure 2.1) that is visited is expected to be visited. Simply put, if a visited node is surprising, response times may be higher. This effect is observed in prior literature where surprisal is measured as $logk$ where k is the number of connections in the node prior to the current node. Lynn, Papadopoulos, et al., 2020 found that for this surprisal significantly impacted reaction times for a full length of random walk.

While all nodes in the graph used have the sam number of connections, a similar model of surprisal was used to simulate potential differences between the model. Briefly, since each cell i, j in a row i of the SR/TCM matrix represents the relative activation of the node associated with that column, a surprisal measure was computed as a $logM_{ij}$ of that cell.

As Figure A.1 shows, first surprisal across cluster increases with increased walk length. Second, there are no qualitative differences across possible parameter values in the effect of surprisal. For that reason, this measure of surprisal was not used to assess the models.

A.2 Comparing first and the last blocks

Similar to comparisons done in the main text, the response times between the boundary and non boundary nodes for the first and last blocks were compared when

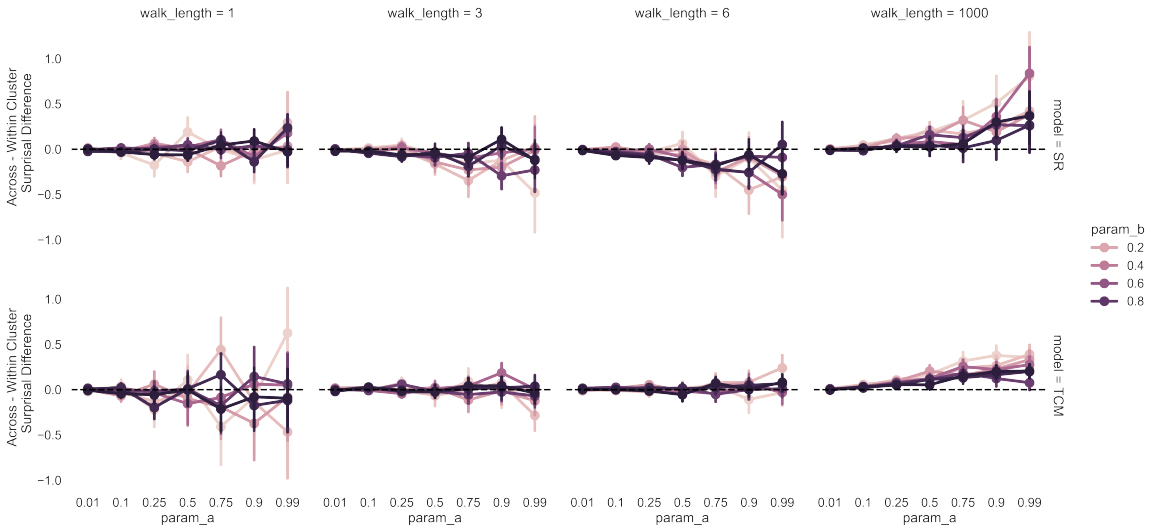


Figure A.1. Surprisal differences between cross-cluster and within-cluster transitions across walk lengths for representations generated by both SR (top row) and TCM (bottom row) models.

transitions leading into those nodes were within cluster (i.e. from another non-boundary nodes).

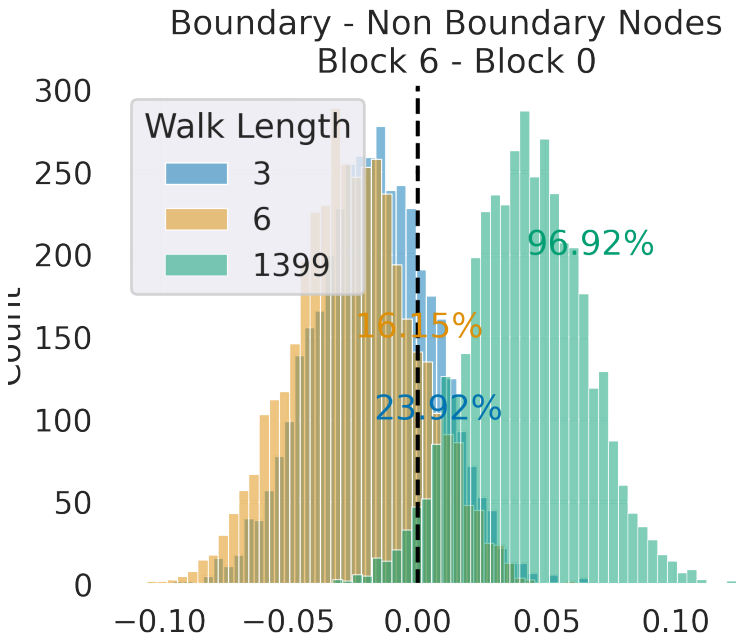


Figure A.2. Posterior estimates of comparisons the slowed down reaction times for boundary nodes relative to non boundary nodes. Reaction times slowed down more with larger walk lengths.

Figure A.2 provides further evidence in support of the SR model. Reaction times decrease across the board. The decrease is lower for boundary nodes than non boundary nodes and this difference increases with walk length providing support for the SR model.

A.3 Fitting the entire learning curve

The entire learning curve was fit using a linear model to estimate differences in decreased response times between boundary and non boundary nodes over time.

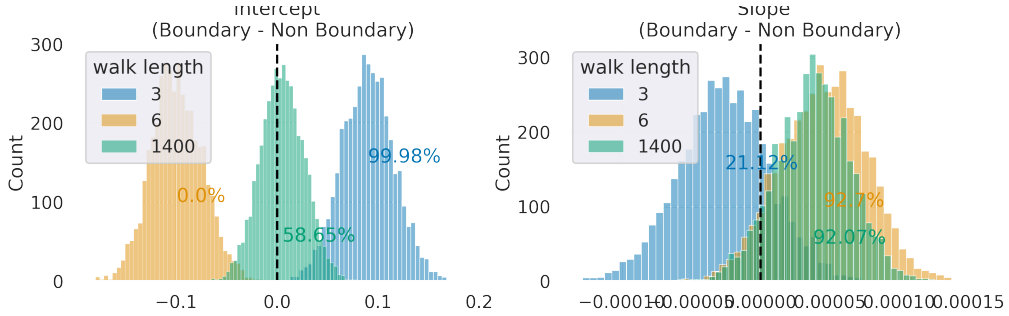


Figure A.3. Intercept (*Left panel*) and Slope (*Right panel*) differences of the linear model fit to all trials.

