Representation of Uncertainty and Recall Precision in Long-Term Episodic and Semantic Memories

Dávid Ádám Magas

Supervisor: József Fiser

Secondary Supervisor: Máté Lengyel

Submitted to Central European University

Department of Cognitive Science

In partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Cognitive Science

Budapest, Hungary

2025

Declaration of Authorship

I, the undersigned, Dávid Ádám Magas, candidate for the PhD degree in cognitive science declare herewith that the present thesis titled "**Representation of Uncertainty and Recall Precision in Long-Term Episodic and Semantic Memories**" is exclusively my own work, based on my research and only such external information as properly credited in notes and bibliography. I declare that no unidentified and illegitimate use was made of the work of others, and no part of the thesis infringes on any person's or institution's copyright. I also declare that no part of the thesis has been submitted in this form to any other institution of higher education for an academic degree.

One part of the dissertation contains work which was done in collaboration.

- Chapter 2 (Experiment 1) with Ádám Ferdinánd Koblinger

Budapest, 30 June 2025

Dávid Ádám Magas

Copyright Notice

Acknowledgments

There are some people that I would like to mention who helped me throughout my PhD studies.

First, I would like to thank my supervisor, József Fiser, for his help and assistance in the last few years. His remarks and guidance certainly helped me work in a more precise and thorough way.

I would like to thank Máté Lengyel for providing useful ideas and inputs for my dissertation.

I would also like to thank the members of the CEU visionlab for providing a stimulating research environment. Special thanks to Ádám Koblinger and Márton Nagy for helping me with ideas and research organization.

I am thankful to Réka Finta for constant and immediate help in connection with the administrative side of the PhD journey.

Further, I would like to thank Bálint Lente and Zoltán Molnár, two wonderful friends, for providing fun, cheerful discussions and time, outside of my research environment, and taking an interest in what it is that I am doing.

Thanks to Vani, for many unforgettable memories.

Last but definitely not least, I would like to thank my parents for their immense support in every aspect of my life. I am one very lucky person.

Abstract

Episodic memory has often been characterized as detailed autonoetic awareness of one's past events. In my dissertation, I reconceptualize episodic memory as part of a general knowledge structure or long-term semantic memory. I offer a common framework in which the recall precision and the representation of uncertainty in short-term and long-term episodic and semantic memory can be investigated. As a result, my work bridges important gaps between perception, long-term episodic and semantic memory, and provides insights into the detailed form in which items in perception and long-term memory are encoded and recalled.

In Chapter 2, I analyze recall precision and the representation of uncertainty in perceptual decision-making and in long-term episodic memories without any semantic regularity imposed on them. I show that items in perception and long-term episodic memory are encoded and recalled in a probabilistic manner. In Chapter 3, I organize episodic elements into simple scenes with both perceptual and semantic connections between the elements. I demonstrate that semantic connections are dominant as opposed to perceptual ones in increasing recall precision. Furthermore, I show that the structure in which scene elements are stored in long-term memory corresponds to the recurring input schema of the scenes. In Chapter 4, I introduce overarching semantic regularity into the input and analyze how it affects recall precision and the representation of uncertainty. I show that semantic regularity improves overall recall precision. In addition, I show that this increase was a result of true semantic learning, where people learnt the structure of the input and used that knowledge exclusively in several responses. Furthermore, I point out major individual differences in episodic and semantic learning ability across participants. Lastly, I

show that the fundamentally probabilistic representation of individual items does not change despite learning the overarching semantic regularity. In Chapter 5, I analyze the effect of attention on episodic and semantic learning and show that semantic but not episodic learning remains intact with divided attention.

Table of Contents

CHAPTER 1

INTRODUCTION

Episodic memory is arguably one of the broadest topics when it comes to areas in cognitive investigation and it has been studied in several domains, such as visual and auditory processing (Koeritzer et al., 2018; Lad et al., 2020; Leist, Lachmann & Klatte, 2025). It has been characterized both as short-term and long-term episodic memory (Cowan, 2008; Wood, Baxter & Belpaeme, 2012), and has been linked to semantic memory, attention and language (Baddeley, 2000; Baddeley, 2003; Schroeder & Marian, 2012; Irish & Piguet, 2013; Wang et al., 2025). Despite its broad nature, episodic memory can be characterized specifically as a partially distinct memory system that is used for storing specifics and details about all types and levels of memories (such as natural objects, scenes, or whole episodes) (Rubin & Umanath, 2015; Zotow, Brisby & Burgess, 2020). Classical definition of episodic memory describes: a subtype of long-term memory that involves the conscious and explicit recollection of specific events or experiences with their contextual details, the famous “what”, “where”, “when” formulation of episodic memory (Tulving, 2002).

One way to investigate episodicness is to study autobiographical memories characterized by autonoetic awareness of detailed past events (Tulving, 1993; Tulving, 2002). However, recently there has been a shift and expansion in understanding the nature and relevance of episodic memory in multiple processes in the brain that have led to the emergence of the concept of generative episodic memory: viewing it as the active process of constructing rather than retrieving exact episodic events, by flexibly recombining stored memory elements, blending actual experiences and inferred (semantic) details (Schacter, Addis & Buckner, 2007). Further developments occurred that downplayed the role of “self” and "subjective time" by emphasizing the strong overlap in active brain networks during remembering the past and other processes such as navigation, thinking about the future, and imaging fictious experiences (Hassabis & Maguire, 2007). In this thesis, I focus on the aspect of episodic memory that is more closely related to general knowledge representation in the brain rather than the classical conscious retrieval of specific autobiographical self-centered events from memory. This aspect is closely related to the question of how solving any task is based on generalization where some stored episodic and semantic snippets are used in a constructive manner in a novel task situation. Such a formulation of the approach to episodic memory is supported by the strong functional analogy between the conceptualization of executing a perceptual task and generative episodic memory retrieval.

The functional analogy

Stage	Perceptual Task	Episodic memory
Creating/Obtaining Stored Knowledge	Sensory experiences and learned patterns build a library of perceptual and semantic "snippets"	Episodic experiences and semantic knowledge are encoded and stored as memory traces
Using it	Retrieve relevant perceptual and conceptual features to interpret new input	Retrieve specific episodic details and semantic scaffolds to reconstruct or simulate a memory
Generalization	Blend prior percepts and concepts to recognize or interpret novel stimuli	Flexibly recombine memory fragments to recall or imagine related events or predict future scenarios
End product	A coherent perception or categorization of a new situation	A constructed or simulated episodic scenario with contextual detail and narrative coherence

Figure 1.1 Investigating generative episodic (episodic memory incorporated into semantic knowledge) memory in perception and long-term memory.

In more detail, in this dissertation, episodic and semantic memories are imagined as components of a hierarchical (generative) framework where semantic memory is equivalent to a probabilistic internal model of the world. The main objective of the dissertation is to behaviorally investigate the form in which episodic and semantic memories are represented and uncover the interaction between the two types of memory systems. Throughout this dissertation, the two major topics that are covered in connection with long-term visual memories are recall precision and the subjective representation of uncertainty in memories. The main goal is to offer a comprehensive and precise computationally motivated behavioral framework in which visual long-term memories with various complexities can be investigated. Although there is no computational modelling presented in the dissertation, probabilistic/computational ideas manifest in the design of the experiments, particularly in measuring uncertainty. Consequently, experimental results from the thesis can support or question already existing computational ideas about episodic and semantic memory.

According to the dominant views in the field, perceptual processes are best formulated as "unconscious inference" (Dayan et al., 1995) where all information is handled and combined with its uncertainty operationalized by probabilistic computation in the brain. This conceptualization must assume that representations of knowledge stored in long-term memory must also be probabilistic (Steyvers, Griffiths & Dennis, 2006; Fiser et al., 2010). However, long-term memories are derived from short-term episodic snippets and as argued above, generative episodic memories themselves are utilized in a process similar to using knowledge encoded in long-term memory. This leaves open the question of what the format of episodic memory is in building long-term memory and/or in solving tasks. Is it probabilistic, or is it highly faithful to the original input? To establish a feasible connection between the two types of memories the link between the representation of uncertainty and episodic memories must be clarified. Therefore, the first major goal of the dissertation is to establish whether there is a representation of uncertainty in visual episodic memory during perceptual decision-making and long-term recall tasks.

Apart from the representation of uncertainty in episodic memories, the concept of semantic memory and the interaction between episodic and semantic memory is also treated vaguely within the long-term memory literature. It is not straightforward what is meant by semantic regularity and representation as it can take several forms (associative connections, high-level statistical regularity) just like episodic memory can be interpreted on several levels of cognition. Therefore, it is unclear what people learn exactly when they acquire semantic knowledge, and how episodic memories interact with certain types of semantic regularities. If episodic memory is to be incorporated into the probabilistic internal models of perception, then a thorough investigation of probabilistic uncertainty representations of episodic and semantic memory must be also carried out to get a deeper understanding of the dynamics and potential interconnectedness of the two seemingly separate memory systems. Consequently, the second major goal of the dissertation is to study episodic and semantic learning along with distinct kinds of semantic regularities in the input.

In addition to these fundamental theoretical questions, there are also unavoidable methodological issues linked to investigating related topics. Working and long-term memory are investigated by distinct empirical methods (Rahnev et al., 2020), and the position of non-classical episodic memories is straddling between those two, raising the question of how to investigate generative episodic memories. To address this, I present a novel general paradigm in which both the content of the memory and the observer's uncertainty are investigated jointly.

The current chapter presents the main topics and arguments in the dissertation and places it within the long-term memory literature. First, I present a method that was developed in perceptual decision making to measure the recall accuracy and subjective uncertainty in simple perceptual representations, then I apply this method and connect it to simple long-term episodic memory. Then increasingly complex and distinct types of regularities (semanticity) are systematically imposed on episodic items, starting with local associations in simple scenes, then moving on to overarching statistical regularity, observing its effects on recall accuracy and uncertainty representation. Lastly, noise and varying levels of attention are added to the input, in order to give an even more comprehensive understanding of the dynamics of episodic and semantic memory, in ecologically valid settings.

1.1 Recall Accuracy and Representation of Uncertainty in Simple Episodic Memories

An overall, clear-cut definition of episodic memory in the literature is missing, partly because episodic memory can appear in more than one form in cognition. This dissertation approaches episodic memory from the aforementioned semantic internal model building perspective and investigates primarily its functionality, in that context. Consequently, here episodic memories are not whole past events that connect to autonoetic awareness but smaller elements of long-term memory, mostly real-world objects. In more detail, in a case when an object detail (such as orientation, color) is recalled with high enough precision and it is not plagued by semantic regularity, then it is understood as an episodic memory trace.

When it comes to long-term visual memories, one of the most investigated questions is: after encoding, what remains in long-term memory, and how much can people store from the incoming stimuli? There have been several arguments for the so-called 'massive' nature of visual long-term memory (Standing, 1973; Brady et al., 2008; Miner et al., 2020; Wolfe et al., 2023). The recurring general argument in connection with massivity and its meaning is that people can remember details in long-term memory, and in certain contexts, a large number of the recalled details are precise. However, massivity can heavily depend on several things such as input structure (Konkle et al., 2010b; Shoval et al., 2023; Persaud et al., 2024), recall testing method and interpretation (Yonelinas, 2002; Brady et al., 2008; Wolfe at al., 2023). For instance, some of the beforementioned studies used input that varied in complexity and in terms of how meaningful it was. In addition, some used free recall (Wolfe at al., 2023), others used an alternative forced choice method (Brady et al., 2008) to test memories, which aligns with recall being based only on familiarity or on true recollection (Yonelinas, 2002). In Chapters 2 and 3, I argue and confirm previous results that massive and precise recall is dependent on context and interpretation. More importantly, if episodic memories are interpreted as parts of semantic internal models, then the question arises: do they need to be massive in the first place. Chapter 4 shows that while not all episodes are forgotten, one of their primary function is to build semantic internal models.

Partly related to massivity, another topic that is investigated in connection with long-term visual memories is boundedness. This line of study revolves around whether parts of a memory (such as parts of a real-world object like color, state etc.), are connected or separate in long-term memory representations. There are arguments for independent storage or representation for object parts (e.g., Brady et al., 2013). At the same time, there are arguments for bounded or holistic object representation (Horner et al., 2015; Balaban et al., 2020). However, just like massivity, boundedness can also be context dependent (Kuhbandner, 2020). The dominant view in the field seems to be that the default form of representing object parts in long-term memory is independent, but whenever a memory item is more meaningful or task relevant then there is more binding between its elements (Utochkin & Brady, 2019; Markov et al., 2021). In Chapter 2, I utilize subjective certainty reports to analyze the boundedness of object features and argue for default boundedness, at least at the level of uncertainty representation. Apart from recall accuracy, another major line of study is concerned with subjective confidence or the representation of uncertainty in long-term visual memory. The representation of subjective certainty has been investigated in various kinds of memories with different age groups. It has been shown that some kind of meta-awareness for the correctness of memories increasingly emerges early in childhood, around 3-6 years (Hembacher & Ghetti, 2014; Liu et al., 2018). Further, there have been some arguments for the fallibility of people's memories and consequently, about false confidence in those memories (Dodson & Krueger, 2006; Dodson & Krueger, 2007; Kim & Kabeza, 2007; Bona & Silvanto, 2014). However, these studies include extreme cases, where memories were highly typical but false examples, or where the fallibility of subjective memory judgements were connected to age as well. Other studies show that humans utilize uncertainty appropriately for optimal decisions (Körding & Wolpert, 2004; Fiser et al., 2010; Ma & Jazayeri, 2014; Lengyel et al., 2015; Li et al., 2021; Koblinger, Fiser & Lengyel, 2021).

Nevertheless, the representation of uncertainty has been investigated more in perceptual than in long-term visual memory studies. Therefore, a thorough investigation of the representation of subjective uncertainty in long-term episodic and semantic memory is valuable for several reasons. First, there is a shortage of studies that investigate uncertainty in long-term visual memory from a computationally motivated standpoint. Second, partly due to the lack of studies, the connection between perception and long-term memory in terms of uncertainty representation is also missing. Further, the method with which subjective uncertainty is tested can matter a lot. A discrete high-level, language-based testing method can paint a different picture about people's confidence judgement than a low-level continuous testing method. Importantly, not only is there a shortage of studies that investigate subjective uncertainty in long-term memory, the method with which uncertainty is tested also varies and differs noticeably in perception and in long-term memory research. Proportionally, there are a lot more perceptual studies that apply a continuous scale when testing uncertainty, whereas long-term memory research tends to stay with discrete scales (Rahnev et al., 2020).

Besides the connection between visual perception and long-term memory, and the testing method, it is vital to state what a given study means by the representation of uncertainty. It is because uncertainty can potentially be represented in several ways such as discretely (low, mid, high uncertainty). It can be a point estimate for the variance of a target variable. In this dissertation, the way the representation of uncertainty is conceptualized and computationally motivated is that upon recall, several values of a certain variable are represented with corresponding probability values for each. For example, when the task is to recall the orientation of a given object, then when recalled, several orientation values are represented with corresponding probability values. Testing uncertainty in episodic and semantic memories from this approach allows to establish whether episodic and semantic memories are encoded and treated probabilistically. In Chapter 2 and Chapter 3, using a fine-grained, continuous testing method for the representation of subjective uncertainty, I argue for the probabilistic nature of simple long-term episodic memories, and I also show that the probabilistic manner of representation is not modularity dependent i.e.: it is the same in perceptual decision-making and in long-term episodic memory.

1.2 Perceptual and Semantic Regularity in Simple Scenes Containing Natural Objects

Studying episodic memory for real-world objects alone is useful, as it is one of the simplest potential forms of long-term episodic memory, however, real life episodes often contain more complex and structured elements such as scenes or events. This dissertation investigates recall accuracy and the representation of uncertainty in episodic memories with increasingly more complex input regularity. After studying simple, individual objects, the next logical step is to consider simple scenes, or, in this case, multiple objects presented simultaneously in one display. One topic that is studied in the literature, and is interesting in its own right, is whether the simple fact that objects are organized into a scene changes anything compared to when they appear on their own. This topic is connected to the previously mentioned massivity, and there are arguments that memory for scenes is detailed as well (Konkle et al., 2010a). More importantly, once episodic memories become more complex (for example, scenes instead of individual objects), the concept of the 'gist' emerges. Simply put, the gist means quick extraction of a meaningful, overall idea about a complex input (e.g., Reyna & Brainerd, 1998) oftentimes by combining lower-level perceptual features with higher-level semantic ones (Oliva, 2005; Oliva, 2006).

When elements such as real-world objects are organized into a structure i.e. a scene, one of the most important questions is how certain scene schemas and structure alter people's encoding, representation and recall performance for the individual elements. Scenes structures and schemas can come in several forms such as object category co-occurrences (Brady & Oliva, 2008), higher order-structure or ensembles (Brady, Konkle & Alvarez, 2011; Brady & Tanenbaum, 2013; Haberman, Brady & Alvarez, 2015; Persaud & Hemmer, 2024) or recurring scene schemas (Richter at al., 2019; Hu & Jacobs, 2021). These studies demonstrate the key idea that people are sensitive to and effective in extracting several kinds of regularities from the input which a scene or other structures facilitate. Further, it has been shown that people can allocate their encoding resources as a function of the acquired semantic regularity (Bates et al., 2019). However, many of the previously mentioned studies analyzed working memory in terms of encoding various kinds of scenes with structure. What is less known is how certain types of scene regularities contribute to later recall precision in a long-term memory task. In Chapter 3, I introduce both perceptual and semantic regularities to simple scenes and treat them on equal terms in encoding. Namely, neither type of regularity is explicitly prioritized by scene inspection time or task requirements. By doing so, I demonstrate the importance of semantic structure and meaning in the input showing that semantic connections facilitate more precise long-term encoding than perceptual connections in scenes. Regarding the scene structure, another important but even less investigated question is the form in which elements are encoded from a scene with complex regularity. In Chapter 2, I argue for bounded representation in the case of single objects. However, boundedness can also be interpreted in the context of scenes, and it is unknown whether the principle of boundedness or unboundedness apply when the input is structured in the form of a scene. In Chapter 3, I show that elements of scenes that are connected through sufficiently complex input structure, behave as chunks in people's representation. In this sense, an opposite principle of unboundedness applies for complex scenes as opposed to the bounded nature of object representation.

As the input becomes more complex, not only memory precision for the parts of the scene but the representation of uncertainty can also change. The relationship of episodic and semantic memory in naturalistic environments has been studied previously (Hemmer & Steyvers, 2012). However, the uncertainty aspect of the representation is less highlighted. Importantly, analyzing the representation of uncertainty in simple scenes with real world objects allows for the most accurate connection between perceptual decision-making and long-term episodic memory. It is because usually, in perceptual studies, there are several elements in the display. This way, a further question can be asked about the representation of uncertainty in long-term memory for scenes: is it item-based or general/scene based. In other words, is there a separate representation of uncertainty for each scene element in long-term memory, just like in perceptual decision-making (Lengyel et al., 2015), or only a global uncertainty representation for the whole scene is stored. Further, an important question is whether semantic scene regularity alters or biases the fundamental form of uncertainty representation for scene elements. In Chapter 3, I argue that without scene regularity, uncertainty is represented in an item-based manner, which serves as proof for further similarity between perceptual decision-making and long-term episodic memory representations. On the other hand, Chapter 3 provides evidence that as the complexity of scene regularity increases, the representation of uncertainty for individual scene elements becomes biased, nevertheless remains item-based and fundamentally probabilistic.