As shown in figure A.3, boundary and non boundary intercepts differed across walk lengths. This difference is likely due to randomly assigned response keys. The key metric of differences is the reduction in response times over time as measured by the slopes. As expected (from the SR model), walk lengths of 3 and 1399 lead to slower decrease in response times for boundary nodes than non boundary nodes (when transitioned to from another within cluster (non boundary) node).

Interestingly, it appears that these slope differences are not meaningfully different between walk lengths of 6 and 1399. However, the findings reported in the main text (Figure 2.10), it is likely that these differences were apparent earlier in the learning process for the walk length of 1399 than for walk length of 6. Conflicting results with the main text likely implies a need for a more complex model (such as a dual rate model (McDougle et al., 2015; Savalia et al., 2022; M. A. Smith et al., 2006)) which allows for a quick initial decrease in reaction times followed by a slow asymptote. Future modeling work should investigate the impact of walk lengths on different aspects of the learning process.

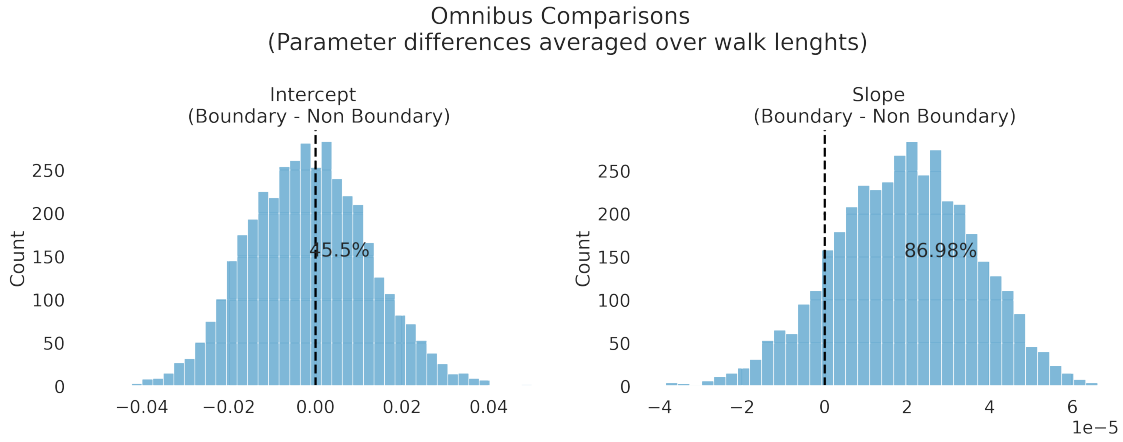


Figure A.4. Omnibus comparisons between boundary and non-boundary nodes for linear models fit to all trials.

Finally, Figure A.4 presents comparisons between intercepts and slopes averaged across all walk lengths. Overall, boundary and non-boundary intercepts are largely similar (45% samples above 0), however the average slope for boundary nodes is larger than that for the non-boundary nodes indicating response times to non-boundary nodes decrease faster than that to boundary nodes.

A.4 Model Statistics

Stats for comparisons between the RTs for the first two blocks

	mean	sd	hdi 2.5%	hdi 97.5%
beta block[0.0]	-0.066	0.176	-0.409	0.269
beta block[1.0]	-0.080	0.177	-0.417	0.271
beta lag	0.039	0.007	0.026	0.053
beta transition exp	-0.018	0.004	-0.025	-0.011
beta trial ntr[boundary, 0.0]	-0.555	0.024	-0.602	-0.508
beta trial ntr[boundary, 1.0]	-0.629	0.028	-0.685	-0.574
beta trial ntr[non-boundary, 0.0]	-0.596	0.023	-0.639	-0.548
beta trial ntr[non-boundary, 1.0]	-0.650	0.028	-0.702	-0.596

Table A.1. Posterior parameter statistics for model fit to walk length 3 data comparing the first two blocks

	mean	sd	hdi 2.5%	hdi 97.5%
beta block[0.0]	0.143	0.173	-0.191	0.482
beta block[1.0]	0.124	0.181	-0.225	0.475
beta lag	0.043	0.008	0.028	0.058
beta transition exp	-0.031	0.004	-0.038	-0.023
beta trial ntr[boundary, 0.0]	0.252	0.033	0.190	0.319
beta trial ntr[boundary, 1.0]	0.184	0.037	0.111	0.257
beta trial ntr[non-boundary, 0.0]	0.316	0.033	0.252	0.383
beta trial ntr[non-boundary, 1.0]	0.241	0.037	0.165	0.310

Table A.2. Posterior parameter statistics for model fit to walk length 6 data comparing the first two blocks

	mean	sd	hdi 2.5%	hdi 97.5%
beta block[0.0]	-0.043	0.172	-0.378	0.294
beta block[1.0]	-0.076	0.177	-0.427	0.267
beta lag	0.066	0.007	0.053	0.080
beta transition exp	-0.010	0.003	-0.016	-0.005
beta trial ntr[non-boundary, 0.0]	-0.476	0.023	-0.522	-0.431
beta trial ntr[non-boundary, 1.0]	-0.635	0.027	-0.688	-0.581
beta trial ntr[boundary, 0.0]	-0.499	0.024	-0.546	-0.455
beta trial ntr[boundary, 1.0]	-0.602	0.027	-0.655	-0.548

Table A.3. Posterior parameter statistics for model fit to walk length 1399 data comparing the first two blocks

Stats for comparisons between the RTs for the first and last blocks

	mean	sd	hdi 2.5%	hdi 97.5%
beta block[0.0]	-0.070	0.173	-0.392	0.274
beta block[6.0]	-0.100	0.172	-0.431	0.236
beta lag	0.029	0.006	0.017	0.041
beta transition exp	-0.007	0.002	-0.010	-0.003
beta trial ntr[boundary, 0.0]	-0.561	0.020	-0.601	-0.523
beta trial ntr[boundary, 6.0]	-0.704	0.037	-0.778	-0.634
beta trial ntr[non-boundary, 0.0]	-0.604	0.020	-0.642	-0.563
beta trial ntr[non-boundary, 6.0]	-0.730	0.039	-0.808	-0.656

Table A.4. Posterior parameter statistics for model fit to walk length 3 data comparing the first and the last blocks

	mean	sd	hdi 2.5%	hdi 97.5%
beta block[0.0]	-0.054	0.175	-0.406	0.279
beta block[6.0]	-0.099	0.178	-0.460	0.228
beta lag	0.038	0.006	0.026	0.050
beta transition exp	-0.006	0.002	-0.010	-0.003
beta trial ntr[boundary, 0.0]	-0.538	0.019	-0.574	-0.501
beta trial ntr[boundary, 6.0]	-0.734	0.038	-0.809	-0.660
beta trial ntr[non-boundary, 0.0]	-0.528	0.020	-0.568	-0.489
beta trial ntr[non-boundary, 6.0]	-0.700	0.038	-0.777	-0.629

Table A.5. Posterior parameter statistics for model fit to walk length 6 data comparing the first and the last blocks

	mean	sd	hdi 2.5%	hdi 97.5%
beta block[0.0]	-0.053	0.178	-0.404	0.306
beta block[6.0]	-0.122	0.176	-0.458	0.230
beta lag	0.060	0.006	0.049	0.072
beta transition exp	-0.003	0.001	-0.006	-0.001
beta trial ntr[non-boundary, 0.0]	-0.515	0.019	-0.553	-0.478
beta trial ntr[non-boundary, 6.0]	-0.845	0.034	-0.910	-0.778
beta trial ntr[boundary, 0.0]	-0.538	0.020	-0.576	-0.499
beta trial ntr[boundary, 6.0]	-0.825	0.034	-0.891	-0.762

Table A.6. Posterior parameter statistics for model fit to walk length 1399 data comparing the first and the last blocks

Parameter statistics for linear model including all trials

	mean	sd	hdi 2.5%	hdi 97.5%
alpha ntr[3, boundary]	0.466	0.138	0.215	0.760
alpha ntr[3, non-boundary]	0.376	0.138	0.106	0.650
alpha ntr[6, boundary]	0.440	0.138	0.166	0.699
alpha ntr[6, non-boundary]	0.539	0.138	0.261	0.789
alpha ntr[1400, boundary]	0.489	0.134	0.281	0.826
alpha ntr[1400, non-boundary]	0.484	0.133	0.263	0.802
beta ntr[3, boundary]	-0.000	0.000	-0.001	-0.000
beta ntr[3, non-boundary]	-0.000	0.000	-0.001	-0.000
beta ntr[6, boundary]	-0.000	0.000	-0.001	-0.000
beta ntr[6, non-boundary]	-0.001	0.000	-0.001	-0.000
beta ntr[1400, boundary]	-0.001	0.000	-0.001	-0.001
beta ntr[1400, non-boundary]	-0.001	0.000	-0.001	-0.001

Table A.7. Parameter statistics for linear model fitting all trials. Parameter 'alpha' is the intercept, and parameter 'beta' is the slope of the linear model.

APPENDIX B

CHAPTER 3

Table B.5. Bayesian SDT Model results for boundary nodes from experiment 3a.

	mean	sd	hdi 3%	hdi 97%
accuracy exposure	-0.805	0.716	-2.088	0.591
true old—structured	4.088	0.332	3.472	4.711
true old—unstructured	4.200	0.308	3.611	4.783
true old—condition sigma	5.228	2.233	1.934	9.368

Table B.6. Bayesian SDT Model results for non-boundary nodes from experiment 3a.

	mean	sd	hdi 3%	hdi 97%
accuracy exposure	-1.347	0.715	-2.665	-0.020
true old—structured	3.885	0.269	3.398	4.406
true old—unstructured	4.593	0.299	4.017	5.122
true old—condition sigma	5.156	2.002	2.010	8.873

Table B.1. Accuracy and Response time Means and Standard Deviations for exposure and recognition phases in experiment 2

				accuracy		rt	
				mean	std	mean	std
phase	block	stimulus type	condition				
exposure	0	boundary	structured	0.683	0.465	1.644	1.911
			unstructured	0.739	0.439	1.710	1.732
		non-boundary	structured	0.669	0.471	1.619	1.728
			unstructured	0.728	0.445	1.713	1.689
	1	boundary	structured	0.735	0.441	1.244	1.322
			unstructured	0.796	0.403	1.283	1.243
		non-boundary	structured	0.739	0.439	1.250	1.343
			unstructured	0.790	0.408	1.263	1.217
	2	boundary	structured	0.757	0.429	1.168	1.657
			unstructured	0.843	0.363	1.111	1.334
		non-boundary	structured	0.765	0.424	1.192	1.656
			unstructured	0.838	0.369	1.085	1.546
memory	0	boundary	structured	0.796	0.404	1.681	1.490
			unstructured	0.878	0.328	1.805	1.712
		new	structured	0.721	0.449	1.899	1.569
			unstructured	0.782	0.413	1.770	1.666
		non-boundary	structured	0.782	0.414	1.715	1.602
			unstructured	0.904	0.296	1.570	1.291
	1	boundary	structured	0.833	0.374	1.564	2.573
			unstructured	0.911	0.285	1.212	0.959
		new	structured	0.800	0.400	1.420	1.486
			unstructured	0.873	0.333	1.338	1.056
		non-boundary	structured	0.860	0.348	1.618	1.812
			unstructured	0.874	0.332	1.139	0.709
	2	boundary	structured	0.907	0.291	1.196	1.194
			unstructured	0.900	0.301	1.075	0.694
		new	structured	0.751	0.433	1.454	2.322
			unstructured	0.844	0.363	1.292	1.127
		non-boundary	structured	0.893	0.310	1.370	1.869
			unstructured	0.926	0.262	1.045	0.729