1.3 Semantic Regularity in the Form of Long-Term, Overarching Statistical Pattern

Another important type of regularity/memory is overarching semantic regularity. Previously, semantic regularity was understood to be local, scene based, with recurring scene schemas or scene ensembles. Showing that semanticity affects item-based recall in a scene-based, working (or long-term) memory task can be taken as proof that people are capable of learning and using internal models to interpret incoming stimuli. However, a more general kind of regularity that spans through all stimuli in a memory experiment, is arguably closer to what is understood to be long-term prior knowledge and to what is taken to be long-term semantic memory. Consequently, the interplay of episodic and semantic memory can be best interpreted in that context.

Considering long-term episodic and semantic memory, studies first intended to investigate whether the two are separate or intertwined systems (Graham et al., 2000; Tulving, 2002; Sadeh et al., 2016; Schapiro et al., 2017; Renoult et al., 2019). The consensus is that episodic and semantic memory are partly distinct systems, but they affect and are connected to each other. Numerous studies investigated the interaction between episodic and semantic memory (Schacter, Addis & Buckner, 2007; Hemmer & Steyvers, 2009; Hemmer & Steyvers, 2009; Greenberg & Verfaellie, 2010; Steyvers & Hemmer, 2012; Brod, Werkle-Bergner & Shing, 2013; Hemmer & Persaud, 2014; Bae et al., 2015; Brady et al., 2018; Fang et al., 2018; Tompary & Thompson-Schill, 2021; Xu et al., 2024). These studies demonstrate one recurring feature, which is also confirmed in the domain of working memory research, that general semantic knowledge biases (sometimes heavily) individual episodic memory traces. For example, when the size of objects (such as fruits and vegetables in Hemmer & Steyvers (2009)) are studied and remembered, people usually recall the size of a particular fruit or vegetable closer to the category mean, in other words, recall is semantically biased. In addition, the function of episodic memory in the connection is to support the building of semantic knowledge and update it (Nagy & Orbán, 2016) whereas acquired semantic memory allows people to make meaningful guesses about individual items for which episodic memory is weak or lost (e.g., Nagy et al., 2020).

While long-term semantic memory biases episodic memory, it is not fully known whether this bias is beneficial or not. Or at least the exact nature of the bias is not known. Some argue that memory distortions improve recall and/or they are adaptive, and, in that sense, they are beneficial (Hemmer & Steyvers, 2009; Schacter et al., 2011). In addition, Huttenlocker et al. (2000) argued that bias is a part of a Bayesian recall procedure, intended to maximize accuracy. In Chapter 4, first I show that overarching semantic regularity boosts overall recall. Thus, even though individual memories become more biased with pervasive semantic regularity, ultimately, these memories are recalled more precisely than memories in a task without overarching semantic regularity. Importantly, the exact nature of the boosts is not known. In particular, it is not known how much semantic regularity people actually learn in a fairly simple memory setup. Specifically, it has not been investigated whether an observer's performance increase due to semantic input regularity is a consequence of true semantic learning (acquiring and representing input regularity) or alternatively, people simply remember individual items better episodically. Furthermore, it is not fully known that if true semantic learning happens, in what form is it represented. There exist studies showing that people learn full distributions and characteristics of distributions, when it comes to semantic memory (Chetverikov et al., 2017; Chetverikov et al., 2017; Chetverikov et al., 2020). In these studies, people were able to learn the detailed shape of distractor distributions (such as bimodal) in a working memory task. However, it is not known how semantic learning, in a more general, long-term memory task, behaves. Furthermore, it is not known whether and how individual differences in long-term episodic and semantic learning manifest themselves. In Chapter 4, I argue that the increased overall memory performance with semantic input regularity is due to true semantic learning. People learn the characteristics of the input distribution and use that exclusively in certain responses. Importantly, this behavior characterizes only a subset of the participants, consequently, there are significant individual differences in long-term episodic and semantic learning.

In order to investigate episodic and semantic memory thoroughly in a common framework, the representation of uncertainty in memories that contain overarching semantic regularity should also be considered. This topic is even less investigated than the representation of uncertainty in pure episodic memory. Regarding uncertainty in long-term semantic memory, the few existing studies either discuss the potential probabilistic nature of the representation (Griffiths & Steyvers, 2002) or they show that high-confidence errors can occur in “false” (semantically constructed) memories. What is missing is investigating recalled memory items that are partly or fully semantic constructions and comparing the representation of uncertainty in those cases to the ones where there is no overarching semanticity imposed on the input. Although, relatedly, Persaud & Hemmer (2024) show a comprehensive interplay of episodic and semantic elements in natural scenes but highlighting the recall accuracy aspect of performance. In Chapter 4 and Chapter 5, I argue that despite learning the overall semantic regularity from the input, people’s representation of uncertainty remains attached to the individually recalled item, just like in scenes, but with even stronger semantic bias from input regularity. All this results in a peculiar mix of episodic and semantic elements in the recalled representation.

1.4 The Effect of Noise and Attention on Recall Precision and Uncertainty Representation

To gain a comprehensive view about the dynamics and the form of representation of long-term episodic and semantic memory, the effect of added noise and varying levels of attention at encoding should not be neglected. It is known that attention and memory (let it be working or long-term memory) is just as much, if not more, intertwined than episodic and semantic memory. For instance, attention can prioritize a certain segment of a long-term memory trace depending on task demands (Cowan et al., 2024). On the other hand, long-term semantic knowledge structures can direct attention when new stimuli are encoded (Cowan et al., 2024). Along with this idea, it has been shown that attention modulates later memory familiarity strength (Ramey et al., 2020). Thus, attention not only selects but also boosts the encoding of certain items. It has been shown that lack of attention affects both episodic and semantic recall under divided attention (Greene & Naveh-Benjamin, 2022). These results suggest that when attention is limited, the quality of the internal model also becomes impaired in certain conditions. In addition, investigating the forgetting patterns and subjective confidence under full and divided attention in young and older adults, Greene & Naveh-Benjamin. (2022) showed that young adults are more susceptible to an overall impairment in recall with divided attention, while older adults show specific episodic impairments. In terms of confidence, this study confirmed earlier results that subjective confidence predicts memory performance less accurately with age, as older adults are more likely to make high-confidence memory errors.

What is missing from these earlier studies is investigating attention, recall precision (with overarching regularity in the input) and the representation of uncertainty together in a common framework. In Chapter 5, besides adding overarching semantic regularity to the input as before, I also selectively add more noise, divided attention, and more attention to object detail in the encoding phase. I argue that with noise and divided attention semantic learning remains intact compared to performance in normal attentional states, contrary to episodic learning which decreases. I also find that modulating noise and attention at encoding does not change the fundamental strategy and form of representing uncertainty.

1.5 Summary of Key Points, Results

In this dissertation, I paint an image of internal representation in which episodic memory is intertwined with general semantic knowledge structures and is represented along with its subjective uncertainty, in a probabilistic manner. Throughout the dissertation I rely on analyzing the recall precision of episodic memories and the representation of their uncertainty. First, with pure episodic memories without any input regularity, I show how uncertainty is understood and measured upon recall with a fine-grained continuous response design, and argue for the imprecise, noisy, but probabilistic recall of episodic memories (Chapter 2). Next, I add different kinds and systematically more complex regularities to this pure episodic input by organizing objects into a scene with perceptual and semantic schemas (Chapter 3), and by adding overarching semantic structure to episodic stimuli (Chapter 4). By doing so, I demonstrate true semantic learning of the input, along with the previously shown item-based uncertainty representation of memory traces. Finally, in Chapter 5, I systematically modulate the level of input noise and attention at encoding to get an even more comprehensive picture of episodic and semantic memory dynamics and show that true semantic memory but not episodic details are robust to noise and attentional changes. In sum, this dissertation offers a comprehensive framework in which the dynamics and features of long-term episodic and semantic memory can be studied unitedly to shed more light on the general principles of knowledge representation.

CHAPTER 2

REPRESENTATION OF UNCERTAINTY IN PERCEPTUAL DECISION-MAKING AND IN LONG-TERM PURE EPISODIC MEMORY

2.1 Recall Performance in Long-term Episodic Memories

What is encoded and recalled in long-term visual memory and why? This is a question which has been the main focus of several long-term memory studies in the past (Standing, 1973; Brady et al., 2008; Konkle et al., 2010). Some of the mentioned studies reported that people are able to store a large number (several hundreds or even thousands) of images and objects in long-term memory and recall them with high fidelity. However, it is relatively unknown whether this so-called massive capacity of memory remains stable across input and test style manipulations. In the reports mentioned above, the encoded objects are oftentimes categorically distinct, and participants recall them in an alternative forced choice design, plus the test of the memory is done on the object level (object identity recognition). Given these encoding and recall criteria, memory capacity may be more massive but in real world environments the encoding and recall circumstances are oftentimes much more delicate where a massive episodic storage for object and scene detail is not even necessarily needed. As there has been some previous evidence to show that humans are effective at capturing and learning regularities of the environment (Hemmer & Steyvers, 2009; Fiser et al., 2010; Persaud & Hemmer, 2024) more so than remembering specific details of it with high precision.

In my dissertation, I systematically add more regularity to episodic input and observe how it affects recall and the relationship between episodic and semantic memory. In Chapter 2, I start with one of the simplest inputs that can be encoded in long-term memory, sequentially presented and semantically unrelated individual objects which I call "pure episodic input/memories". In terms of recall performance, the main concept I discuss is the aforementioned massivity. I argue that it is a vague concept and recall efficiency and vividity of episodic memories depend a lot on the encoding and testing contexts. I show that when the input lacks possible episodic or semantic associations, context or any kind of overarching regularity, then episodic memory for object detail is rather poor with even low set sizes and a two alternative forced choice testing method.

Apart from the capacity of long-term visual memory, Chapter 2 also discusses the nature of representations in terms of boundedness. Utochkin and Brady (2020) showed that parts of natural objects can be represented independently, as individual parts in long-term memory. It is possible that memory for a particular feature is lost while others are preserved. In Chapter 2, utilizing separate confidence measurements for object identity and object orientation, I show that while it may be possible that people recall the identity of an object and not its orientation or vice versa, but in those cases, subjective confidence in all parts of the representation drops compared to when both parts of the objects are recalled successfully, which suggests a holistic representation for objects, at least on the level of uncertainty representation.

2.2 Subjective (Un)certainty in Working Memory and In Long-term Episodic Memory

Another major topic that Chapter 2 investigates is the probabilistic nature of representations in perceptual decision-making and in episodic long-term memory. Several studies suggested that humans use their representation of uncertainty for optimal decisions (Körding & Wolpert, 2004; Fiser et al., 2010; Ma & Jazayeri, 2014; Li et al., 2021). However, just like in the case of recall performance, there are important subtleties to discuss regarding uncertainty representations. First, there may be several ways to handle uncertainty in probabilistic computations. A probability distribution can simply be represented by its mean and variance. Alternatively, it is possible that at any one time in a perceptual or recall process, a full probability distribution is represented about a given variable. Discussing the various ways in which probabilistic computations can be done is not the main focus of this dissertation, however, with utilizing the so-called calibratedness measure, I implicitly make the claim that when variables are recalled probabilistically, a whole distribution of values are represented with their corresponding probability values.

Defining calibratedness

Calibratedness is the measure of the level of linear correlation between objective task performance and subjective (un)certainty reports (task performance being object identity and object orientation recall accuracy in this dissertation). Well-calibratedness means that the level of correlation between objective performance and subjective certainty is significant. To my knowledge, the first instance in the literature where the correlation between error and uncertainty was directly linked to some form of probabilistic representation was in Lengyel at el. (2015). In their preprint, they used time related patterns in the correlation between error and uncertainty to pinpoint probabilistic sampling in the brain. The term calibratedness was first used in Magas et al. (2022) and in Magas & Fiser (2023) where it meant exactly what has been described before, the level of linear correlation between accuracy and subjective uncertainty. In more detail, the correlation is measured for each subject throughout the trials. And in each experiment 'well-calibratedness' is established if the mean correlation of the subjects is significantly above chance (0).

The way calibratedness or well-calibratedness leads to probabilistic encoding/representation is simple. To describe the relation, it is best to walk through *figure 1.2* below. Let us say that one encodes an object and later the task is to recall its orientation or color. This dissertation imagines probabilistic recall the following way. When the object is recalled, several orientation or color values are represented with corresponding probability values. The reported orientation value is the expected value of the probability distribution. In addition, there are two reasonable assumptions that are made. One is that the reported value is not necessarily centered on the true stimulus value. Second, the wider the probability distribution, on average, the farther the reported value is going to be from the true stimulus value. The more uncertain a participant is in a given trial, the more values they are going to represent for orientation or color. With more uncertainty, the represented probability distribution is going to be wider and flatter. And with that, the reported value is going to be farther from the true stimulus value, i.e., error will be larger with more uncertainty. This way, there is a straightforward connection between the width of the probability distribution and accuracy. Consequently, if subjects use the aforementioned probabilistic representation, then, by definition, they should be well-calibrated. It might not be the only type of probabilistic representation that leads to well-calibratedness. However, it is certain from the view above that if there is well-calibratedness, there is also intelligible way of representing and handling uncertainty.

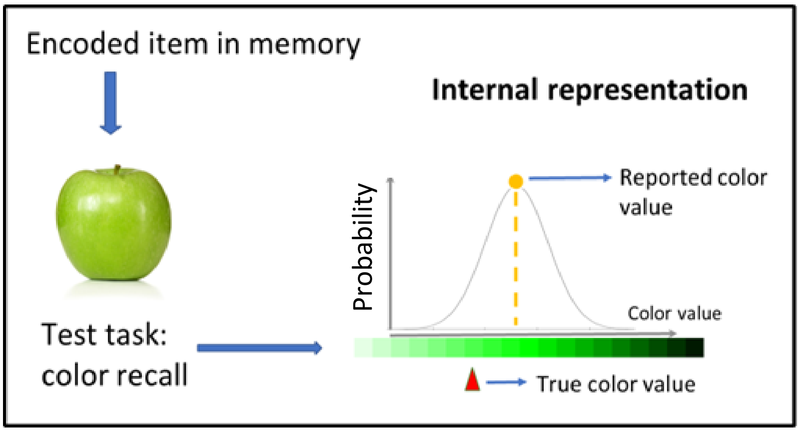

Figure 1.2 Probabilistic representation of episodic memory: upon recall, humans represent potential values and the corresponding probabilities about an object category or object feature (e.g.: size, color, orientation).

In long-term memory, significantly fewer studies have investigated the representation of uncertainty than in perception (Rahnev et al., 2020). What is more, confidence (or certainty) measurements in long-term memory tended to use continuous or more fine-grained scales less frequently than studies in perception. Therefore, the experiments presented in Chapter 2 are valuable for several reasons. First, it gives a comprehensive picture about the representation of uncertainty in pure episodic memories with a fine grained, continuous, sophisticated confidence/certainty measure, which arguably gives a more veridical picture of people's uncertainty representation. Second, it also bridges the gap between working and long-term memory regarding the representation of uncertainty which, to my knowledge, has not been explicitly done before. Nevertheless, for example, Persaud & Hemmer (2016) studied time related dynamics of episodic memory but rather from an accuracy standpoint.

In a series of five experiments, I demonstrate that, most importantly, uncertainty is represented in pure episodic memories, in a well-calibrated manner, at all levels of certainty. This suggests probabilistic encoding and recall for the details of an episodic item. I also show that the level of calibratedness, therefore the probabilistic form of representation, is robust to the number of items that are memorized, as long as there is above chance recall performance. And crucially, the probabilistic manner of representation is robust to the method with which memories are tested, whether it is a 2AFC or free recall design. These results suggest that even if a particular memory trace is not available without any recall cue, once retrieved, it is done probabilistically.

2.3.1 Experiment 1: Probabilistic Representations in Perceptual Decision-Making

Experiment 1 investigates whether representations in working memory, in a perceptual decision-making task are probabilistic. Although perception and working memory are not the main theme of this dissertation, presenting Experiment 1 is useful for two reasons. First, I demonstrate one key element in its design that is used throughout the long-term memory experiments, too. Second, results from Experiment 1 serve as the basis for comparison, regarding the probabilistic nature of recall. The main indicator of probabilistic encoding and representations here is the aforementioned well-calibratedness of (orientation) responses, a significant linear correlation between objective accuracy and subjective certainty.

Participants

Fifteen people participated in Experiment 1 (PERCEPTUAL DECISION-MAKING). All of them were recruited through the Hungarian MADS student organization and were mostly university students from Hungary. They received hourly compensation of 3000 HUF (~7.5$) for their participation. The experiment lasted for about one hour. The experiment took place in person, in the visionlab of CEU, in Budapest. All participants had normal or corrected to normal vision. The study was approved by the Hungarian United Ethical Review Committee for Research in Psychology (EPKEB), and all participants provided informed consent.

Materials and Procedure

Experiment 1 consisted of 2 practice and 3 test sessions. Throughout the experiment participants worked on a Wacom tablet of 20 by 16 cm, using a pen. The study was done offline, in the lab, therefore all participants viewed the images on the same sized monitor screen (~17 inches). In the first practice session, participants drew wedges on the tablet with a single line. In each trial, a single, orange wedge appeared in the middle of the screen with a particular width. Participants saw 34 wedges altogether, and the widths were sampled randomly from [20, 180] degrees. The size of the wedge depended on its width, the smallest one was around 30 by 30 pixels, and the largest wedge was around 100 by 100 pixels. Each wedge width had a corresponding line length which participants had to draw on the tablet. The drawn line opened another wedge automatically with which they had to match the stimulus wedge with sufficient precision. In each trial, until the error was sufficiently small, they had to repeat their responses.

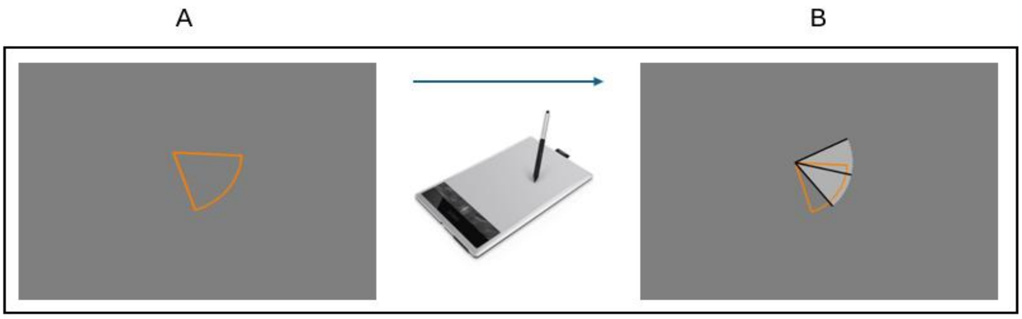

Figure 2.1. First practice session: wedge drawing. The first practice session helps participants learn the connection between line length and wedge width**. A:** Stimulus wedge. **B:** Stimulus wedge + wedge drawn by the participant (by a single line stroke on a tablet).

The second practice session matched the real task in Experiment 1 almost exactly. Here, after a fixation dot that appeared in the middle of the screen, participants saw a varying number of oriented Gabor patches, presented simultaneously on a circular organization, centered in the middle of the screen with a diameter of ~210 pixels. The orientation of the Gabor patches was sampled randomly from [0, 180] degrees. For reference, in psychtoolbox (which was used for all the experiments in the dissertation) a 0 degree orientation (for objects as well) is a vertical orientation. The number of Gabor patches varied randomly between 3 and 6. Contrast levels were sampled randomly for each trial from [0 0.1353 0.2231 0.3679 0.6065 1], scale [0, 1]. Presentation time also varied from trial to trial sampled randomly from [33 133 600] milliseconds. The size of each Gabor patch was 64 pixels in length. After the stimulus, a mask appeared on the screen with an orange circle, indicating the location of one of the Gabor patches. At this time, participants had to match the orientation of the circled Gabor patch with a line on the tablet and give their subjective certainty in their response. They did the subjective certainty rating by drawing the response orientation line to a certain length. The shorter the line the larger the certainty, the longer the line the smaller their certainty was in their orientation response.