Table B.2. Accuracy increases with block during exposure. The table shows an estimate of the block effect on overall accuracy. The hdi does not include 0 implying a reliable increase in accuracy with more exposure.

	mean	sd	hdi 2.5%	hdi 97.5%
Intercept	0.886	0.017	0.852	0.918
block	0.269	0.014	0.240	0.295

Table B.3. Response times decrease with block during exposure. The table shows an estimate of the block effect on overall response times. The hdi does not include 0 implying a reliable decrease in response times with more exposure.

	mean	sd	hdi 2.5%	hdi 97.5%
Intercept	1.624	0.012	1.603	1.649
block	-0.268	0.009	-0.287	-0.251
rt sigma	1.544	0.005	1.534	1.555

Table B.4. No apparent effect of condition on accuracy (hdi for the condition factor includes 0) when accounting for between subject variability through a hierarchical model.

	mean	sd	hdi 2.5%	hdi 97.5%
Intercept	1.193	0.195	0.821	1.583
condition[unstructured]	0.326	0.268	-0.215	0.824
1—participant sigma	0.998	0.096	0.810	1.186

Table B.7. Drift diffusion model parameters for experiment 3a.

parameter, condition, block	mean	sd	hdi 3%	hdi 97%
a structured, 0	1.259	0.022	1.220	1.300
a structured, 1	1.144	0.021	1.104	1.183
a structured, 2	1.121	0.021	1.081	1.159
a unstructured, 0	1.306	0.023	1.263	1.349
a unstructured, 1	1.237	0.025	1.193	1.284
a unstructured, 2	1.149	0.023	1.107	1.191
v boundary, structured, 0	0.385	0.092	0.202	0.546
v boundary, structured, 1	0.476	0.096	0.294	0.653
v boundary, structured, 2	0.688	0.097	0.506	0.873
v boundary, unstructured, 0	0.554	0.108	0.357	0.753
v boundary, unstructured, 1	0.761	0.101	0.565	0.944
v boundary, unstructured, 2	0.830	0.106	0.637	1.029
v new, structured, 0	-0.733	0.062	-0.848	-0.618
v new, structured, 1	-1.124	0.077	-1.267	-0.980
v new, structured, 2	-1.011	0.075	-1.159	-0.879
v new, unstructured, 0	-0.872	0.062	-0.986	-0.754
v new, unstructured, 1	-1.414	0.081	-1.560	-1.259
v new, unstructured, 2	-1.267	0.079	-1.408	-1.112
v non-boundary, structured, 0	0.236	0.069	0.102	0.362
v non-boundary, structured, 1	0.468	0.081	0.326	0.631
v non-boundary, structured, 2	0.570	0.084	0.411	0.730
v non-boundary, unstructured, 0	0.619	0.079	0.471	0.768
v non-boundary, unstructured, 1	0.830	0.092	0.663	1.005
v non-boundary, unstructured, 2	0.927	0.097	0.749	1.114
z 0	0.499	0.010	0.481	0.517
z 1	0.505	0.010	0.485	0.523
z 2	0.489	0.010	0.470	0.507

Table B.8. Bayesian model results for Experiment 3b.

True Distance, Condition	mean	sd	hdi 3%	hdi 97%
1, structured	0.222	0.191	-0.155	0.562
1, unstructured	0.096	0.224	-0.338	0.496
2, structured	-0.026	0.205	-0.401	0.358
2, unstructured	0.113	0.234	-0.329	0.543
3, structured	-0.320	0.280	-0.857	0.184
3, unstructured	-0.505	0.337	-1.137	0.106

APPENDIX C

CHAPTER 4

C.1 Model Statistics

Table C.1. Bayesian model statistics for experiment 4a

Num Features, Condition	mean	sd	hdi 3%	hdi 97%
1.0, structured	0.324	0.308	-0.240	0.927
1.0, unstructured	0.236	0.304	-0.300	0.832
2.0, structured	0.661	0.309	0.077	1.253
2.0, unstructured	0.031	0.304	-0.508	0.621
3.0, structured	0.563	0.311	-0.004	1.152
3.0, unstructured	0.189	0.302	-0.350	0.753
4.0, structured	0.564	0.306	-0.017	1.138
4.0, unstructured	0.126	0.303	-0.444	0.684

Table C.2. Bayesian model statistics for experiment 4b.

Num Features, Category Determinant	mean	sd	hdi 3%	hdi 97%
1.0, A	1.005	0.372	0.303	1.693
1.0, B	-1.529	0.438	-2.361	-0.713
2.0, A	1.338	0.378	0.620	2.036
2.0, B	-1.803	0.443	-2.638	-0.977
3.0, A	1.299	0.378	0.551	1.987
3.0, B	-1.709	0.439	-2.508	-0.848
4.0, A	1.687	0.385	0.977	2.435
4.0, B	-1.706	0.442	-2.596	-0.948

C.2 Feature Differences

Experiment 3b provides an indication that categorization (regardless of if it is dependent on temporal consistency) depends on the importance of each visual feature of the stimuli. At test, participants categorized the previously seen stimulus by

matching its feature to the two on-screen options. On each test trial, the on-screen options differ from each other in one to four features such that for the 'category' option all the category diagnostic features (those that remained temporally consistent during exposure) remained the same as the test stimulus. For the 'non category' option, the features that were not category diagnostic (those that frequently changed during exposure) remained the same as the test stimulus. Thus, in order to categorize based on category diagnostic features, participants must *ignore* the different valued non-category diagnostic features in the category option. Furthermore, participants must use the different category diagnostic feature in the non-category option to decide to not pick that option.

A logistic regression model was fit where each feature was coded as whether it was identical in the non-category option (relative to the test stimulus). If a feature changed in the non category option, (hence was identical in the category option), it was coded as '1' otherwise it was coded as '0'. Fitting this model thus provided a measure of the weights a participant placed on that feature to select the category option.

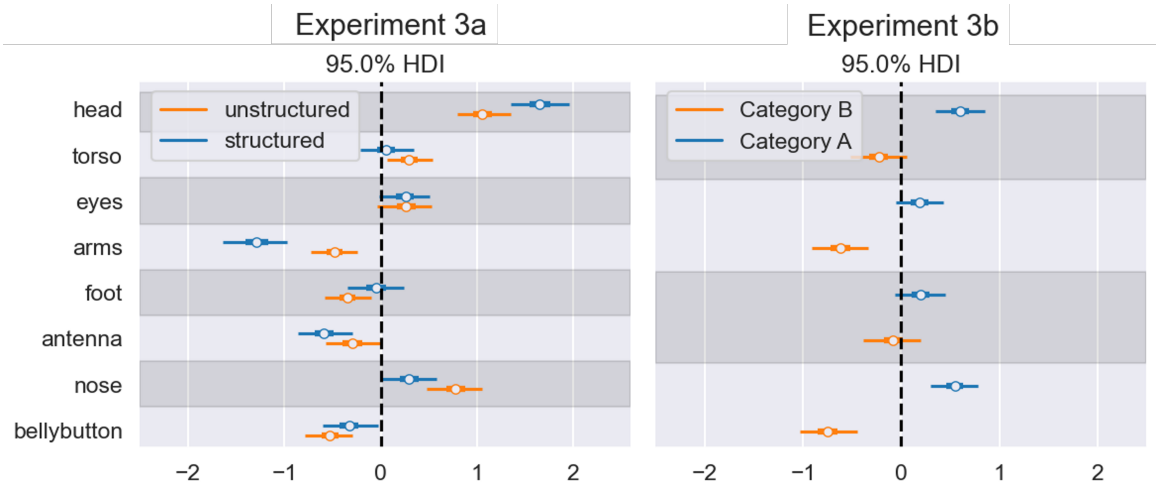


Figure C.1. Relative weights placed by participants to categorize test stimuli. Values above 0 indicate features that are category diagnostic (i.e. remained consistent during exposure) are chosen more often to categorize whereas values below 0 indicate features that are category non-diagnostic (i.e. changed frequently) are used to categorize.

As seen in Figure C.1, some features are grouped together for being temporally consistent whereas some features are grouped together for being temporally inconsistent. Interestingly, this is true even in the unstructured case. Regardless of the temporal exposure, participants will group aliens together if they have the same head color or nose shape. On average, features that were chosen to be category diagnostic for Category A participants in experiment 3b (head, eyes, foot, and nose), were inherently grouped together for sharing feature values. Whereas features that were chosen to be category diagnostic for Category B participants in experiment 3b (torso, arms, antenna, bellybutton) were inherently grouped together if the did *not* share feature values. While this analysis does not address covariances among features, it provides a possible explanation for patterns in Experiment 3b – visual features carry different weights in categorization. Based on their weights, attention may be drawn to them either for being temporally consistent or for being temporally inconsistent. Future modeling will aim to assess whether baseline attention weights modulate the effect of order of presentation in implicit categorization tasks.

BIBLIOGRAPHY

Aslin, R. N., & Newport, E. L. (2012). Statistical learning: From acquiring specific items to forming general rules. *Current directions in psychological science*, *21*(3), 170–176.

Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., & Norman, K. A. (2017). Discovering event structure in continuous narrative perception and memory. *Neuron*, *95*(3), 709–721.

Baldwin, D., Andersson, A., Saffran, J., & Meyer, M. (2008). Segmenting dynamic human action via statistical structure. *Cognition*, *106*(3), 1382–1407.

Behrens, T. E., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. (2007). Learning the value of information in an uncertain world. *Nature neuroscience*, *10*(9), 1214–1221.

Bera, K., Shukla, A., & Bapi, R. S. (2021). Motor chunking in internally guided sequencing. *Brain Sciences*, *11*(3), 292.

Billman, D., & Knutson, J. (1996). Unsupervised concept learning and value systematicitiy: A complex whole aids learning the parts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(2), 458.

Bonner, M. F., & Epstein, R. A. (2021). Object representations in the human brain reflect the co-occurrence statistics of vision and language. *Nature communications*, *12*(1), 4081.

Botvinick, M. M. (2012). Hierarchical reinforcement learning and decision making. *Current opinion in neurobiology*, *22*(6), 956–962.

Brady, T. F., & Oliva, A. (2008). Statistical learning using real-world scenes: Extracting categorical regularities without conscious intent. *Psychological science*, *19*(7), 678–685.

Brunec, I. K., Moscovitch, M., & Barense, M. D. (2018). Boundaries shape cognitive representations of spaces and events. *Trends in cognitive sciences*, *22*(7), 637–650.

Carvalho, P. F., & Goldstone, R. L. (2014). Putting category learning in order: Category structure and temporal arrangement affect the benefit of interleaved over blocked study. *Memory & cognition*, *42*, 481–495.

Carvalho, P. F., & Goldstone, R. L. (2017). The sequence of study changes what information is attended to, encoded, and remembered during category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(11), 1699.

Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology: General*, *120*(3), 235.

Clewett, D., DuBrow, S., & Davachi, L. (2019). Transcending time in the brain: How event memories are constructed from experience. *Hippocampus*, *29*(3), 162–183.

Cohen, A. O., Glover, M. M., Shen, X., Phaneuf, C. V., Avallone, K. N., Davachi, L., & Hartley, C. A. (2022). Reward enhances memory via age-varying online and offline neural mechanisms across development. *Journal of Neuroscience*, *42*(33), 6424–6434.