With a certain line length, they automatically opened a wedge, exactly like in the first practice session. There was no specified minimum line length, however, after a certain length (around 150 pixels) the wedge was fully opened to 180 degrees indicating maximum uncertainty. This is a fast and instantaneous measure of subjective uncertainty that arguably reflects participants' representation more veridically than discrete language-based scales that involve higher-level, conscious processing. In each trial, participants received feedback. The true orientation of the Gabor patch appeared on the screen along with participants' response orientation and subjective certainty, shown by wedge width. The function describing the connection between line length and wedge width was identical to the first practice session. Participants' error was the angular difference between the true and response orientation. Participants completed 72 trials in this practice session.

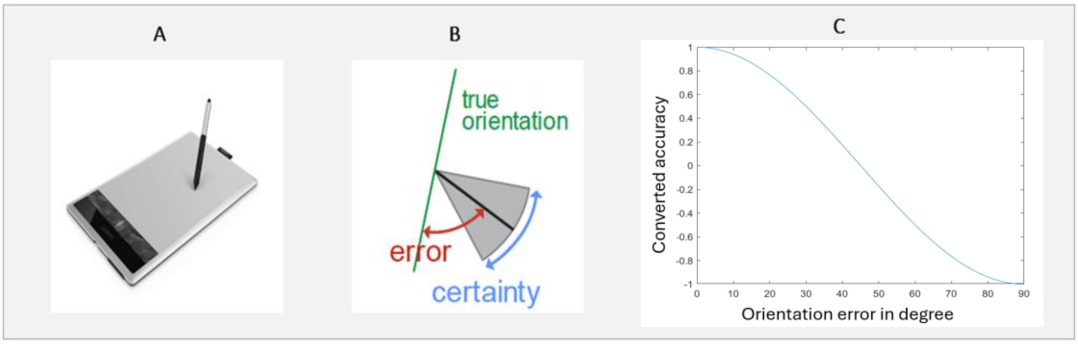

Figure 2.2 Second practice session: orientation response. The general orientation response method used throughout the dissertation. A: orientation response: single line stroke on a tablet with a pen. **B:** black line: reported orientation. Line length: subjective (un)certainty in the orientation response. **Measures:** accuracy: trial average of the (cosine) error, scale: [-1, 1]. Certainty: subjective estimate of the expected accuracy derived from wedge width, scale: [0, 1]. **C:** cosine function converting the orientation error from: [0, 90] in degrees, to a scale: [-1, 1], where -1 is the largest and 1 is the smallest possible error.

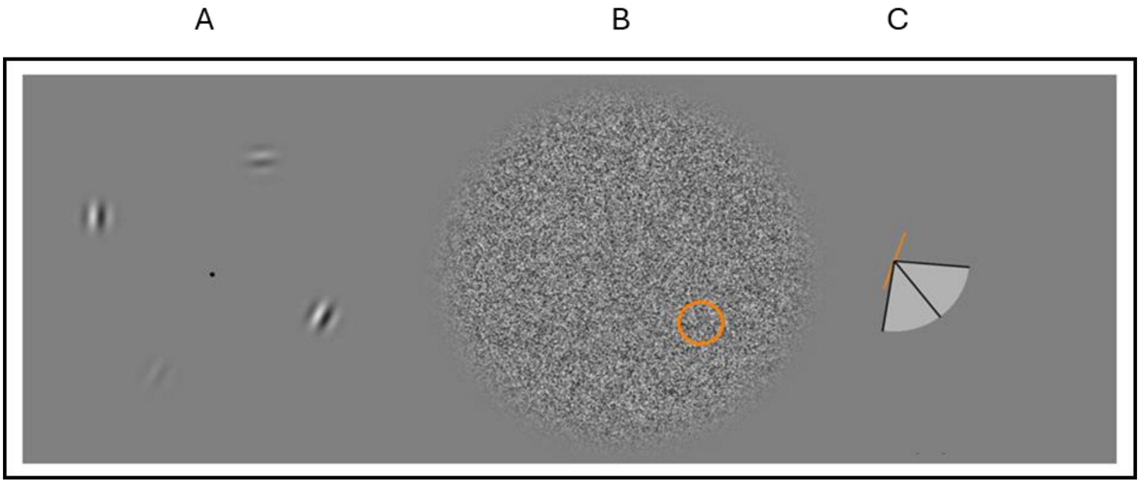

Figure 2.3 Sequence of orientation response practice with Gabor-patches. A: stimulus: oriented Gabor patches (four in the figure only for illustration purposes) with varying contrast levels and presentation time. **B:** mask with orange circle locating the target patch. **C:** feedback about the true orientation (orange line) and the drawn orientation (middle line within the wedge) and subjective certainty reported in a given trial (wedge width).

After practice, three test sessions followed. The procedure was identical to the second practice phase, with the exception that participants did not receive feedback about the true orientation of the target Gabor patch. The orientation of the Gabor patches was sampled randomly from [0, 180] degrees. The set size varied between 3 and 6. Contrast levels were sampled randomly from [0 0.0498 0.0821 0.1353 0.2231 0.3679 0.6065 1]. Presentation times were selected randomly from [33 50 83 133 200 600] milliseconds. Each test phase consisted of 192 trials. Between test phases, participants could optionally take a few minute break.

Throughout the dissertation, there is one main exclusion criteria from the data analysis: below chance level (0 on the converted scale) performance for object orientation accuracy. These cases indicate that subjects had zero or close to zero overall memory of the input.

Results

In Experiment 1, the correlation between orientation accuracy and subjective certainty, across subjects, was significantly above chance (mean_pearson_r = 0.195, SD = 0.190; t(14) = 3.976, p = 0.001, Cohen's d = 1.027, BF = 29.404).

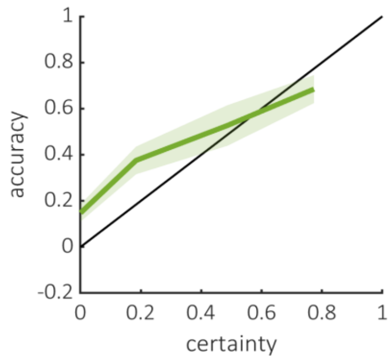

Figure 2.4 Well-calibratedness in Experiment 1. There was a significant correlation between accuracy and certainty indicating that subjective certainty reports were well-calibrated. They predicted participants' accuracy, at all levels of certainty, suggesting probabilistic representations.

Discussion

In Experiment 1, orientation responses were well-calibrated, indicating a probabilistic representation of the stimulus in working memory. Participants could use their (un)certainty to make meaningful predictions for their accuracy, at all levels of certainty. It means that in the representation, instead of a single point estimate, a whole distribution of potential orientation values is represented with the corresponding probabilities. I argue that this result is strong evidence for the probabilistic nature of representations in perceptual decision-making. Further, it serves as the basis for comparison with long-term visual memory which is discussed in upcoming experiments.

2.3.2 Experiment 2/a: Probabilistic Representations in Pure Episodic Memories

Participants

Twenty-six[1] people participated in Experiment 2/a (60 SINGLE RANDOM). All of them were recruited through the Hungarian MADS student organization and were mostly university students from Hungary. They received hourly compensation of 3000 HUF (~7.5$) for their participation. The experiment lasted for about one hour. The experiment took place in person, in the visionlab of CEU in Budapest. All participants had normal or corrected to normal vision. The study was approved by the Hungarian United Ethical Review Committee for Research in Psychology (EPKEB), and all participants provided informed consent.

Materials

The stimuli for the long-term memory experiments consisted of a set of 190 long and narrow natural objects selected from a larger pool of around 800 images/objects[2]. The selection of the stimuli was done mostly by google search or by making images of real-world objects that fit the long and narrow quality. The presented objects were mostly categorically distinct and with few semantic relations between them to avoid associations and structure learning. Naturally, it is not possible to exclude all connections and meaning from the input, when they are real world objects. The monitors used were identical to those of Experiment 1. The images/objects appeared in the middle of the screen in 200x200 pixels

[1] Based on pilot experiments, meaningful patters started to emerge in the data with around 10-15 participants, so the target number was at least 10, or when the data started to crystalize. The number of participants varied based on whether a given experiment was an important base (e.g., Exp 2/a, Exp 4/a) or 'only' a supporting experiment. Also, in some cases patterns started to emerge in the data later. Importantly, all experiments were done offline, consequently, the number of volunteers from the MADS organization was limited.
[2] The stimulus set for the entire thesis is available on this link.

embedded in a grey circular mask with a diameter of 300 pixels. Outside the image and the mask, other parts of the screen were white and empty.

Procedure

Experiment 2/a consisted of 3 practice sessions, and 1 test session. The first 2 practice sessions were almost identical to Experiment 1. First, participants learned the mapping between wedge width and line length in 34 trials, then in the second session, they practiced the orientation response and subjective certainty feedback with Gabor patches (see figures 2.2 and 2.3). Here participants only completed 50 trials instead of 72. Overall, these two seemingly unrelated practice sessions were necessary for my long-term memory experiments, too, because the orientation recall of real-world objects used the same method as with the Gabor patches. In addition, the orientation feedback could not be practiced with real objects because they could cause potential interference in encoding and recall.

The long-term memory part of the experiment consisted of a practice and test session, each containing a study and a recall phase. In the study phase, participants passively watched 60[3] randomly selected objects (5 during practice) in various orientations. The objects were presented individually for 1 second with a 2 second interstimulus interval. The instruction was to remember the presented objects and their orientation. The orientation of the objects was sampled from a uniform distribution from [-90, 90] degrees. This ensured that objects were not shown in an upside-down position, yet the entirety of possible orientations on a circular pattern was covered.

[3] Throughout the dissertation 60 objects seemed to be a sweet spot, where recall performance is meaningful and above chance. In addition, there is time for structure learning and for prior knowledge to accumulate.

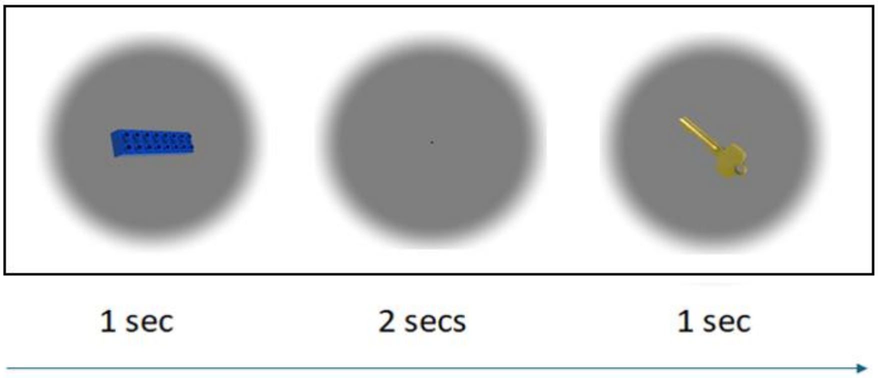

Figure 2.5 Study phase of Experiment 2/a. The general study phase structure of long-term memory experiments in Chapter 2. Objects are presented sequentially and individually with random orientation. Participants are instructed to remember the object and their orientation.

After all the objects were shown, participants had the option to take a few minute break before the recall phase or continue right away (this was an option only during the test session). In the recall phase, participants had to answer two questions about 120 objects (10 during practice). Thus, the recall phase consisted of 50% old (previously presented) and 50% new (previously not presented) objects. The testing order of the objects was randomized in all long-term memory experiments in the dissertation. These objects were shown for 1 second, always in a vertical position to ensure that a recall cue did not remain present on the screen, and that there is as little interaction as possible with the orientation in which the object appeared in the study phase. Participants had to answer two questions about each object. First, whether they had seen the given object in the study phase or not and how confident they were in their responses. This first answer concerning object identity was given on a horizontal scale. If participants dragged a pointer to the left from the middle it meant that they did not see the object before, and if they dragged the pointer to the right, it meant that they had seen it before. In addition, the distance (how far they dragged the pointer from the middle) corresponded to their subjective confidence in their answer.

Confidence was on a scale [0, 1] in this measure, and it was the linear function of the distance from the middle either to the left or to the right. The second question concerned the orientation of the objects with which they were presented during the study phase. Participants had to recall the orientation in a free recall setup and draw a line on the tablet that matched the orientation of that particular object in the study phase. In addition, they once again had to indicate their subjective certainty in the orientation response. They did this identically to Experiment 1, by drawing the response orientation line to a length corresponding to their subjective certainty. The longer the line, the smaller their certainty was in their response, just like in Experiment 1. Both during the practice and test sessions, participants received feedback about their drawn orientation line and wedge width, which appeared on the screen after they responded.

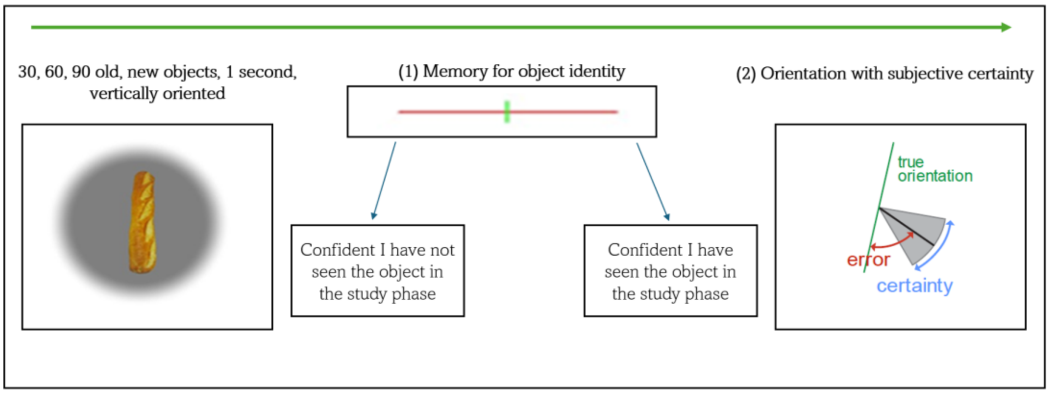

Figure 2.6 Recall phase in experiments 2/a, b, and c. The recall phase for individually presented objects that are tested in a free recall design. Participants respond about object identity and object detail (orientation). Both object identity and orientation responses are given along with subjective certainty reports, although the two responses use slightly different measures.

In the long-term memory experiments, there are three main scores along which participants' performance is analyzed. Performance for object identity is shown by d prime from classic signal detection theory, the strength of the signal calculated from participants hit and false alarm rates. Further, performance for object orientation. This is the cosine error between the true and drawn orientation. Lastly, the correlation between error and subjective (un)certainty, just like in Experiment 1.

Results[4]

Regarding performance for object identity, a one sample t-test yielded significantly above chance performance (chance = 0) considering participants signal strength (d') ($M = 2.281$, $SD = 0.773$; $t(25) = 15.039$, $p < 0.001$, $BF > 1000$). The corresponding subjective confidence levels were high across subjects ($M = 0.88$, $SD = 0.072$).

As for orientation accuracy, another one sample t-test showed significantly above chance performance (chance = 0) ($M = 0.29$, $SD = 0.168$; $t(25) = 8.806$, $p < 0.001$, $BF > 1000$). This result is for the objects that were presented during the study phase. As a sanity check, accuracy for objects that were not presented was not significantly different from chance level ($M = 0.011$, $SD = 0.059$; $t(25) = 0.938$, $p = 0.357$, $BF = 0.308$).

Importantly, a Pearson correlation between orientation accuracy and subjective certainty across subjects was significant (mean_pearson_r = 0.256, $SD = 0.160$; $t(25) = 8.173$, $p < 0.001$, Cohen's $d = 1.603$, $BF > 1000$). Comparing the subject-wise correlation with Experiment 1, with an independent sample t-test, there was no significant difference

[4] In all further experiments in the dissertation, the result figures are presented together at the end of each chapter.

between Experiment 1 and Experiment 2/a (M_exp1 = 0.195, SD = 0.190; M_exp2/a = 0.256, SD = 0.160; $t(39) = 1.090$, $p = 0.282$, Cohen's d = 0.353, BF = 0.502).

Discussion

Experiment 2/a showed that even though participants can recognize well whether they have seen a particular object before, they perform more purely when an object detail (orientation) must be recalled. On the one hand, this part of the results is not surprising as there has been previous evidence to show that people tend to lose the details of a memory trace faster than the overall idea about it (Lew et al., 2016; Greene & Naveh-Benjamin, 2022). However, the relatively low (nevertheless above chance) performance for object orientation is surprising given previous results claiming massive memory for even object details (Brady et al., 2008; Wolfe et al., 2023). Importantly, the significant correlation between orientation accuracy and subjective certainty indicates that long-term episodic memories (or at least their orientation) are encoded and recalled in a probabilistic manner, just like representations in working memory.

2.3.3 Experiment 2/b: Probabilistic Representations in Pure Episodic Memories with a Decreased Set Size

One of the main questions of my dissertation is how robust probabilistic representations in long-term memory to certain input changes are. The simplest, yet an important change to stimulus characteristics is an altered set size and therefore the potential manipulation of participants' memory strength for individual items. Does the structure of representations remain similar to that of Experiment 1 and Experiment 2/a with different set sizes?

Experiment 2/b (30 SINGLE RANDOM) and Experiment 2/c (90 SINGLE RANDOM) intend to investigate this question.

Participants

Thirteen people took part in Experiment 2/b. They were recruited through the Hungarian MADS student organization and were mostly university students from Hungary. They received hourly compensation of 3000 HUF (~7.5$) for their participation. The experiment was about one hour long. The study was conducted in person, in the visionlab of CEU in Budapest. Every participant had normal or corrected to normal vision. The study was approved by the Hungarian United Ethical Review Committee for Research in Psychology (EPKEB), and all participants provided informed consent.

Materials

The materials in Experiment 2/b were identical to that of Experiment 2/a.

Procedure

The procedure was also identical to Experiment 2/a, except now 30 objects were presented during the study phase of the test session of the long-term memory part of the experiment which were tested along with 30 new items in the recall phase.

Results

Object identity performance was significantly above chance ($M = 3.633$, $SD = 1.393$; $t(11) = 9.032$, Cohen's $d = 2.607$, $p < 0.001$, $BF > 1000$) in Experiment 2/b. The corresponding subjective confidence for object identity was high ($M = 0.88$, $SD = 0.081$). Further, object identity performance was significantly higher in Experiment 2/b (30 SINGLE RANDOM) than in Experiment 2/a (60 SINGLE RANDOM) (M_exp2/a = 2.281, SD = 0.773; M_exp2/b = 3.633, SD = 1.393; $t(36) = 3.858$, $p < 0.001$, Cohen's d = 1.346, BF = 58.154).

Accuracy for object orientation in Experiment 2/b was highly significantly above chance (M = 0.418, SD = 0.193; $t(11) = 7.508$, $p < 0.001$, BF > 1000). Further, there was a significant difference between Experiment 2/a (60 SINGLE RANDOM) and Experiment 2/b (30 SINGLE RANDOM) in orientation recall performance (M_exp2/a = 0.290, SD = 0.168; M_exp2/b = 0.418, SD = 0.193; $t(36) = 2.082$, $p = 0.044$, Cohen's d = 0.727, BF = 1.686).

The Pearson correlation between orientation accuracy and certainty was significant in Experiment 2/b (mean_pearson_r = 0.204, SD = 0.213; $t(11) = 3.320$, $p = 0.007$, Cohen's d = 0.958, BF = 8.075).

One participant was excluded from analyses from Experiment 2/b for below chance orientation accuracy.