Cohen, A. L., Nosofsky, R. M., & Zaki, S. R. (2001). Category variability, exemplar similarity, and perceptual classification. *Memory & Cognition*, *29*(8), 1165–1175.

Cowell, R. A., Barense, M. D., & Sadil, P. S. (2019). A roadmap for understanding memory: Decomposing cognitive processes into operations and representations. *Eneuro*, *6*(4).

Cox, G. E., & Criss, A. H. (2020). Similarity leads to correlated processing: A dynamic model of encoding and recognition of episodic associations. *Psychological Review*, *127*(5), 792.

Davis, T., Love, B. C., & Preston, A. R. (2012). Striatal and hippocampal entropy and recognition signals in category learning: Simultaneous processes revealed by model-based fmri. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(4), 821.

Dayan, P. (1993). Improving generalization for temporal difference learning: The successor representation. *Neural computation*, *5*(4), 613–624.

Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., & Pallier, C. (2015). The neural representation of sequences: From transition probabilities to algebraic patterns and linguistic trees. *Neuron*, *88*(1), 2–19.

Dezfouli, A., & Balleine, B. W. (2012). Habits, action sequences and reinforcement learning. *European Journal of Neuroscience*, *35*(7), 1036–1051.

Dezfouli, A., Lingawi, N. W., & Balleine, B. W. (2014). Habits as action sequences: Hierarchical action control and changes in outcome value. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1655), 20130482.

Dolan, R. J., & Dayan, P. (2013). Goals and habits in the brain. *Neuron*, *80*(2), 312–325.

DuBrow, S., & Davachi, L. (2013). The influence of context boundaries on memory for the sequential order of events. *Journal of Experimental Psychology: General*, *142*(4), 1277.

Estes, W. K. (1955). Statistical theory of spontaneous recovery and regression. *Psychological review*, *62*(3), 145.

Ezzyat, Y., & Davachi, L. (2011). What constitutes an episode in episodic memory? *Psychological science*, *22*(2), 243–252.

Ezzyat, Y., & Davachi, L. (2014). Similarity breeds proximity: Pattern similarity within and across contexts is related to later mnemonic judgments of temporal proximity. *Neuron*, *81*(5), 1179–1189.

Fengler, A., Bera, K., Pedersen, M. L., & Frank, M. J. (2022). Beyond drift diffusion models: Fitting a broad class of decision and rl models with hddm. *bioRxiv*, 2022–06.

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 458.

Fisher, A., Rudin, C., & Dominici, F. (2019). All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *Journal of Machine Learning Research*, *20*(177), 1–81.

Fitts, P. M., & Peterson, J. R. (1964). Information capacity of discrete motor responses. *Journal of experimental psychology*, *67*(2), 103.

Gabay, Y., Dick, F. K., Zevin, J. D., & Holt, L. L. (2015). Incidental auditory category learning. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(4), 1124.

Gershman, S. J. (2018). The successor representation: Its computational logic and neural substrates. *Journal of Neuroscience*, *38*(33), 7193–7200.

Gershman, S. J., Moore, C. D., Todd, M. T., Norman, K. A., & Sederberg, P. B. (2012). The successor representation and temporal context. *Neural Computation*, *24*(6), 1553–1568.

Gershman, S. J., & Niv, Y. (2010). Learning latent structure: Carving nature at its joints. *Current opinion in neurobiology*, *20*(2), 251–256.

Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition*, *7*(2), 95–112.

Griffiths, B. J., & Fuentemilla, L. (2020). Event conjunction: How the hippocampus integrates episodic memories across event boundaries. *Hippocampus*, *30*(2), 162–171.

Hahamy, A., Dubossarsky, H., & Behrens, T. E. (2023). The human brain reactivates context-specific past information at event boundaries of naturalistic experiences. *Nature neuroscience*, *26*(6), 1080–1089.

Heusser, A. C., Ezzyat, Y., Shiff, I., & Davachi, L. (2018). Perceptual boundaries cause mnemonic trade-offs between local boundary processing and across-trial associative binding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(7), 1075.

Hicks, J. L., & Starns, J. J. (2006). Remembering source evidence from associatively related items: Explanations from a global matching model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(5), 1164.

Horner, A. J., Bisby, J. A., Wang, A., Bogus, K., & Burgess, N. (2016). The role of spatial boundaries in shaping long-term event representations. *Cognition*, *154*, 151–164.

Howard, M. W., Fotedar, M. S., Datey, A. V., & Hasselmo, M. E. (2005). The temporal context model in spatial navigation and relational learning: Toward a common explanation of medial temporal lobe function across domains. *Psychological review*, *112*(1), 75.

Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in science & engineering*, *9*(03), 90–95.

Kahn, A. E., Karuza, E. A., Vettel, J. M., & Bassett, D. S. (2018). Network constraints on learnability of probabilistic motor sequences. *Nature human behaviour*, *2*(12), 936–947.

Karuza, E. A. (2022). The value of statistical learning to cognitive network science. *Topics in Cognitive Science*, *14*(1), 78–92.

Karuza, E. A., Kahn, A. E., & Bassett, D. S. (2019). Human sensitivity to community structure is robust to topological variation. *Complexity*, *2019*(1), 8379321.

Karuza, E. A., Kahn, A. E., Thompson-Schill, S. L., & Bassett, D. S. (2017). Process reveals structure: How a network is traversed mediates expectations about its architecture. *Scientific reports*, *7*(1), 12733.

Knowlton, B. J., Ramus, S. J., & Squire, L. R. (1992). Intact artificial grammar learning in amnesia: Dissociation of classification learning and explicit memory for specific instances. *Psychological science*, *3*(3), 172–179.

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological science*, *19*(6), 585–592.

Kornell, N., Castel, A. D., Eich, T. S., & Bjork, R. A. (2010). Spacing as the friend of both memory and induction in young and older adults. *Psychology and aging*, *25*(2), 498.

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, *2*, 249.