Discussion

Decreased set size significantly increased both object identity and orientation performance. The increase is not surprising overall, however, the fact that set size had a larger effect on object identity performance than orientation performance is puzzling. With 60 instead of 30 objects, identity performance suffered a larger decrease. It may be that in my experiments, even in the case of 30 objects, orientation detail could not be encoded with high enough precision, therefore the smaller drop in performance from 30 to 60 in terms of orientation accuracy. If it is the case, then it is further proof against the massive nature of long-term visual memory, in the case of random objects. Most importantly, the calibratedness of orientation responses showed a very similar pattern with a lowered set size, indicating that the form of representation remains the same, i.e. probabilistic.

2.3.4 Experiment 2/c: Probabilistic Representations in Pure Episodic Memories with an Increased Set Size

Participants

Eighteen people took part in Experiment 2/c, recruited through the Hungarian MADS student organization, and as before, they were mostly university students from Hungary. Their hourly compensation was 3000 HUF (~7.5$) for their participation. The experiment was about one hour long. The study was conducted in person, in the visionlab of CEU in Budapest. Every participant had normal or corrected to normal vision. The study was approved by the Hungarian United Ethical Review Committee for Research in Psychology (EPKEB), and all participants provided informed consent.

Materials

The materials were the same as in experiments 2/a and 2/b.

Procedure

The procedure was also identical to that of experiments 2/a and 2/b, however, now, 90 objects were selected to be shown in the study phase of the test session of the long-term memory part of the experiment and an additional 90 new objects were tested with the old ones, in the recall phase.

Results

Object identity performance was significantly above chance in Experiment 2/c ($M = 2.480$, $SD = 1.163$; $t(17) = 9.048$, $p < 0.001$, $BF > 1000$). Subjective confidence levels for object identity were very high across participants ($M = 0.902$, $SD = 0.087$). However, there was no significant difference between Experiment 2/a (60 SINGLE RANDOM) and Experiment 2/c (90 SINGLE RANDOM) in identity performance. ($M_{exp2/a} = 2.281$, $SD = 0.773$; $M_{exp2/c} = 2.480$, $SD = 1.163$; $t(42) = 0.683$, $p = 0.498$, Cohen's $d = 0.210$, $BF = 0.364$).

Orientation recall accuracy was significantly above chance in Experiment 2/c ($M = 0.236$, $SD = 0.140$; $t(17) = 7.126$, Cohen's $d = 1.680$, $p < 0.001$, $BF > 1000$). Comparing orientation accuracy to Experiment 2/a (60 SINGLE RANDOM) there was no significant difference ($M_{exp2/a} = 0.290$, $SD = 0.168$; $M_{exp2/c} = 0.236$, $SD = 0.140$; $t(42) = 1.122$, $p = 0.268$, Cohen's $d = 0.344$, $BF = 0.498$).

Orientation responses in Experiment 2/c were well-calibrated indicated by the significant Pearson correlation between accuracy and certainty across subjects (mean_pearson_r = 0.211, $SD = 0.135$; $t(17) = 6.602$, $p < 0.001$, Cohen's $d = 1.556$, $BF > 1000$).

Discussion

An increased set size did not significantly change either object identity performance or orientation accuracy compared to the base experiment (60 SINGLE RANDOM). It indicates that after a stronger initial encoding with a smaller set size, seen in Experiment 2/b (30 SINGLE RANDOM), meaningful memory remains both for the object as a whole and for object detail, even with 90 objects presented. Most importantly, the well-calibratedness of orientation responses did not change either. It suggests that as long as there is above chance or meaningful memory performance then encoding and recall is done in a probabilistic manner.

2.3.5 Experiment 2/d: Probabilistic Pure Episodic Memories with an Alternative Forced Choice Testing Method

In all previously presented experiments, participants gave their orientation responses in a free recall setup. After showing a test object for 1 second they had to remember the orientation without any cue. However, as previously discussed, even when people cannot freely recall a memory trace, with a helping cue, otherwise inaccessible memories can be retrieved (Yonelinas, 2002). The question arises: does the structure of the recalled memory change with a forced choice instead of a free recall testing method. In other words, whether familiarity and remembering in memory follow the same principles regarding the probabilistic nature of the representations.

Participants

Twenty people took part in Experiment 2/d, recruited through the Hungarian MADS student organization. They were mostly university students from Hungary. Their hourly compensation was 3000 HUF (~7.5$) for their participation. The experiment was about one hour long. The study was conducted in person, in the visionlab of CEU in Budapest. Every participant had normal or corrected to normal vision. The study was approved by the Hungarian United Ethical Review Committee for Research in Psychology (EPKEB), and all participants provided informed consent.

Materials

The materials used were the same as in Experiment 2/a.

Procedure

The procedure in the study phase was similar to Experiment 2/a, b and c. 60 objects were presented individually, and their orientation was sampled from a uniform distribution from [0, 180] degrees. However, now in the recall phase, at the orientation answer, the same object appeared simultaneously in two different orientations and participants had to select the correct one in a 2 alternative forced choice manner. One orientation was always the correct one, and the other one was 70 degrees plus or minus from the real one. The rationale behind 70 degrees was to avoid potential confusion that can originate from a 90-degree difference, a perpendicular presentation. In that case, it is arguably easier to mix the two orientations. Here, the objects remained present on the screen until subjects gave their response. Importantly, participants choose the orientation by copying the orientation of the object which they believed was correct. They indicated their subjective certainty the same way as before, by drawing the response line to a certain length.

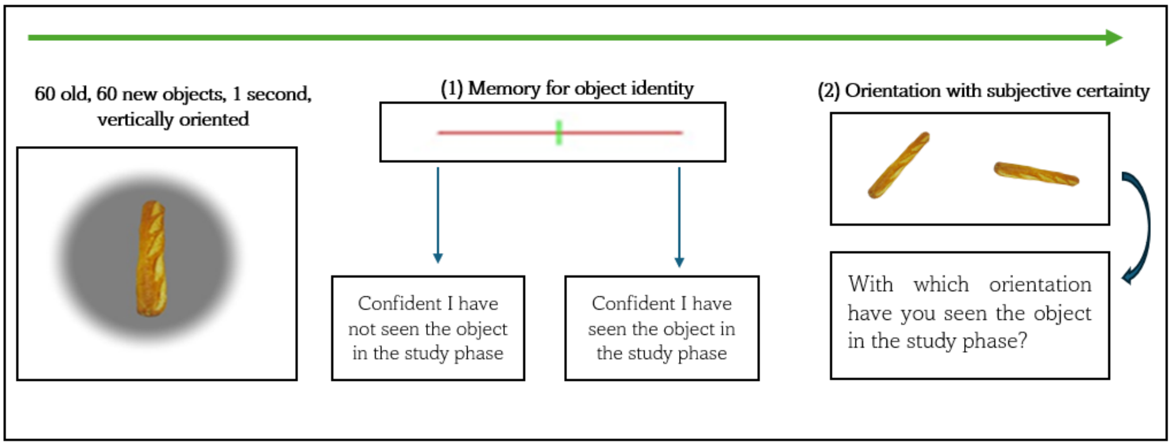

Figure 2.7 Recall phase of Experiment 2/d. The recall phase of Experiment 2/d differs only in that at the orientation response participants choose the correct orientation from two options by copying the orientation that is believed to be the true one. Importantly, this response is drawn on the tablet, the same way as before, with a corresponding subjective certainty report given by line length.

Results

Object identity performance was significantly above chance (M = 2.887, SD = 1.404; t(19) = 9.195, p < 0.001, Cohen's d = 2.056, BF > 1000). Mean confidence rating for object identity was very high (M = 0.903, SD = 0.082). The difference in object identity recognition between Experiment 2/a (60 SINGLE RANDOM) and 2/d (60 SINGLE RANDOM AFC) approached significance (M_exp2/a = 2.281, SD = 0.773; M_exp2/d = 2.887, SD = 1.404; t(44) = 0.683, p = 0.069 Cohen's d = 0.555, BF = 1.175).

Chance level for orientation recall in experiment 2/d was not zero but 0.12 because the alternative orientation was only 70 degrees away from the correct one, so the analysis needed to correct for this feature in the design. It was simply done by subtracting 0.12 from the mean accuracy of each subject. Taking this into account, orientation performance was not significantly different between Experiment 2/a (60 SINGLE RANDOM) and experiment 2/d (60 SINGLE RANDOM AFC) (M_exp2/a = 0.290, SD = 0.168; M_exp2/d = 0.296, SD = 0.253; $t(44) = 0.095$, $p = 0.925$, Cohen's d = 0.028, BF = 0.295). However, orientation accuracy remained significantly above chance in experiment 2/d (M = 0.296, SD = 0.253; $t(19) = 5.237$, $p < 0.001$, Cohen's d = 1.171, BF > 100).

As for calibratedness, a Person correlation was significant between accuracy and certainty in Experiment 2/d, indicating well-calibratedness (mean_pearson_r = 0.163, SD = 0.197; $t(19) = 3.703$, $p = 0.002$, Cohen's d = 0.828, BF = 25.5).

Further analysis for bound, unbound object representations, Experiment 2/d

The following analysis aims to investigate whether a long-term memory representation about an object is bound/holistic, or alternatively it is possible that individual memory parts have their separate and vivid representations, independent of each other. In my experiments, there are two measurable components of object representation: identity and orientation. The question is whether these two can become clearly separate in long-term memory. For this analysis, Experiment 2/d served the best base, as in that experiment both the object identity and the object orientation responses were given in an old/new and in a 2AFC design respectively, consequently, both kinds of responses can be classified into 'hit' or 'miss'. Responses were separated into three main categories. Hit-hit, when both the object identity and the object orientation responses were correct. Hit-miss was when the identity response was correct, but the orientation response was wrong. Miss-hit was when the identity response was wrong, but the orientation response was correct. There is also a fourth kind of response, miss-miss, however, in this analysis, those responses are rather pointless to consider since they imply no meaningful memory trace either for object identity or for object orientation. Of all the trials, hit-hit responses constituted about 65.8%, hit-miss responses 21.08%, miss-hit 6.8%, and miss-miss 6.25%. To be able to decide whether objects are represented in a bound or unbound manner, the subjective confidence ratings for object identity, certainty reports for object orientation and the correlation of these two variables were analyzed for the three cases: hit-hit, hit-miss, and miss-hit.

Considering object identity confidence, there was a significant difference between the hit-hit and hit-miss responses (M_hit_hit = 0.957, SD = 0.038; M_hit_miss = 0.881, SD = 0.127; t(38) = 2.563, p = 0.014, Cohen's d = 0.810, BF = 3.74). The difference between hit-hit and miss-hit was highly significant regarding identity confidence (M_hit_hit = 0.957, SD = 0.038; M_miss_hit = 0.804, SD = 0.158; t(37) = 4.212, p < 0.001, Cohen's d = 1.349, BF = 148.119). The difference in identity confidence between hit-miss and miss-hit was not significant, but the effect size was in the mid-range, showing a drop in miss-hit compared to hit-miss (M_hit_miss = 0.881, SD = 0.127; M_miss_hit = 0.804, SD = 0.158; t(37) = 1.689, p = 0.100, Cohen's d = 0.541, BF = 0.947).

Regarding orientation certainty, the difference between hit-hit and hit-miss responses approached significance (M_hit_hit = 0.579, SD = 0.154; M_hit_miss = 0.473, SD = 0.193; t(38) = 1.908, p = 0.064, Cohen's d = 0.603, BF = 1.270). The difference between the hit-hit and miss-hit responses in orientation certainty was highly significant (M_hit_hit = 0.579, SD = 0.154; M_miss_hit = 0.108, SD = 0.118; t(37) = 10.672, p < 0.001, Cohen's d = 3.419, BF > 100 000). Likewise, the difference between hit-miss and miss-hit was highly significant in orientation certainty (M_hit_miss = 0.473, SD = 0.193; M_miss_hit = 0.108, SD = 0.118; t(37) = 7.096, p < 0.001, Cohen's d = 2.273, BF > 100 000).

Finally, the correlation between identity confidence and orientation certainty across subjects was significant in the case of hit-hit responses (mean_pearson_r = 0.163, SD =

0.254; $t(19) = 2.867$, $p = 0.010$, Cohen's $d = 0.641$, $BF = 5.180$). Further, considering hit-miss responses, the correlation between identity confidence and orientation certainty was not significant (mean_pearson_r = 0.106, SD = 0.468; $t(19) = 1.013$, $p = 0.324$, Cohen's $d = 0.226$, $BF = 0.365$). Lastly, in the case of the miss-hit responses, the correlation was not significant either, and possibly even meaningless (mean_pearson_r = -0.252, SD = 0.581; $t(15) = 1.739$, $p = 0.103$, Cohen's $d = 0.435$, $BF = 0.874$).

Discussion

A two-alternative forced choice testing method had very little effect on the measures. Orientation recall accuracy was not significantly improved. Similarly, object identity performance was only mildly improved by the 2AFC method. This is surprising because in the 2AFC recall, a cue (the right orientation) was present during the response. On the other hand, it shows that for several objects, not even dim traces remained in memory which could have been accessed via a recall cue. Further, the significant correlation between accuracy and certainty shows that once memories become accessible either through free recall or with 2AFC, they are recalled in a probabilistic manner. This suggests that the probabilistic form of representation seems to be robust to testing style. Further, Experiment 2/d showed that parts of object representation (identity and orientation) are interconnected through the representation of confidence/uncertainty, indicating at least partly holistic representations for objects in long-term visual memory.

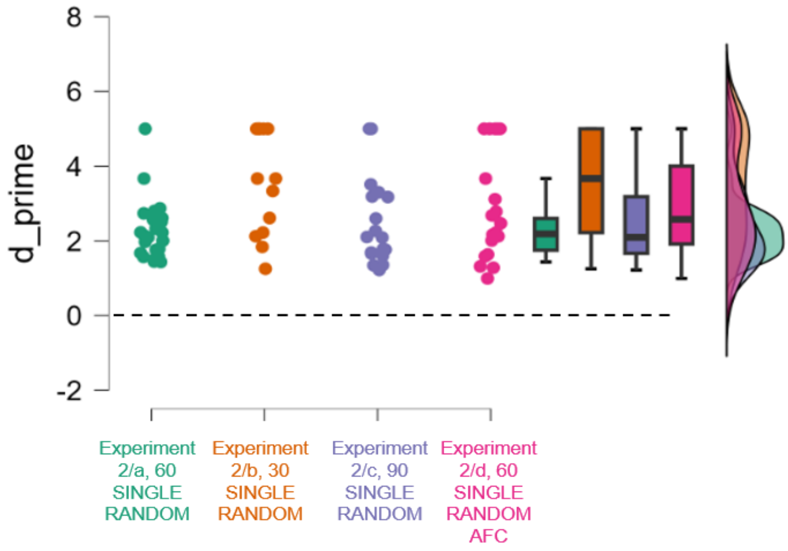

Figure 2.8 Object identity performance in Experiment 2/a, b, c, d. Performance for object identity was high throughout the long-term memory experiments. The dashed line indicates chance level performance.

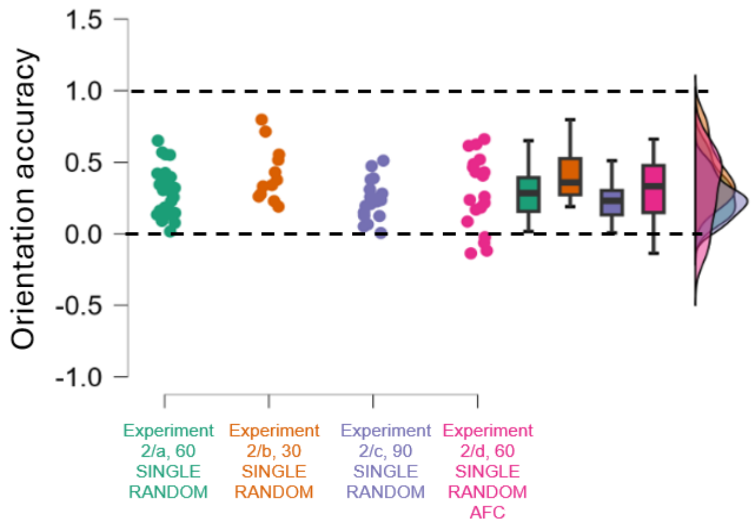

Figure 2.9 Object orientation recall performance in Experiment 2/a, b, c, d. Object orientation recall performance was above chance level (accuracy of 0) in all experiments. However, performance was not close to massive (accuracy of 1) in any of the experiments.

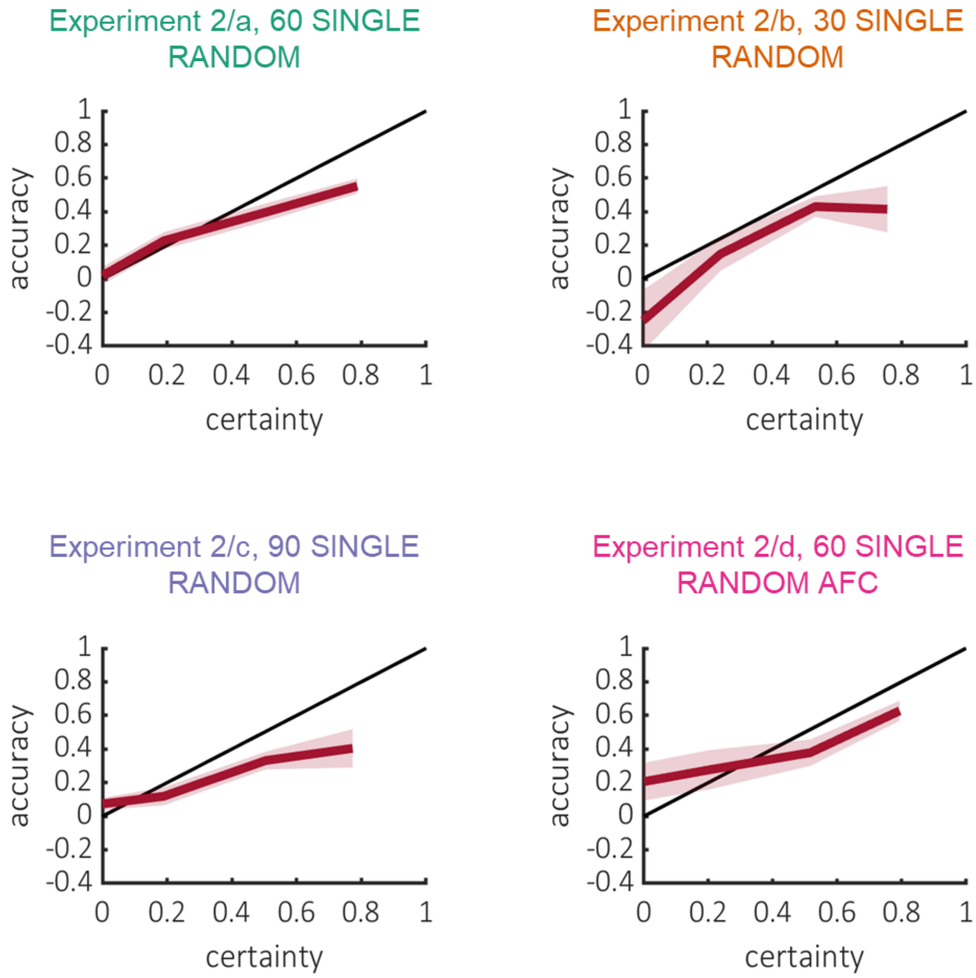

Figure 2.10 Calibratedness in Experiment 2/a, b, c, d. Subjective certainty reports were well-calibrated to accuracy throughout the four experiments, indicating probabilistic representations and recall.