Kruschke, J. K. (2020). Alcove: An exemplar-based connectionist model of category learning. In *Connectionist psychology* (pp. 107–138). Psychology Press.

Lositsky, O., Chen, J., Toker, D., Honey, C. J., Shvartsman, M., Poppenk, J. L., Hasson, U., & Norman, K. A. (2016). Neural pattern change during encoding of a narrative predicts retrospective duration estimates. *elife*, *5*, e16070.

Love, B. C. (2002). Comparing supervised and unsupervised category learning. *Psychonomic bulletin & review*, *9*(4), 829–835.

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). Sustain: A network model of category learning. *Psychological review*, *111*(2), 309.

Luce, R. D. (1977). The choice axiom after twenty years. *Journal of mathematical psychology*, *15*(3), 215–233.

Lynn, C. W., & Bassett, D. S. (2020). How humans learn and represent networks. *Proceedings of the National Academy of Sciences*, *117*(47), 29407–29415.

Lynn, C. W., Kahn, A. E., Nyema, N., & Bassett, D. S. (2020). Abstract representations of events arise from mental errors in learning and memory. *Nature communications*, *11*(1), 2313.

Lynn, C. W., Papadopoulos, L., Kahn, A. E., & Bassett, D. S. (2020). Human information processing in complex networks. *Nature Physics*, *16*(9), 965–973.

Mack, M. L., & Palmeri, T. J. (2015). The dynamics of categorization: Unraveling rapid categorization. *Journal of Experimental Psychology: General*, *144*(3), 551.

Marr, D., & Poggio, T. (1976). From understanding computation to understanding neural circuitry.

McDougle, S. D., Bond, K. M., & Taylor, J. A. (2015). Explicit and implicit processes constitute the fast and slow processes of sensorimotor learning. *Journal of Neuroscience*, *35*(26), 9568–9579.

McDougle, S. D., & Collins, A. G. (2021). Modeling the influence of working memory, reinforcement, and action uncertainty on reaction time and choice during instrumental learning. *Psychonomic bulletin & review*, *28*, 20–39.

Medin, D. L., & Bettger, J. G. (1994). Presentation order and recognition of categorically related examples. *Psychonomic bulletin & review*, *1*(2), 250–254.

Medin, D. L., Wattenmaker, W. D., & Hampson, S. E. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive psychology*, *19*(2), 242–279.

Mensink, G.-J., & Raaijmakers, J. G. (1988). A model for interference and forgetting. *Psychological Review*, *95*(4), 434.

Michelmann, S., Hasson, U., & Norman, K. A. (2023). Evidence that event boundaries are access points for memory retrieval. *Psychological Science*, *34*(3), 326–344.

Momennejad, I. (2020). Learning structures: Predictive representations, replay, and generalization. *Current Opinion in Behavioral Sciences*, *32*, 155–166.

Momennejad, I., Russek, E. M., Cheong, J. H., Botvinick, M. M., Daw, N. D., & Gershman, S. J. (2017). The successor representation in human reinforcement learning. *Nature human behaviour*, *1*(9), 680–692.

Murdock, B. B. (1997). Context and mediators in a theory of distributed associative memory (todam2). *Psychological Review*, *104*(4), 839.

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive psychology*, *19*(1), 1–32.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of experimental psychology: General*, *115*(1), 39.

Nosofsky, R. M. (2011). The generalized context model: An exemplar model of classification. *Formal approaches in categorization*, 18–39.

Nosofsky, R. M., Gluck, M. A., Palmeri, T. J., McKinley, S. C., & Glauthier, P. (1994). Comparing modes of rule-based classification learning: A replication and extension of shepard, hovland, and jenkins (1961). *Memory & cognition*, *22*(3), 352–369.

Nosofsky, R. M., Little, D. R., Donkin, C., & Fific, M. (2011). Short-term memory scanning viewed as exemplar-based categorization. *Psychological review*, *118*(2), 280.

Nosofsky, R. M., Sanders, C. A., Gerdom, A., Douglas, B. J., & McDaniel, M. A. (2017). On learning natural-science categories that violate the family-resemblance principle. *Psychological science*, *28*(1), 104–114.

Nosofsky, R. M., Sanders, C. A., Meagher, B. J., & Douglas, B. J. (2018). Toward the development of a feature-space representation for a complex natural category domain. *Behavior research methods*, *50*, 530–556.

Osth, A. F., & Dennis, S. (2020). Global matching models of recognition memory.

Ostlund, S. B., Winterbauer, N. E., & Balleine, B. W. (2009). Evidence of action sequence chunking in goal-directed instrumental conditioning and its dependence on the dorsomedial prefrontal cortex. *Journal of Neuroscience*, *29*(25), 8280–8287.

Peirce, J. W. (2007). Psychopy—psychophysics software in python. *Journal of neuroscience methods*, *162*(1-2), 8–13.

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in cognitive sciences*, *10*(5), 233–238.

Pitt, B. S. (2018). *Metaphorical mappings of time and number: How cultural experience shapes cognitive universals*. The University of Chicago.

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological review*, *116*(1), 129.

Radvansky, G. A., & Zacks, J. M. (2017). Event boundaries in memory and cognition. *Current opinion in behavioral sciences*, *17*, 133–140.

Ratcliff, R., Scharre, D. W., & McKoon, G. (2022). Discriminating memory disordered patients from controls using diffusion model parameters from recognition memory. *Journal of Experimental Psychology: General*, *151*(6), 1377.

Ratcliff, R., & Starns, J. J. (2009). Modeling confidence and response time in recognition memory. *Psychological review*, *116*(1), 59.

Ratcliff, R., Thapar, A., & McKoon, G. (2004). A diffusion model analysis of the effects of aging on recognition memory. *Journal of Memory and Language*, *50*(4), 408–424.

Roark, C. L., & Holt, L. L. (2018). Task and distribution sampling affect auditory category learning. *Attention, Perception, & Psychophysics*, *80*, 1804–1822.