2.4 Summary

In summary, Chapter 2, in a series of five experiments, investigated the massive nature of long-term visual memory, probabilistic recall in perceptual decision-making and long-term simple episodic memory, and boundedness in long-term object memory representation. Experiments in Chapter 2 showed meaningful but non-massive performance for object detail. Further, Experiment 2/a showed that long-term episodic memory is encoded and recalled probabilistically, and that regarding the probabilistic nature of representations, working memory and long-term episodic memory follow the same principles (Experiment 1, and 2/a). Even though I am not aware of indications in the literature that working and long-term memory operate on distinct principles, these results are surprising because in long-term memory there are potentially more things that can bias later recall such as mixing with other memories, or loss of memory vividity. In addition, experiments 2/b, 2/c, and 2/d showed that probabilistic encoding and recall are robust to changes in set size and testing method. Lastly, results from Experiment 2/d demonstrated that object parts (identity, orientation) in long-term memory, are bounded, on the level of uncertainty representation.

CHAPTER 3

RECALL EFFICIENCY AND REPRESENTATION OF UNCERTAINTY IN EPISODIC MEMORIES WITH WITHIN SCENE PERCEPTUAL AND SEMANTIC REGULARITIES

3.1 Recall Performance and Uncertainty Representations in Elementary Scenes

Chapter 2 discussed what is remembered, why, and in what form considering simpler forms of episodic memories, namely individual objects. Chapter 3 goes a step further in complexity and asks similar questions about elementary scenes. Regarding scenes, one of the main discussion points in the literature revolves around the concept of the gist which is understood to be the essence of a scene, mostly containing real world elements with a meaningful connection between them. This essence can come from spatial relations between scene elements, semantic associations (Persaud & Hemmer, 2024), or even statistical patterns in simple shapes (Brady & Alvarez, 2011). What is usually studied in connection with scenes, and the gist of scenes is the order by which certain elements and levels of a scene are perceived. In other words, the relative importance of scene elements (in certain contexts), and consequently whether these elements are in a hierarchy (Fei et al., 2007; Oliva & Torralba, 2008; Wu et al., 2014).

Chapter 3 also indirectly touches upon the sequence and the hierarchy in which elements are processed but the main question is rather whether and how certain regularities within a scene change the form and precision with which elements are encoded and recalled. In four experiments, I demonstrate that predominantly semantic but not perceptual connections facilitate recall precision. I also show that more complex within scene regularity not only improves overall recall accuracy but facilitates the formation of chunks, in long-term memory, that follow the structure of the input. Further, I show that regardless of within scene perceptual or semantic regularity, the fundamental probabilistic form of encoding, established in Experiment 1 and Experiment 2/a remains. In all four experiments, 3 individual objects are presented in one scene with various kinds and amounts of connections between them. Experiment 3/a serves as the bridge between Chapter 2 and Chapter 3, as in it the objects are random in the scenes, meaning there is no connection between them either perceptually or semantically. Through Experiment 3/a I demonstrate that with multiple objects in one trial, calibratedness is just as high as it was in Experiment 1, in perceptual decision-making. The remaining three experiments systematically show that semantic connections are dominant compared to perceptual ones in terms of contributing to higher recall performance by adding perceptual (Experiment 3/b), semantic (Experiment 3/c), and perceptual and semantic (Experiment 3/d) relations to objects in the scenes. Further, the four experiments together demonstrate that despite increasingly complex within scene regularity, the representation of uncertainty for individual objects remains well-calibrated throughout (although slightly more biased with more regularity). All these results show that the representation of a simple scene can be understood as semantic where elements are connected based on input structure, on the other hand, regarding uncertainty, the representation remains anchored to the individual objects, suggesting a mixture of episodic and semantic strategies and elements in encoding and recall.

3.2.1 Experiment 3/a: Recall Performance and Representation of Uncertainty in Random Scenes

In Chapter 2, results showed a relatively poor recall performance for object detail. When random, unrelated objects were presented individually and sequentially, participants had almost zero opportunity to make meaningful associations between the objects and therefore achieve better memory performance. In Experiment 3/a, still random (regarding perceptual and semantic connections) but instead of individual objects, elementary scenes (multiple objects) are presented in each trial. It has been shown that in working memory, an increased set size affects recall performance negatively (Lengyel et al., 2015). However, in long-term memory, an opposite effect can emerge, as it is presumably easier to make associations between objects when they are organized into a simple scene, even though the objects are unrelated both perceptually and semantically. As for the representation of uncertainty, an interesting question arises: does a well-calibrated object-based uncertainty representation remain, just like with individual presentations, or showing multiple objects in scenes result in an overall/scene-based representation of uncertainty.

Participants

Twenty-seven people participated in Experiment 3/a (60 MULTIPLE RANDOM). All of them were recruited through the Hungarian MADS student organization and were mostly Hungarian university students. They received hourly compensation of 3000 HUF (~7.5$) for their participation. The experiment lasted for about one hour. The experiment took place in person, in the visionlab of CEU in Budapest. All participants had normal or corrected to normal vision. The study was approved by the Hungarian United Ethical Review

Committee for Research in Psychology (EPKEB), and all participants provided informed consent.

Materials and Procedure

Experiment 3/a used the same set of 190 objects that was also used in Experiment 2/a, b, c, and d. Experiment 3/a consisted of 3 practice sessions and of 1 test session, similar to previous experiments. The first two practice sessions were identical to Experiment 2/a, b, c, d. Participants learned the function between drawn line length and wedge width on the tablet. Then, they learned and practiced the orientation response with subjective certainty reports in the second practice session with Gabor-patches. The long-term memory section of the experiment contained a practice and test session, each comprising a study and recall phase. In the study phase, in each trial, 3 objects were shown simultaneously for 6 seconds, on a circular pattern with a diameter of 200 pixels. Naturally, images appeared in a smaller size (about 150 x 150 pixels) than in experiments 2/a, b, c, and d, however, the objects and their detail were still clearly visible and identifiable. The orientation of the 3 objects was generated in the following way. The orientation of one object was chosen randomly without replacement from: [45, 135] degrees. Then a number was chosen randomly from: [25, 45]. This number then was added to the first number, to get the orientation for the second object. For the orientation of the third object, the same procedure was repeated. A number was selected randomly from: [25, 45], but now, it was subtracted from the orientation of the first object. The rationale behind this orientation generation process was to avoid potential 'gluing' effects between the 3 objects in a particular scene. For instance, if 2 (or even 3) objects had a similar orientation by a fully random selection, then for that given scene, an unintended structure would have been introduced, and therefore, the orientation of objects could be potentially remembered with higher accuracy. The same applies if there had been a 90 degree difference between 2 of the objects accidentally. The interstimulus gap between the scenes was 2 seconds. Participants saw 20 scenes altogether, so the total number of objects was 60, identical to the base experiment of Chapter 2 (Experiment 2/a, 60 SINGLE RANDOM). The instruction for participants was to remember the presented objects and their orientation. After all the images were presented, subjects could optionally take a few minute break before the recall phase.

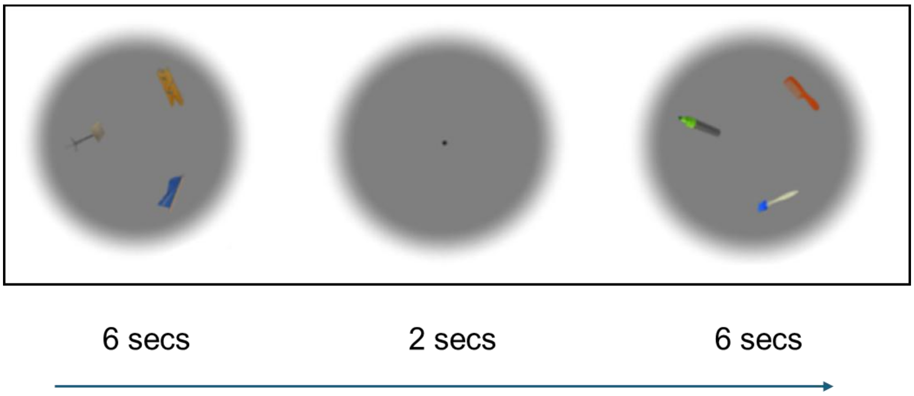

Figure 3.1 Study phase of Experiment 3/a. 3 objects are presented in each trial/scene that are unrelated perceptually (orientation) and also semantically. Participants are instructed to remember the objects and their orientation.

In the recall phase, 60 old (previously presented) and 60 new (previously not presented) objects were shown individually, in a vertical position, for 1 second. Subjects answered on a continuous scale whether they had seen the object in the study phase. In the second response, they had to recall the orientation of the object from the study phase, along with their subjective certainty, identical to Experiment 2/a, b, c, and d.

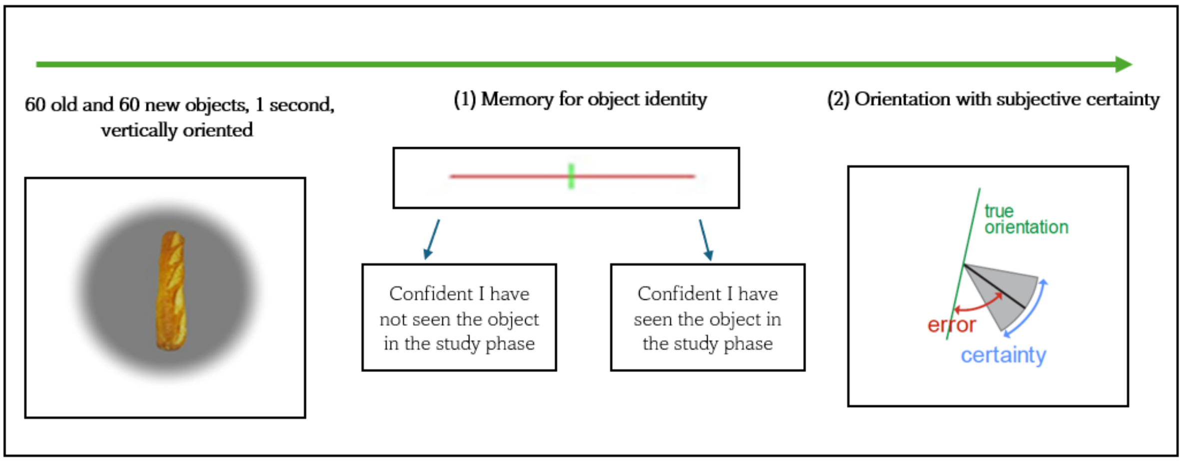

Figure 3.2 Recall phase of Experiment 3/a. Participants give two responses about 50% old, 50% new vertically oriented objects that are shown for 1 second. They indicate on a scale whether they had seen a given object in the study phase, then they draw the orientation with which they remember the object was presented in the study phase. Both responses include participants' subjective confidence/certainty.

Results

Object identity recognition performance was high with mean d' = 2.342, SD = 1.057, with high corresponding subjective confidence responses across subjects (mean = 0.84, SD = 0.13). There was no significant difference between Experiment 2/a (60 SINGLE RANDOM) and Experiment 3/a (60 MULTIPLE RANDOM) in object identity recognition ($M_{exp2/a}$ = 2.281, SD = 0.773; $M_{exp3/a}$ = 2.342, SD = 1.057; $t(49) = 0.237$, $p = 0.814$, Cohen's d = 0.066, BF = 0.287).

Orientation recall accuracy was significantly above chance in Experiment 3/a (M = 0.226, SD = 0.186; $t(24) = 6.088$, $p < 0.001$, BF > 1000). What is more, there was no significant difference between the base experiment from Chapter 2 (Experiment 2/a, 60 SINGLE RANDOM) and the base experiment from Chapter 3 (Experiment 3/a, 60 MULTIPLE

RANDOM) (M_exp2/a = 0.290, SD = 0.168; M_exp3/a = 0.226, SD = 0.186; $t(49) = 1.290$, $p = 0.203$, Cohen's d = 0.361, BF = 0.554).

Orientation responses were well-calibrated in Experiment 3/a, shown by a Pearson correlation between orientation accuracy and certainty across subjects (mean_pearson_r = 0.163, SD = 0.163; $t(24) = 5.019$, $p < 0.001$, Cohen's d = 1.004, BF > 100).

Two people were excluded from the analyses from Experiment 3/a for below chance orientation accuracy.

Discussion

Experiment 3/a echoed the patterns that were shown in Experiment 2/a as well. Subjects could decide whether they had seen a particular object before with high accuracy and subjective confidence. Further, there was a meaningful, above chance performance for object orientation, however, precision was relatively low. In addition, subjective certainty reports for orientation were well-calibrated in the two experiments. These results suggest that even though the display of objects was fundamentally different (multiple as opposed to individual), participants' encoding strategy remained similar in both Experiment 2/a and 3/a. It seems that encoding and recall was done in an episodic, item-based manner, despite the simultaneous presentation in Experiment 3/a.

3.2.2 Experiment 3/b: Recall Efficiency and Representation of Uncertainty in Scenes with Perceptual Connections

There are several components of natural scenes and in all likelihood, there is a hierarchy of the certain possible connections between the elements (let them be perceptual or semantic)

(Kadar & Ben-Shadar, 2012). As a consequence, regularity on the various levels can affect encoding to various extents both in terms of accuracy and uncertainty representation. Experiment 3/b introduces a so-called low-level perceptual, within scene regularity. Here, two of the three objects in a scene has the same orientation, called perceptual gluing. The main question is whether this perceptually salient and elementary level scene structure influences recall accuracy and the representation of uncertainty.

Participants

Thirty-two people participated in Experiment 3/b (60 MULTIPLE PERCEPTUAL GLUING). All of them were recruited through the Hungarian MADS student organization and were mostly Hungarian university students. They received hourly compensation of 3000 HUF (~7.5$) for their participation. The experiment lasted for approximately one hour. The experiment took place in person, in the visionlab of CEU in Budapest. All participants had normal or corrected to normal vision. The study was approved by the Hungarian United Ethical Review Committee for Research in Psychology (EPKEB), and all participants provided informed consent.

Materials and Procedure

Regarding the objects, the same set of 190 objects were used as before. 3 objects were presented simultaneously in each trial. Here, the orientation of 2 of the objects were identical in each scene, which is called perceptual gluing. The objects with the same orientation in the scene are called glued objects. The orientation for both of the glued objects were sampled randomly from: [0, 180] degrees. The third object was unrelated to the other two in terms of its orientation and is called an unglued object. Importantly, it was ensured that the orientation of the unglued object was neither closer than 30 degrees nor further than 60 degrees from the glued ones. In practice, it meant that the orientation of the third object was sampled from: [30, 60] degrees, in distance, from the glued objects. This was in order to avoid the unglued object being perceptually attached to the glued ones by accidentally having a very similar orientation or a 90 degree difference in orientation to the glued objects. Participants saw 20 scenes altogether, each presented for 6 seconds, with a 2 second intertrial gap. The instruction was the same as before, remember the objects and their orientation. Subjects were not informed that there is structure in the input, and they must be attentive to that.

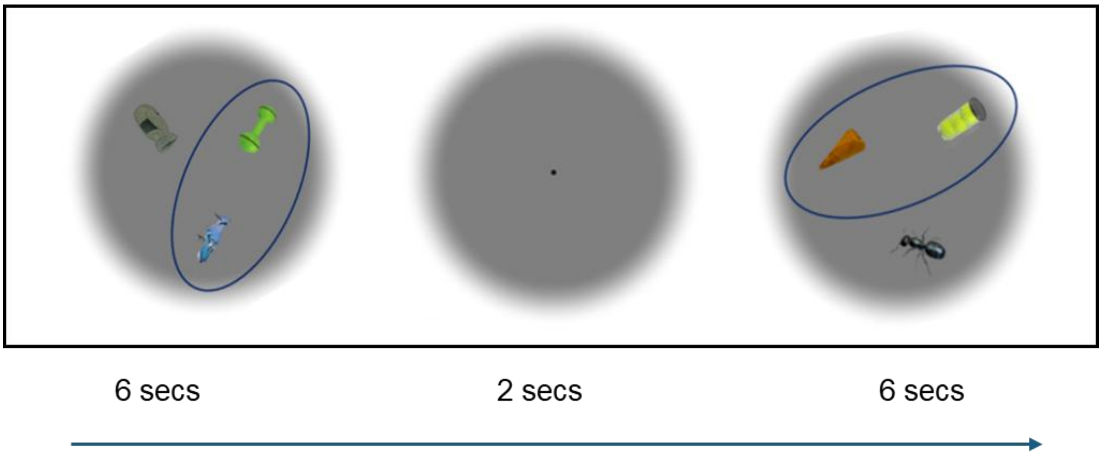

Figure 3.3 Study phase of Experiment 3/b. 3 objects are presented in each scene. 2 of them are related perceptually, they have the same orientation (perceptual gluing). Participants are instructed to remember the objects and their orientation.

The recall phase was identical to Experiment 3/a. Subjects gave the same kinds of responses about 120 objects (60 old, 60 new), namely, object identity recall with a confidence response, and object orientation recall with a certainty response.

Results

Object identity performance was moderately high with mean d' = 2.172, SD = 1.149. Mean confidence across participants was 0.83 with SD = 0.11. There was no significant difference between Experiment 3/a (60 MULTIPLE RANDOM) and Experiment 3/b (60 MULTIPLE PERCEPTUAL GLUING) regarding object identity performance (M_exp3/a = 2.342, SD = 1.057; M_exp3/b = 2.172, SD = 1.149; $t(53) = 0.568$, $p = 0.572$, Cohen's d = 0.154, BF = 0.312).

Orientation recall performance was significantly above chance in Experiment 3/b (M = 0.223, SD = 0.149; $t(29) = 8.197$, $p < 0.001$, BF > 1000). Overall, there was no significant difference between Experiment 3/a and Experiment 3/b regarding orientation performance (M_exp3/a = 0.226, SD = 0.186; M_exp3/b = 0.223, SD = 0.149; $t(53) = 0.078$, $p = 0.938$, Cohen's d = 0.021, BF = 0.274). Further, orientation accuracy for the glued and unglued objects, in Experiment 3/b, was not significantly different, shown by a pairwise t-test. (M_glued = 0.235, SD = 0.161; M_unglued = 0.198, SD = 0.157; $t(29) = 1.592$, $p = 0.122$, Cohen's d = 0.291, BF = 0.601). There was no significant difference between Experiment 3/a and the accuracy of the glued objects from Experiment 3/b (M_exp3/a = 0.226, SD = 0.186; M_exp3/b_glued = 0.235, SD = 0.161; $t(53) = 0.183$, $p = 0.856$, Cohen's d = 0.049, BF = 0.277). Lastly, there was no significant difference between Experiment 3/a and the accuracy of the unglued objects from Experiment 3/b (M_exp3/a = 0.226, SD = 0.186; M_exp3/b_unglued = 0.198, SD = 0.157; $t(53) = 0.598$, $p = 0.553$, Cohen's d = 0.162, BF = 0.317).

The overall correlation between orientation accuracy and certainty was significant in Experiment 3/b (mean_pearson_r = 0.167, SD = 0.148; $t(29) = 6.166$, $p < 0.001$, Cohen's d = 1.126, BF > 1000). Moreover, a pairwise t-test across subjects showed no significant difference in correlation between the glued and unglued objects (M_{glued} = 0.175, SD = 0.185; $M_{unglued}$ = 0.149, SD = 0.212; $t(29) = 0.515$, $p = 0.611$, Cohen's d = 0.094, BF = 0.220).

Two people were excluded from the analysis from Experiment 3/b for below chance orientation accuracy.

Discussion

The results from Experiment 3/b are surprising as the introduced perceptual regularity did close to nothing in terms of recall accuracy and representation of uncertainty. Identity and orientation memory performance was not significantly different from Experiment 3/a (60 MULTIPLE RANDOM), and subjective certainty reports did not become biased quantitatively or qualitatively either. It is surprising because theoretically treating the glued objects as one chunk and encoding only one orientation value for the two could have decreased memory load and led to better recall performance. Even though orientation is a rather arbitrary feature, it was the only regularity in the display, and it was salient. Therefore, the opportunity was there to utilize it for later recall. However, overall results from Experiment 3/b suggest (fully) episodic, item-based encoding that is shown by the non-significant difference in recall accuracy compared to random scenes and by the well-calibrated orientation responses both in the case of glued and unglued objects.

3.2.3 Experiment 3/c: Recall Efficiency and Representation of Uncertainty in Scenes with Semantic Connections

Experiment 3/c went further and investigated whether higher-level semantic connections between the objects within a scene had an effect on the usual measures of recall performance and uncertainty representations.

Participants

Twenty-three people participated in Experiment 3/c (60 MULTIPLE SEMANTIC GLUING). All of them were recruited through the Hungarian MADS student organization, and, as usual, were mostly Hungarian university students. They received hourly compensation of 3000 HUF (~7.5$) for their participation. The experiment was around 1 hour long. The experiment took place in person, in the visionlab of CEU in Budapest. All participants had normal or corrected to normal vision. The study was approved by the Hungarian United Ethical Review Committee for Research in Psychology (EPKEB), and all participants provided informed consent.

Materials and Procedure

The only difference between Experiment 3/a (60 MULTIPLE RANDOM) and Experiment 3/c (60 MULTIPLE SEMANTIC GLUING) is that in the latter 2 out of the 3 objects were semantically related to each other. Consequently, now the stimulus set was slightly different because 2 objects in each scene were selected to be semantically related from the largest stimulus set. But the objects still had to fit the long and narrow criteria, just like before. Otherwise, the presentation times, orientation generation process, and instructions were identical. The semantic relation between two of the objects was always straightforward, yet slightly different across trials. Semantic gluing could mean that the same kind of object was presented (such as 2 guitars with slightly different overall shape, color etc.). Or it could mean objects with a higher-level functional connection between them, for example, a tennis racket and a set of tennis balls, or bird and feather.

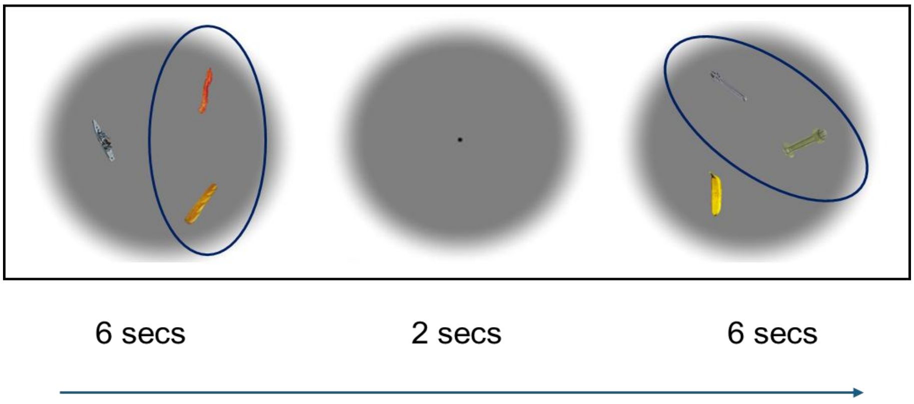

Figure 3.4 Study phase of Experiment 3/c. 3 objects are presented in each scene. 2 of them are related semantically (semantic gluing, bacon-baguette and screw-screwdriver, in the figure above). Participants are instructed to remember the objects and their orientation.

The recall phase design was identical to Experiment 3/a and Experiment 3/b.

Results

Object identity recognition was higher in Experiment 3/c (60 MULTIPLE SEMANTIC GLUING) than in Experiment 3/a (60 MULTIPLE RANDOM). The difference was not significant, although the effect size was moderate, and the Bayes factor could not decipher the difference (M_exp3/a = 2.342, SD = 1.057; M_exp3/c = 2.939, SD = 1.128; t(44) = 1.849, p = 0.071, Cohen's d = 0.547, BF = 1.143). The corresponding confidence ratings across subjects were high with mean = 0.88 and SD = 0.1, in Experiment 3/c.

Overall orientation accuracy was higher in Experiment 3/c than in Experiment 3/a. This effect was borderline significant (M_exp3/a = 0.226, SD = 0.186; M_exp3/c = 0.333, SD

= 0.188; $t(44) = 1.930$, $p = 0.060$, Cohen's $d = 0.571$, $BF = 1.287$). This effect was larger when only the unglued objects were compared from Experiment 3/c to Experiment 3/a ($M_{exp3/a} = 0.226$, $SD = 0.186$; $M_{exp3/c\_unglued} = 0.411$, $SD = 0.187$; $t(44) = 3.361$, $p = 0.002$, Cohen's $d = 0.995$, $BF = 21.023$). Orientation accuracy for the glued objects alone was not significantly higher than overall orientation performance in Experiment 3/a ($M_{exp3/a} = 0.226$, $SD = 0.186$; $M_{exp3/c\_glued} = 0.293$, $SD = 0.228$; $t(44) = 1.105$, $p = 0.275$, Cohen's $d = 0.327$, $BF = 0.480$). Further, within Experiment 3/c, there was a significant difference between the glued and unglued objects ($M_{glued} = 0.293$, $SD = 0.228$; $M_{unglued} = 0.411$, $SD = 0.187$; $t(20) = 2.426$, $p = 0.025$, Cohen's $d = 0.529$, $BF = 2.384$).

Overall correlation between accuracy and certainty was significant in Experiment 3/c ($mean\_pearson\_r = 0.149$, $SD = 0.141$; $t(19) = 4.713$, $p < 0.001$, Cohen's $d = 1.054$, $BF = 192.4$). In addition, there was no significant difference in correlation between the glued and unglued objects across subjects ($M_{glued} = 0.108$, $SD = 0.167$; $M_{unglued} = 0.201$, $SD = 0.231$; $t(19) = 1.501$, $p = 0.150$, Cohen's $d = 0.336$, $BF = 0.610$).

Two participants were excluded from the analyses from Experiment 3/c for below chance orientation accuracy and one further subject from the correlation analyses for giving maximum uncertainty in every single response.

Discussion

Experiment 3/c showed that as opposed to perceptual, semantic regularity did have an effect on recall accuracy. Not only was there a significantly increased overall orientation recall performance compared to Experiment 3/a (60 MULTIPLE RANDOM), there was a significant difference also between glued and unglued objects in Experiment 3/c. Further, interestingly the orientation of unglued objects was recalled with higher accuracy than the orientation of glued objects. This indicates that participants did utilize within scene semantic regularity. What is more, the advantage of unglued objects suggests that they utilized an odd one out encoding strategy, even though the scene regularity seemingly benefited the glued objects. Further, the fact that responses were well-calibrated both in the case of glued and unglued objects, with no significant difference between the two suggests something that the pure episodic experiments also showed that uncertainty representations remained predominantly on an item-based level. On the other hand, it must be mentioned that although there was no significant difference between glued and unglued in calibratedness, the level of correlation was noticeably lower for the glued objects than it used to be in previous experiments (1.5-2.5 range). It indicates that representations did become noisy due to input regularity.

3.2.4 Experiment 3/d: Recall Efficiency and Representation of Uncertainty in Scenes with Perceptual and Semantic Connections

Experiment 3/d went further in adding more within-scene regularity to the input. Experiment 3/b showed that perceptual gluing had almost zero overall effect on how scenes and individual objects are encoded compared to the random baseline experiment (60 MULTIPLE RANDOM). However, it is possible that perceptual and semantic gluing have a synergistic effect when both are present in a scene, increasing memory performance further than semantic connections alone. Experiment 3/d investigated this question by adding both perceptual and semantic regularity to simple scenes.

Participants

Thirty-four people participated in Experiment 3/d (60 MULTIPLE PERCEPTUAL AND SEMANTIC GLUING). All of them were recruited through the Hungarian MADS student organization and were predominantly Hungarian university students. They received hourly compensation of 3000 HUF (~7.5$) for their participation. The experiment lasted for approximately 1 hour. The experiment took place in person, in the visionlab of CEU in Budapest. All participants had normal or corrected to normal vision. The study was approved by the Hungarian United Ethical Review Committee for Research in Psychology (EPKEB), and all participants provided informed consent.

Materials and Procedure

Experiment 3/d was identical to Experiment 3/c (60 MULTIPLE SEMANTIC GLUING) except that here 2 objects were not only semantically related but also perceptually glued, just like in Experiment 3/b (60 MULTIPLE PERCEPTUAL GLUING). The orientation generation process was identical to Experiment 3/b, 2 out of the 3 objects had identical orientations and the third one was in a 30 to 60 degree distance from them. The recall phase design was identical to Experiment 3/a, b, and c.

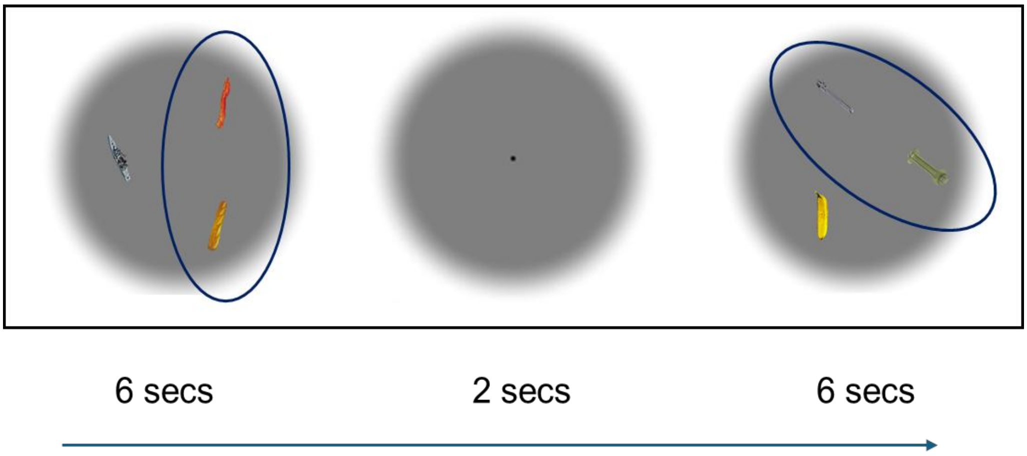

Figure 3.5 Study phase of Experiment 3/d. 3 objects are presented in each scene. 2 of them are related both perceptually (orientation) and semantically. Participants are instructed to remember the objects and their orientation.

Results

Identity recognition performance was not significantly different in Experiment 3/a (60 MULTIPLE RANDOM) and Experiment 3/d (M_exp3/a = 2.342, SD = 1.057; M_exp3/d = 2.605, SD = 1.190; t(56) = 0.873, p = 0.386, Cohen's d = 0.231, BF = 0.369). Similarly to previous experiments, the subjective confidence ratings for object identity were high across subjects (mean = 0.87, SD = 0.098), in Experiment 3/d.

Regarding orientation accuracy, there was an overall significant difference between Experiment 3/a and Experiment 3/d (M_exp3/a = 0.226, SD = 0.186; M_exp3/d = 0.33, SD = 0.225; t(56) = 1.995, p = 0.051, Cohen's d = 0.529, BF = 1.373). This effect was even more pronounced when only the glued objects were compared to Experiment 3/a (M_exp3/a = 0.226, SD = 0.186; M_exp3/d_glued = 0.367, SD = 0.238; t(56) = 2.454, p = 0.017, Cohen's d = 0.651, BF = 3.1). Further, there was no significant difference between Experiment 3/a and the unglued objects from Experiment 3/d (M_exp3/a = 0.226, SD =

0.186; M_exp3/d_unglued = 0.275, SD = 0.251; t(56) = 0.817, p = 0.417, Cohen's d = 0.217, BF = 0.354). Comparing orientation accuracy between the glued and unglued objects within Experiment 3/d, a pairwise t-test yielded a significant difference (M_glued = 0.367, SD = 0.238; M_unglued = 0.275, SD = 0.251; t(32) = 2.759, p = 0.010, Cohen's d = 0.480, BF = 4.5).

A Pearson correlation between orientation accuracy and certainty was significant in Experiment 3/d (mean_pearson_r = 0.175, SD = 0.182; t(32) = 5.523, p < 0.001, Cohen's d = 0.961, BF > 1000). In addition, there was no significant difference comparing the correlation between the glued and unglued objects by a pairwise t-test across the mean correlation of subjects (M_glued = 0.194, SD = 0.206; M_unglued = 0.101, SD = 0.259; t(32) = 1.747, p = 0.090, Cohen's d = 0.304, BF = 0.727).

One participant was excluded from the analyses from Experiment 3/d for below chance orientation accuracy.

Discussion

Overall identity and orientation performance did not increase by the extra regularity introduced in Experiment 3/d compared to Experiment 3/c. However, in Experiment 3/d, the orientation of the glued objects was remembered significantly better as opposed to Experiment 3/c where the unglued objects had significantly higher recall accuracy. Overall, these results indicate that semantic gluing alone carries the weight in stronger encoding, although perceptual gluing changes the strategy in terms of the type of objects (glued or unglued) that are encoded with higher accuracy. In that way, perceptual and semantic gluing did have a synergistic effect. What is more, certainty representations for unglued objects were not as well calibrated in Experiment 3/d as they were with glued objects, so the dissociation with Experiment 3/c was apparent in calibratedness as well. Perceptual and semantic gluing together biased calibratedness for the unglued objects, just like semantic gluing alone biased calibratedness for the glued objects.

Further results, analyses, experiments 3/a, b, c, d altogether

So far, I have demonstrated that regarding orientation recall accuracy, participants are more susceptible to higher level input regularity as it was only in Experiment 3/c and 3/d that either glued or unglued objects had a significant recall advantage. At this point, the question arises, why did the recall advantage happen in those experiments? Was it because glued objects were simply perceptually and conceptually salient and, as a consequence, participants simply spent proportionally more time looking at glued objects. Or some kind of chunking happened in people's representation as well, which followed the structure of the input. The following analysis attempts to decipher this by measuring the within scene correlation in recall accuracy between object pairs in the scenes, in all experiments. The logic is that if it can be shown that there is significant correlation in recall accuracy between object pairs from scenes, then it is a strong indication that those objects are linked (or chunked) in long-term memory also. Measuring the correlation in accuracy between the two glued objects (glued chunk) within a scene is straightforward. However, since there is only one unglued object, defining the 'unglued chunk' is a bit more difficult. In this analysis, in each scene, it is randomly selected which glued object is correlated with the unglued object, in terms of orientation recall accuracy. In Experiment 3/a (60 MULTIPLE RANDOM) where there is no gluing in the input, it is again straightforward to define 'chunks', in each scene, each object pair is measured regarding the correlation in accuracy.

Starting with Experiment 3/a, there were 3 object pairs within the scenes: object1-object2, object1-object3, object2-object3. The mean correlation in accuracy between object 1 and object 2 within scenes across subjects was not significantly above chance (M = 0.002, SD = 0.229; t(24) = 0.036, p = 0.971, Cohen's d = 0.007, BF = 0.211). The same measure for object 1 and object 3 was not significantly above chance either (M = -0.032, SD = 0.203; t(24) = -0.793, p = 0.435, Cohen's d = -0.159, BF = 0.280). Finally, the correlation in accuracy between object 2 and object 3 was likewise not significantly above chance (M = -0.018, SD = 0.321; t(24) = -0.286, p = 0.778, Cohen's d = -0.057, BF = 0.219).

In Experiment 3/b, the correlation between the two glued objects was significantly above chance (M = 0.122, SD = 0.240; t(29) = 2.795, p = 0.009, Cohen's d = 0.509, BF = 4.777). However, this was not the case for the correlation between the unglued object and the glued objects (M = 0.049, SD = 0.239; t(29) = 1.116, p = 0.274, Cohen's d = 0.204, BF = 0.342). Further, the difference between the glued chunk (correlation of the two glued objects) and the unglued chunk (correlation of the unglued object and the randomly selected glued object) was not significant in Experiment 3/b (M_glued_chunk = 0.122, SD = 0.240; M_unglued_chunk = 0.049, SD = 0.239; t(29) = 1.210, p = 0.236, Cohen's d = 0.221, BF = 0.377).

In Experiment 3/c, the correlation in accuracy of the two glued objects was not significantly above chance (M = 0.089, SD = 0.262; t(20) = 1.555, p = 0.136, Cohen's d = 0.339, BF = 0.642). Further, it was the case for the unglued object and the randomly selected glued object as well (M = 0.060, SD = 0.254; t(20) = 1.077, p = 0.294, Cohen's d = 0.235, BF = 0.380). Correspondingly, the difference between the glued chunk and the unglued chunk was not significant either (M_glued_chunk = 0.089, SD = 0.262; M_unglued_chunk = 0.060, SD = 0.254; t(20) = 0.367, p = 0.717, Cohen's d = 0.080, BF = 0.242).

In Experiment 3/d, the correlation in accuracy in the glued chunk was significantly above chance ($M = 0.310$, $SD = 0.286$; $t(32) = 6.213$, $p < 0.001$, Cohen's $d = 1.081$, $BF > 1000$). This was not true for the unglued chunk since their correlation in accuracy was not significantly above chance ($M = -0.020$, $SD = 0.209$; $t(32) = 0.541$, $p = 0.592$, Cohen's $d = 0.094$, $BF = 0.213$). Lastly, the difference in correlation between the glued and unglued chunk was highly significant ($M_{glued\_chunk} = 0.310$, $SD = 0.286$; $M_{unglued\_chunk} = -0.020$, $SD = 0.209$; $t(32) = 5.045$, $p < 0.001$, Cohen's $d = 0.878$, $BF > 1000$).

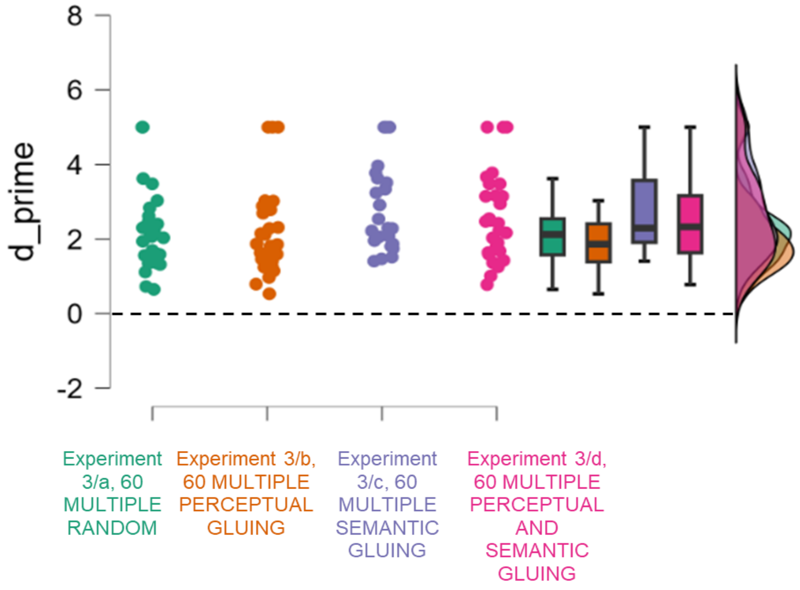

Figure 3.6 Object identity performance in Experiment 3/a, b, c, d. The dashed line indicates chance level performance. There was only a mild difference between the experiments, although the slight advantage of the semantic experiments is apparent in the case of identity recognition too.