Roark, C. L., Lehet, M. I., Dick, F., & Holt, L. L. (2022). The representational glue for incidental category learning is alignment with task-relevant behavior. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(6), 769.

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(6), 906–914.

Rouder, J. N., & Ratcliff, R. (2004). Comparing categorization models. *Journal of Experimental Psychology: General*, *133*(1), 63.

Rouhani, N., Norman, K. A., & Niv, Y. (2018). Dissociable effects of surprising rewards on learning and memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(9), 1430.

Rouhani, N., Norman, K. A., Niv, Y., & Bornstein, A. M. (2020). Reward prediction errors create event boundaries in memory. *Cognition*, *203*, 104269.

Russek, E. M., Momennejad, I., Botvinick, M. M., Gershman, S. J., & Daw, N. D. (2017). Predictive representations can link model-based reinforcement learning to model-free mechanisms. *PLoS computational biology*, *13*(9), e1005768.

Savalia, T., Cowell, R. A., & Huber, D. E. (2022). "leap before you look": Conditions that promote implicit visuomotor adaptation without explicit learning. *bioRxiv*, 2022–07.

Savalia, T., Cowell, R. A., & Huber, D. E. (2024). "leap before you look": Conditions that suppress explicit, knowledge-based learning during visuomotor

adaptation. *Journal of Experimental Psychology: Human Perception and Performance*.

Savalia, T., Shukla, A., & Bapi, R. S. (2016). A unified theoretical framework for cognitive sequencing. *Frontiers in psychology*, *7*, 1821.

Schapiro, A., & Turk-Browne, N. (2015). Statistical learning. *Brain mapping*, *3*(1), 501–506.

Schapiro, A. C., Rogers, T. T., Cordova, N. I., Turk-Browne, N. B., & Botvinick, M. M. (2013). Neural representations of events arise from temporal community structure. *Nature neuroscience*, *16*(4), 486–492.

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological monographs: General and applied*, *75*(13), 1.

Shin, Y. S., & DuBrow, S. (2021). Structuring memory through inference-based event segmentation. *Topics in Cognitive Science*, *13*(1), 106–127.

Smith, K. S., & Graybiel, A. M. (2016). Habit formation. *Dialogues in clinical neuroscience*, *18*(1), 33–43.

Smith, M. A., Ghazizadeh, A., & Shadmehr, R. (2006). Interacting adaptive processes with different timescales underlie short-term motor learning. *PLoS biology*, *4*(6), e179.

Sols, I., DuBrow, S., Davachi, L., & Fuentemilla, L. (2017). Event boundaries trigger rapid memory reinstatement of the prior events to promote their representation in long-term memory. *Current Biology*, *27*(22), 3499–3504.

Stachenfeld, K. L., Botvinick, M. M., & Gershman, S. J. (2017). The hippocampus as a predictive map. *Nature neuroscience*, *20*(11), 1643–1653.

Starns, J. J. (2014). Using response time modeling to distinguish memory and decision processes in recognition and source tasks. *Memory & cognition*, *42*, 1357–1372.

Starns, J. J., & Ratcliff, R. (2014). Validating the unequal-variance assumption in recognition memory using response time distributions instead of roc functions: A diffusion model analysis. *Journal of memory and language*, *70*, 36–52.

Sutton, R. S. (2018). Reinforcement learning: An introduction. *A Bradford Book*.

Swallow, K. M., Zacks, J. M., & Abrams, R. A. (2009). Event boundaries in perception affect memory encoding and updating. *Journal of Experimental Psychology: General*, *138*(2), 236.

Tremblay, P.-L., Bedard, M.-A., Langlois, D., Blanchet, P. J., Lemay, M., & Parent, M. (2010). Movement chunking during sequence learning is a dopamine-dependant process: A study conducted in parkinson's disease. *Experimental brain research*, *205*, 375–385.

Turk-Browne, N. B. (2019). The hippocampus as a visual area organized by space and time: A spatiotemporal similarity hypothesis. *Vision research*, *165*, 123–130.

Turk-Browne, N. B., Isola, P. J., Scholl, B. J., & Treat, T. A. (2008). Multidimensional visual statistical learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(2), 399.

Unger, L., & Sloutsky, V. M. (2022). Ready to learn: Incidental exposure fosters category learning. *Psychological Science*, *33*(6), 999–1019.

Unger, L., Weichart, E. R., Reardon, N., & Sloutsky, V. (2023). Without even trying: How incidental exposure shapes category learning. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *45*.

Vlach, H. A., Sandhofer, C. M., & Kornell, N. (2008). The spacing effect in children's memory and category induction. *Cognition*, *109*(1), 163–167.

Wahlheim, C. N., Dunlosky, J., & Jacoby, L. L. (2011). Spacing enhances the learning of natural concepts: An investigation of mechanisms, metacognition, and aging. *Memory & cognition*, *39*, 750–763.

Whitehead, P. S., Zamary, A., & Marsh, E. J. (2021). Transfer of category learning to impoverished contexts. *Psychonomic Bulletin & Review*, 1–10.

Wilson, J. P., Hugenberg, K., & Bernstein, M. J. (2013). The cross-race effect and eyewitness identification: How to improve recognition and reduce decision errors in eyewitness situations. *Social Issues and Policy Review*, *7*(1), 83–113.

Wixted, J. T., & Mickes, L. (2014). A signal-detection-based diagnostic-feature-detection model of eyewitness identification. *Psychological Review*, *121*(2), 262.

Young, S. G., Hugenberg, K., Bernstein, M. J., & Sacco, D. F. (2012). Perception and motivation in face recognition: A critical review of theories of the cross-race effect. *Personality and Social Psychology Review*, *16*(2), 116–142.

Zacks, J. M. (2020). Event perception and memory. *Annual review of psychology*, *71*(1), 165–191.

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological bulletin*, *133*(2), 273.

Zacks, J. M., & Swallow, K. M. (2007). Event segmentation. *Current directions in psychological science*, *16*(2), 80–84.