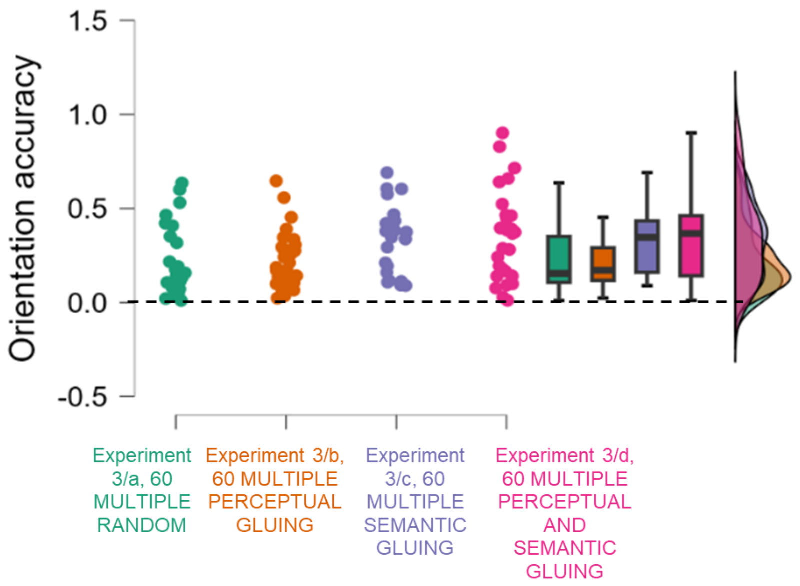

Figure 3.7 Object orientation performance in Experiment 3/a, b, c, d. The dashed line indicates chance level performance. Regarding overall performance, semantic gluing contributed to the majority of the improvement compared to random and only perceptual gluing.

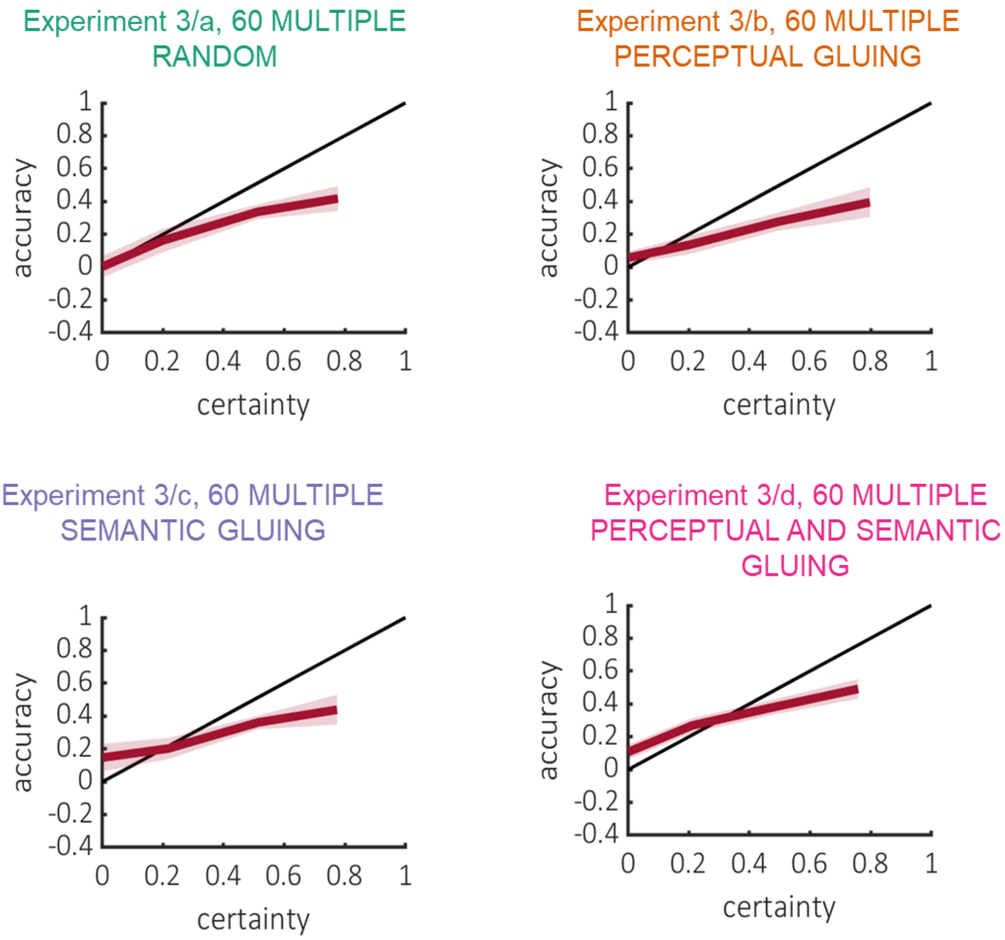

Figure 3.8 Calibratedness in Experiment 3/a, b, c, d. As has been shown with the pure episodic experiments in Chapter 2, the significant correlation between certainty and accuracy indicates probabilistic encoding of the stimulus throughout experiments in Chapter 3 as well.

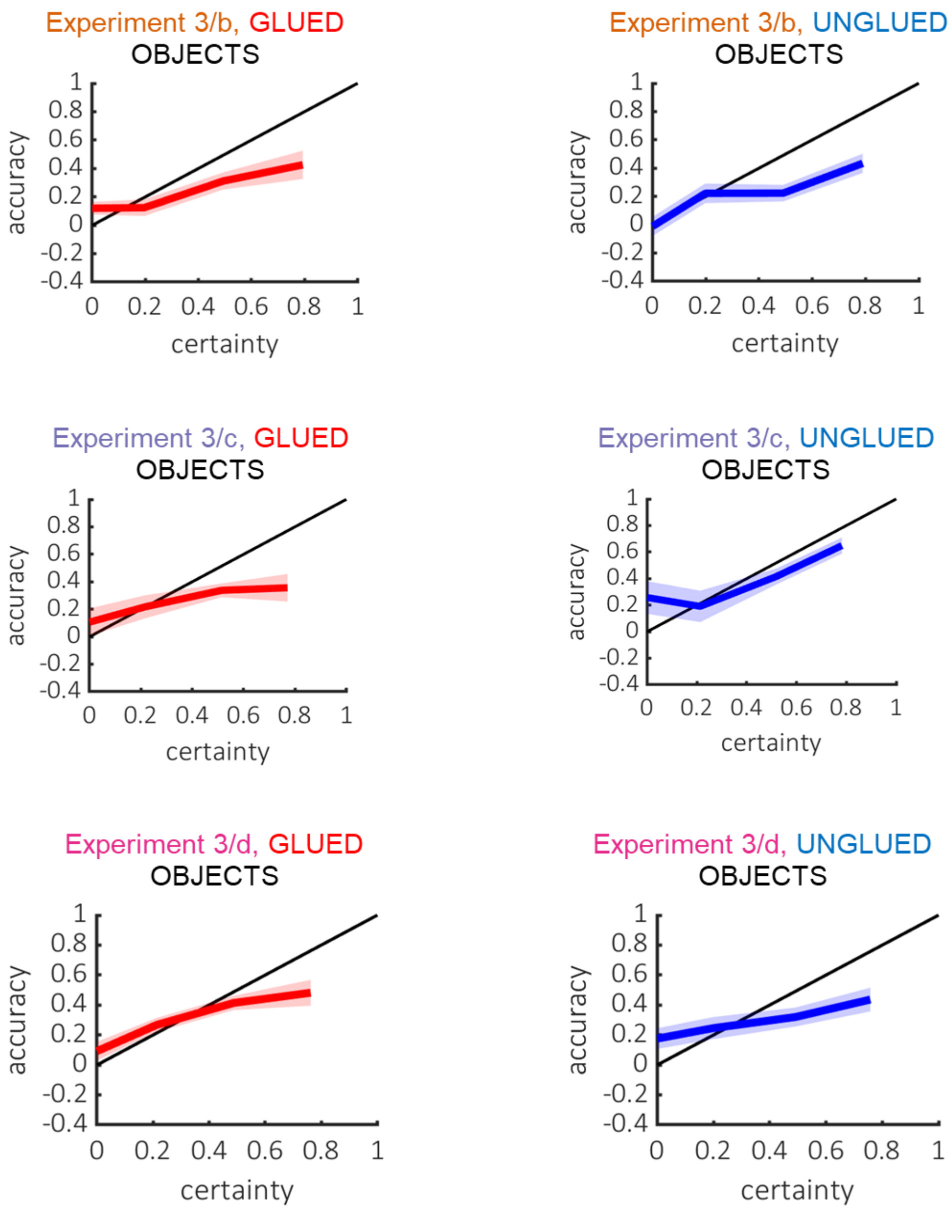

Figure 3.9 Calibratedness in Experiment 3/b, c, d, separated to glued and unglued objects. The separation of the calibration plots for glued and unglued objects revealed that certainty representation remained calibrated to orientation accuracy, although as input regularity increased, the calibration curves tended to be noisier both quantitatively and qualitatively (Exp 3/c GLUED and Exp 3/d UNGLUED).

3.3 Summary

In summary, Chapter 3 presented four experiments. Experiment 3/a showed that simultaneous as opposed to sequential presentation (Experiment 2/a) had very little effect on overall recall performance. Experiments 3/b, c, d demonstrated that from various within scene regularity types (perceptual, semantic, perceptual + semantic), semantic regularity carried most of the weight in significantly improving orientation recall accuracy. Perceptual + semantic gluing had a synergistic effect in terms of which type of object (glued or unglued) is recalled with higher accuracy, as in Experiment 3/c unglued objects had the advantage, whereas in Experiment 3/d, glued objects had the advantage in orientation recall precision. In addition, from further analysis it turned out that complex within scene regularity contributed to helping people form chunks, based on scene structure, which was especially apparent in Experiment 3/d. Lastly, regarding uncertainty, Experiment 3/a provided further proof for the similarity between perceptual decision-making and long-term episodic memory by showing that in long-term memory, in simple scenes, the representation of uncertainty is object-based. Further, as scene regularity got more complex, certainty representations remained well-calibrated to accuracy, but also became noisier, suggesting the incorporation of semanticity into the memory representation.

CHAPTER 4

RECALL EFFICIENCY AND REPRESENTATION OF UNCERTAINTY IN EPISODIC MEMORIES WITH OVERARCHING STATISTICAL REGULARITY

4.1 Introduction

Previous chapters in the dissertation have investigated the recall accuracy and the representation of uncertainty in pure episodic memories or in episodic memories in scenes with local semantic connections. However, in the memory literature what is usually understood to be semantic memory is extracting recurring, meaningful patterns from the input, often over extended periods of time (Brady et al., 2011; Chetverikov et al., 2016). This is the kind of semantic memory or ensemble representation that Chapter 4 also investigates. Regarding the connection between long-term episodic and semantic memory, previous research suggests that they are partly separate memory systems (Graham et al., 2000, Sadeh et al., 2016). Moreover, it is a challenge to uncover the interaction of the two memory systems and point to precise computational functions of each (Hemmer & Steyvers, 2009; Nagy et al, 2020). It is established that statistical regularity in the input oftentimes biases the recall precision of individual items (Hemmer & Steyvers, 2009; Utochkin & Brady, 2020). It has also been demonstrated that learning the semantic regularity of the input, or constructing prior knowledge, improves memory performance for individual items as well, probably because based on general patterns, people can make intelligible guesses about items for which episodic memory is dim (Hemmer & Steyvers, 2009). However, the exact way in which overarching input regularity boosts people's long-term recall accuracy is unclear. It is often assumed that with overarching input regularity, episodic memory for individual items and details is generalized, lost, and becomes biased (e.g., Nagy et al., 2020; Sherman & Turk-Browne, 2020). But alternatively, it can be the case that with structured input, episodic vividness also increases which is then wrongly taken to be ‘semantic memory’, which is a possibility that is not discussed in the literature, to my knowledge. Further, individual differences in long-term memory are less often discussed (e.g., Hemmer, Tauber & Steyvers (2015). It is not clear what kind of representations and strategies people who are more successful or less successful in a given memory task apply. Finally, previous chapters in the dissertation argued that pure episodic memories or episodic memories with local semantic regularity are encoded and recalled probabilistically. It is unknown whether the probabilistic nature of the representation holds when items/objects are recalled individually, at the same time, summary statistics/overarching regularity is imposed on the input.

Chapter 4 shows in alignment with previous research that overarching semantic input regularity does increase long-term recall compared to when there is no regularity in the input. In addition, Chapter 4 uncovers the nature of the increase in recall accuracy and shows that participants who perform well (in the top 50%) do so because they have formed true semantic representations of the input, i.e. learnt the prior and followed its structure upon recall while their episodic memory for the item was dim (based on subjective reports). What is more, learning the regularity of the typical items supposedly helps good performers to remember outliers better than bad memory performers, who extract some of the regularity from the input but not as efficiently, and their lack of semantic learning also presumably inhibits episodic learning. All this demonstrates that more efficient recall when there is structure in the input is due to an acquired semantic representation, at least in the case of good performers. Finally, Chapter 4 shows that with long-term semantic regularity, the probabilistic nature of representations remains but becomes biased by the input structure. The bias originated from subjective certainty representations remaining item-based (even though explicit knowledge of the summary statistics could be utilized in the uncertainty response), while orientation responses benefited from learning the regularity. All this results in long-term memory representations where episodic (object-based uncertainty) and semantic (knowledge of input structure) parts are not fully integrated.

4.2.1 Experiment 4/a: The Effect of Overarching Semantic Regularity on Episodic Memories

Participants

Thirty-six people participated in Experiment 4/a (60 SINGLE BUMPED). All of them were recruited through the Hungarian MADS student organization and were mostly Hungarian university students. They received hourly compensation of 3000 HUF (~7.5$) for their participation. The experiment lasted for around one hour. The experiment took place in person, in the visionlab of CEU in Budapest. All participants had normal or corrected to normal vision. The study was approved by the Hungarian United Ethical Review Committee for Research in Psychology (EPKEB), and all participants provided informed consent.

Materials and Procedure

The materials and procedure used in Experiment 4/a were almost identical to the base experiment of the dissertation, Experiment 2/a (60 SINGLE RANDOM). The experiment consisted of 3 practice sessions and 1 test session. The first session was the usual learning the connection between line length and wedge width on the tablet. The second practice session targeted learning the orientation response with subjective certainty reports. Both practice sessions were identical to Experiment 2/a.

The long-term memory part of the experiment consisted of one practice and one test session, each containing a study and a recall phase. The same set of 190 objects were used, as in previous experiments. 60 (5 during practice) objects were presented individually and sequentially in the study phase for 1 second, with a 2 second intertrial gap. The only difference to Experiment 2/a is that now, the orientation of the objects was sampled from a Gaussian (instead of a uniform) distribution, with mean = 30 degrees, and standard deviation = 30 degrees. Participants were instructed to remember the objects and their orientation. They were not told to look for any regularity in the input or that there will be any regularity, in the first place. In the recall phase, they had to make decisions about 120 objects (10 during practice) in total that were vertically oriented and shown for 1 second. First, they answered whether they had seen an object or not in the study phase. This response was done on the tablet, using the same horizontal scale as before. Second, they had to recall the orientation of the objects during the study phase by drawing a line on the tablet, with line length corresponding to subjective certainty. These responses were identical to previous experiments.

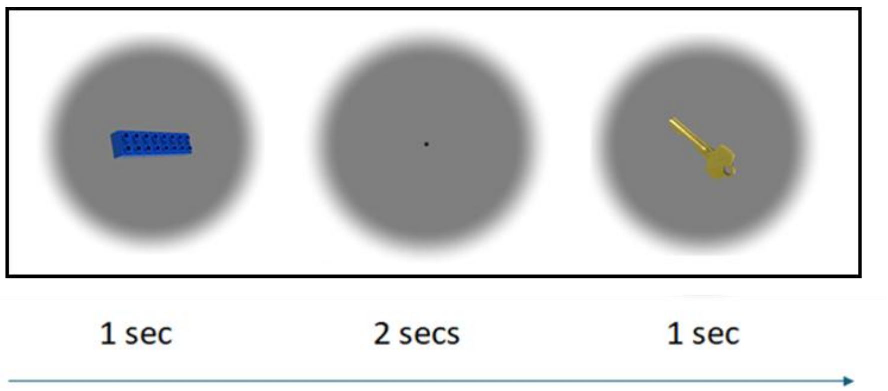

Figure 4.1 Study phase in Experiment 4/a, b, c. Individual objects (60, 30, 90) are presented sequentially. The orientation of the objects is sampled from a normal (mean = 30 degrees, standard deviation = 30 degrees) instead of a uniform distribution in previous experiments. Participants are instructed to remember the objects and their orientation.

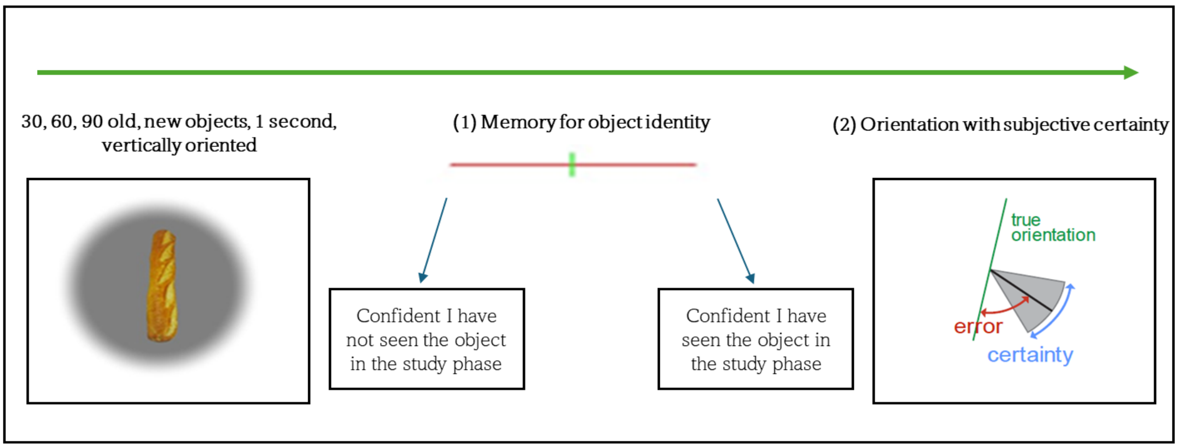

Figure 4.2 Recall phase in Experiment 4/a, b, c. The setup is identical to previous experiments. Participants give two responses about 50% old, 50% new vertically oriented objects that are shown for 1 second. They indicate on a scale whether they had seen a given object in the study phase, then they draw the orientation with which they remember the object was presented in the study phase. Both responses include participants' subjective confidence/certainty.

In addition, in Experiment 4/a participants completed an after-experiment interview investigating how much they have (semantically) learnt from the input. During the interview I read out loud the questions to participants and wrote down the answer they gave.

The idea was to ask increasingly more detailed questions about the experiment and the input (see appendix). Eventually, participants were separated into two groups: implicit/no learners and explicit learners. When a participant said that s/he did not notice anything about the orientation of the objects but later in more specific questions s/he showed signs of awareness of the input structure (for instance, said that the most frequent orientation was around 30 degrees (mean of the input Gauss)) then s/he was classified as an implicit learner. When they said that they did not notice anything at all, not even at more specific questions, they were labeled as 'no learners' but were analyzed together with implicit learners. When a subject claimed right away that s/he noticed that the orientation of most objects was around 30 degrees, in other words, could clearly pinpoint what was going on in the input then they were classified as explicit learners.

Results

Based on the after-experiment questionnaire ~14% of participants showed no learning considering the orientation of the object. ~19% of participants were implicit learners, and the rest ~66% clearly fell into the explicit category.

Object identity memory performance did not increase significantly in Experiment 4/a compared to Experiment 2/a (60 SINGLE RANDOM) ($M\_exp2/a = 2.281$, $SD = 0.773$; $M\_exp4/a = 2.519$, $SD = 1.067$; $t(59) = 0.965$, $p = 0.339$, Cohen's $d = 0.250$, $BF = 0.389$) The corresponding subjective confidence ratings for identity were high with mean = 0.89 and SD = 0.08 in Experiment 4/a.

Orientation recall accuracy, in Experiment 4/a, was significantly above chance (compared to 0) ($M = 0.407$, $SD = 0.164$; $t(34) = 14.706$, $p < 0.001$, Cohen's $d = 2.486$, $BF > 1000$). This result applies when chance is considered to be 0, which can be called an episodic chance level, and it is based on the assumption that subjects do not learn any regularity from the input. However, assuming that the semantic regularity is learnt, the chance level should be higher because subjects do not have to remember each item individually but can use their semantic knowledge. Therefore, besides the episodic baseline, two separate analyses/simulations were run to establish semantic chance levels and compare participants' performance to those as well.

The first simulation assumed that participants learn the mean and the standard deviation of the input distribution and draw random samples from that when they give their orientation responses to a particular object. This kind of learning here is called "full semantic" learning. Orientation recall accuracy was slightly significantly above this baseline (~0.35) in Experiment 4/a ($M = 0.407$, $SD = 0.164$; $t(34) = 2.052$, $p = 0.048$, Cohen's $d = 0.347$, $BF = 1.167$). The second simulation assumed that participants store a single value from the input in memory, the mean of the input distribution, and use this value exclusively when they give their orientation answer. This kind of learning here is called "lazy semantic" learning. Orientation recall accuracy was significantly below this baseline (~0.6) in Experiment 4/a ($M = 0.407$, $SD = 0.164$; $t(34) = 6.986$, $p < 0.001$, Cohen's $d = 1.181$, $BF > 1000$). In addition, in Experiment 4/a overall orientation performance was significantly different from Experiment 2/a (60 SINGLE RANDOM) ($M_{exp2/a} = 0.290$, $SD = 0.168$; $M_{exp4/a} = 0.407$, $SD = 0.164$; $t(59) = 2.729$, $p = 0.008$, Cohen's $d = 0.707$, $BF = 5.4$).

Further analysis was performed on the data from Experiment 4/a. Presented objects from the study phase were divided into two categories, namely, middle and outer. Middle meant objects with an orientation that was one standard deviation within the mean of the input Gaussian. Outer meant objects with an orientation that was more than one standard deviation from the mean. There was a significant difference between orientation recall performance for middle and outer objects (M_middle = 0.446, SD = 0.183; M_outer = 0.324, SD = 0.229; t(34) = 2.921, p = 0.006, Cohen's d = 0.494, BF = 6.471).

Regarding calibratedness, the correlation between orientation recall accuracy and subjective certainty was significantly above chance in Experiment 4/a (mean_pearson_r = 0.126, SD = 0.170, t(34) = 4.402, p < 0.001, Cohen's d = 0.744, BF > 100). The difference in calibratedness between the middle and outer objects was not significant (M_middle = 0.118, SD = 0.244; M_outer = 0.133, SD = 0.269; t(34) = 0.237, p = 0.814, Cohen's d = 0.040, BF = 0.186).

One participant was excluded from the analysis from Experiment 4/a for below chance orientation accuracy.

Considering the after-experiment interview, interestingly, in terms of overall orientation recall accuracy, explicit learners did not perform significantly better than implicit/no learners (M_explicit = 0.39, SD = 0.178; M_implicit_&_no = 0.42, SD = 0.14; t(33) = 0.358, p = 0.722, Cohen's d = 0.128, BF = 0.355). Regarding calibratedness, there was no significant difference between the two types of participants either (M_explicit = 0.112, SD = 0.191; M_implicit_&_no = 0.153, SD = 0.1227; t(33) = 0.665, p = 0.511, Cohen's d = 0.237, BF = 0.4).

Discussion

In terms of object identity performance, Experiment 4/a did not bring surprising results. D' was high, just like in previous experiments, with high corresponding subjective confidence ratings, although participants did not remember every single object, even on the level of identity. Orientation recall accuracy, on the other hand, improved significantly from the case when there was no input regularity (Experiment 2/a, 60 SINGLER RANDOM). The cause of this effect is not straightforward. It can simply occur because the task in Experiment 4/a was simpler, i.e. the range of potential orientations was smaller. It can also be that because of the input regularity (semanticity) participants were able to better remember the objects purely episodically and hence the increase in orientation performance. However, further analyses in Chapter 4 will reveal that it is much more likely that at least some participants learnt the input regularity and used it efficiently when they gave their orientation responses. Regarding calibratedness the overall quantitative pattern, was similar to the previous experiments, namely, that the correlation between orientation accuracy and certainty was significant. However, there was an apparent bias in calibratedness, which was more prominent with middle objects. In the calibration curve, there was a 'horizontal shift', meaning that even when subjects' certainty was low, they could supposedly utilize the input statistics to give reasonable orientation responses, however, their subjective certainty reports did not correspond to this knowledge.

One further interesting result from Experiment 4/a was the after-experiment questionnaire. From that it appears that memory performance does not correspond to what participants verbally claim after the experiment. There was no significant difference between implicit/no learners and explicit learners in overall orientation accuracy and calibratedness. If anything, the overall accuracy and calibratedness of implicit/no learners were slightly higher than explicit learners. This pattern that there is no correspondence between the results and the after-experiment questionnaire will also be apparent in upcoming analyses.

4.2.2 Experiment 4/b: The Effect of Overarching Semantic Bias on Episodic Memories with a Decreased Set Size

Experiment 4/a served as the bridge between Chapter 2 and Chapter 4, as in both cases the base experiment used 60 objects in the study phase. In Chapter 2, the main question behind decreasing and increasing set size was whether it changes recall accuracy. Chapter 4, in addition, asks whether set size changes the encoding strategy in terms of semanticity and episodicness. In other words, Experiment 4/b investigates whether semantic learning occurs the same way with 30 objects as it does with 60, and later with 90 objects.

Participants

Thirty-three people participated in Experiment 4/b (30 SINGLE BUMPED). All of them were recruited through the Hungarian MADS student organization, and most of them were Hungarian and university students. They received hourly compensation of 3000 HUF (~7.5$) for their participation in the one hour long experiment. The experiment took place in person, in the visionlab of CEU in Budapest. All participants had normal or corrected to normal vision. The study was approved by the Hungarian United Ethical Review Committee for Research in Psychology (EPKEB), and all participants provided informed consent.

Materials and Procedure

The materials and procedure were identical to Experiment 4/a, except that now 30 objects were presented in the study phase and participants gave responses about 60 objects in total in the recall phase.

Results

Object identity performance was significantly different in Experiment 4/a (60 SINGLE BUMPED) and in Experiment 4/b (30 SINGLE BUMPED) ($M\_exp4/a = 2.519$, $SD = 1.067$; $M\_exp4/b = 3.309$, $SD = 1.362$; $t(66) = 2.669$. $p = 0.010$, Cohen's $d = 0.648$, $BF = 4.778$). Subjective confidence ratings for identity were high across subjects in Experiment 4/b with mean = 0.88, $SD = 0.12$.

Orientation recall performance was significantly above change, the episodic baseline (0) in Experiment 4/b ($M = 0.498$, $SD = 0.161$; $t(32) = 17.732$, $p < 0.001$, Cohen's $d = 3.087$, $BF > 1000$). Further there was a significant difference between Experiment 4/a (60 SINGLE BUMPED) and Experiment 4/b (30 SINGLE BUMPED) ($M\_exp4/a = 0.407$, $SD = 0.164$; $M\_exp4/b = 0.498$, $SD = 0.161$; $t(66) = 2.321$. $p = 0.023$, Cohen's $d = 0.563$, $BF = 2.368$). Considering the two semantic baselines, orientation accuracy was significantly above the full semantic baseline (~0.35) in Experiment 4/b ($t(32) = 4.602$, $p < 0.001$, Cohen's $d = 0.801$, $BF > 100$) but significantly below the lazy semantic baseline (~0.6) ($t(32) = 3.369$, $p = 0.002$, Cohen's $d = 0.586$, $BF = 17.671$). Further, there was a significant difference between middle and outer objects ($M\_middle = 0.542$, $SD = 0.187$; $M\_outer = 0.394$, $SD = 0.279$; $t(32) = 2.742$. $p = 0.010$, Cohen's $d = 0.477$, $BF = 4.395$) in Experiment 4/b.

The correlation between orientation accuracy and certainty was significantly above chance across subjects in Experiment 4/b ($mean\_pearson\_r = 0.170$, $SD = 0.207$, $t(32) = 4.715$, $p < 0.001$, Cohen's $d = 0.821$, $BF > 100$). In addition, there was no significant difference in correlation between the middle and outer objects in Experiment 4/b ($M\_middle = 0.150$, $SD = 0.252$; $M\_outer = 0.078$, $SD = 0.350$; $t(32) = 1.041$. $p = 0.305$, Cohen's $d = 0.181$, $BF = 0.306$). The outer objects in Experiment 4/b was the first case in the dissertation, when calibratedness was not significantly different from zero ($t(32) = 1.287$, $p = 0.207$, Cohen's $d = 0.227$, $BF = 0.396$).

Discussion

Based on the first set of analyses, the learning strategy in Experiment 4/b (30 SINGLE BUMPED) behaved differently in some regards to learning in Experiment 4/a (60 SINGLE BUMPED). Although, in both cases there was a significant difference between middle and outer objects in accuracy, with 30, the magnitude was smaller. In calibratedness, the difference between middle and outer was not significant, what is more, responses for middle objects were a bit more calibrated than responses for outer objects, unlike in the case of 60 objects. In addition, the calibratedness of outer objects did not reach significance in Experiment 4/b. This result is puzzling, as this kind of calibration pattern is the opposite of every other experiment with overarching regularity. One possible explanation for this is that the study phase of Experiment 4/b was simply too short and therefore more confusing. Participants might have realized already that there is regularity in the input but there was not enough time to adapt to the semantic-episodic pattern fully. This kind of confusion could have caused the loss of significant calibration in the case of outer objects.

4.2.3 Experiment 4/c: The Effect of Overarching Semantic Bias on Episodic Memories with an Increased Set Size

Experiment 4/c investigates the same questions as Experiment 4/a, and b. As set size changes, whether and how do recall accuracy and encoding strategies change, regarding semanticity and episodicness.

Participants

Twenty-seven people participated in Experiment 4/c (90 SINGLE BUMPED). All of them were recruited through the Hungarian MADS student organization, and most of them were Hungarian and university students. They received hourly compensation of 3000 HUF (~7.5$) for their participation in the one hour long experiment. The experiment took place in person, in the visionlab of CEU in Budapest. All participants had normal or corrected to normal vision. The study was approved by the Hungarian United Ethical Review Committee for Research in Psychology (EPKEB), and all participants provided informed consent.

Materials and Procedure

The materials and procedure were identical to Experiment 4/a, except that now 90 objects were presented in the study phase and participants gave responses about 180 objects in total in the recall phase.

Results

Object identity performance decreased but not significantly in Experiment 4/c (90 single BUMPED) compared to Experiment 4/a (60 SINGLE BUMPED) ($M_{exp4/a} = 2.519$, $SD = 1.067$; $M_{exp4/c} = 2.061$, $SD = 1.042$; $t(60) = 1.691$, $p = 0.096$, Cohen's $d = 0.433$, $BF = 0.859$). The corresponding subjective confidence levels in Experiment 4/c were high with mean = 0.857, SD = 0.088.

Orientation recall performance was significantly above chance in Experiment 4/c, considering the episodic baseline (0) ($M = 0.394$, $SD = 0.147$; $t(26) = 13.936$, $p < 0.001$, Cohen's $d = 2.682$, $BF > 1000$). In addition, performance was not significantly above the full semantic baseline (~0.35) ($t(26) = 1.666$, $p = 0.108$, Cohen's $d = 0.321$, $BF = 0.689$) and was significantly below the lazy semantic baseline (~0.6) ($t(26) = 6.927$, $p < 0.001$, Cohen's $d = 1.333$, $BF > 1000$). Further, there was no significant difference in orientation accuracy between Experiment 4/a and Experiment 4/c ($M_{exp4/a} = 0.407$, $SD = 0.164$; $M_{exp4/c} = 0.394$, $SD = 0.147$; $t(60) = 0.316$, $p = 0.753$, Cohen's $d = 0.081$, $BF = 0.272$). However, orientation accuracy in Experiment 4/c was significantly worse than in Experiment 4/b (30 SINGLE BUMPED) ($M_{exp4/b} = 0.498$, $SD = 0.1616$; $M_{exp4/c} = 0.394$, $SD = 0.147$; $t(58) = 2.589$, $p = 0.012$, Cohen's $d = 0.672$, $BF = 4.047$). Further, there was a significant difference between the orientation accuracy of middle and outer objects in Experiment 4/c ($M_{middle} = 0.440$, $SD = 0.155$; $M_{outer} = 0.296$, $SD = 0.212$; $t(26) = 3.747$, $p < 0.001$, Cohen's $d = 0.721$, $BF = 37.3$).

In terms of the calibratedness of orientation uncertainty, a Pearson correlation was significant between orientation accuracy and certainty ($mean\_pearson\_r = 0.2$, $SD = 0.184$, $t(26) = 5.655$, $p < 0.001$, Cohen's $d = 1.088$, $BF > 1000$) in Experiment 4/c. In addition, there was no significant difference in correlation between the middle and outer objects ($M_{middle} = 0.177$, $SD = 0.232$; $M_{outer} = 0.259$, $SD = 0.190$; $t(26) = 1.741$, $p = 0.094$, Cohen's $d = 0.335$, $BF = 0.766$).

Discussion

Results in Experiment 4/c echoed the results from Experiment 4/a. There was a significant difference in orientation accuracy between middle and outer objects, which is one indication of semantic learning. But interestingly, in Experiment 4/c, performance for both middle and outer objects dropped compared to Experiment 4/a. This is puzzling, as the opposite pattern would have been expected because once the input regularity is learned, performance for middle objects should stagnate, whereas outer performance should decrease significantly as more episodic details are lost with an increased set size. These preliminary results indicate that even though there are signs of semantic learning, it is not universal across subjects. Further analyses in upcoming paragraphs intend to uncover the amount of semantic and episodic learning across subjects and their relation to overall orientation performance.

Further results, analyses, Experiment 4/a, b, c altogether

The rationale behind the upcoming set of analysis was that it could provide further insights into the amount and quality of semantic and episodic learning at various performance levels among subjects. Most importantly it can decide whether there is true semantic learning from the input at various performance levels. True semantic learning means that subjects answer almost exclusively based on characteristics of the input regularity (such as mean and standard deviation) and have very limited episodic memory. In the analysis participants in each experiment have been divided into two groups, "good" and "bad" performers. This division was done based on overall orientation accuracy. Good performer meant ~top 50% of subjects considering overall orientation accuracy and bad performer meant ~bottom 50%.

First, the difference between middle and outer object accuracy among good and bad performers was analyzed. In Experiment 4/a (60 SINGLE BUMPED), in the case of middle objects, the difference between good and bad performers was highly significant ($M_{good} = 0.591$, $SD = 0.118$; $M_{bad} = 0.310$, $SD = 0.116$; $t(33) = 7.086$, $p < 0.001$, Cohen's $d = 2.397$, $BF > 100\,000$). This comparison yielded high significance in Experiment 4/b (30 SINGLE BUMPED) as well ($M_{good} = 0.696$, $SD = 0.092$; $M_{bad} = 0.379$, $SD = 0.104$;

$t(31) = 9.286$, $p < 0.001$, Cohen's $d = 3.234$, BF > 1000). The difference in middle object accuracy was significant in Experiment 4/c (90 SINGLE BUMPED) too ($M_{good} = 0.561$, $SD = 0.119$; $M_{bad} = 0.328$, $SD = 0.080$; $t(25) = 5.994$, $p < 0.001$, Cohen's $d = 2.309$, BF > 1000).

The same comparisons between good and bad performers have been made for outer objects for the three experiments. In 60 SINGLE BUMPED, the difference between good and bad performers for outer objects was significant ($M_{good} = 0.444$, $SD = 0.192$; $M_{bad} = 0.212$, $SD = 0.207$; $t(33) = 3.436$, $p = 0.002$, Cohen's $d = 1.162$, BF = 21.058). However, the difference was only approaching significance in 30 SINGLE BUMPED ($M_{good} = 0.482$, $SD = 0.246$; $M_{bad} = 0.301$, $SD = 0.289$; $t(31) = 1.943$, $p = 0.061$, Cohen's $d = 0.667$, BF = 1.364). In 90 SINGLE BUMPED the difference reached significance again ($M_{good} = 0.441$, $SD = 0.166$; $M_{bad} = 0.163$, $SD = 0.155$; $t(25) = 4.498$, $p < 0.001$, Cohen's $d = 1.733$, BF = 158).

Further analyses regarding Experiment 4/a, b, and c have been carried out to investigate whether the learning strategies and quality (episodic, semantic) are different for subjects that are efficient overall performers (considering orientation) and for participants who are not that efficient overall orientation performers. The same division has been done as before. Good performers were roughly top 50% and bad performers were bottom 50% in terms of overall orientation performance. After that, the drawn orientation angles of the two groups of participants were analyzed. In particular, the orientation angles given as a response for the outer objects at low subjective certainty. Low certainty was determined for each subject. With each subject, their mean level of certainty for presented OUTER objects was derived. Then responses that were below the mean certainty were classified as low certainty answers. Ultimately, the drawn orientation angles for outer objects at low certainty were analyzed for both good and bad performers. The aim of this was to analyze the form of the representations of good and bad performers and show (what appeared from previous results too) that better overall performers are first and foremost more efficient semantic learners. In other words, they learn a prior (such as the structure of the input Gaussian) from the input more efficiently than bad overall performers. The logic was to show that at low certainty responses when it is most probable that participants do not have vivid episodic recollection of the stimulus, consequently, they must rely on their prior knowledge to give a reasonable orientation estimate, the response orientation distribution of good performers are significantly different from the distribution of bad performers. What is more, the distribution of good performers is closer to the prior (input Gaussian) both regarding mean and standard deviation. The reason why only outer objects are analyzed is to remove the potential confound which is that the responses of good performers are, on average, by definition, closer to the mean of the input Gaussian distribution, and it is especially true for the middle objects. The actual analysis was done by bootstrap resampling (1000 times) the response orientation angle distribution (only outer objects) of both good and bad performers. Each resampling gave a mean and a standard deviation value. Eventually, the 1000 samples of good and bad performers were compared with independent sample t-tests.

In Experiment 4/a (60 SINGLE BUMPED), regarding the bootstrap means, the orientation angles of good performers were highly significantly different than bad performers, with good performers closer to the prior mean (Bootstrap_mean_good_performers = 25.227, SD = 3.736; Bootstrap_mean_bad_performers = 17.608, SD = 3.926; $t(1998) = 44.748$, $p < 0.001$, Cohen's $d = 2.001$, BF > 100 000). As for the bootstrap standard deviation, the drawn orientation angles of good and bad performers were highly significantly different with good performers closer to the prior standard deviation (Bootstrap_mean_good_performers =

43.591, SD = 2.782; Bootstrap_mean_bad_performers = 53.557, SD = 2.195; t(1998) = 88.925, p < 0.001, Cohen's d = 3.977, BF = Infinite). Interestingly, what participants claimed in the after-experiment questionnaire, did not correspond to whether they were good or bad overall performers either because among good performers 35% were implicit/no learners and 65% explicit learners, and among bad performers 33% were implicit/no learners and 67% explicit learners.

In Experiment 4/b (30 SINGLE BUMPED) regarding the bootstrap means, the orientation angles of good performers were highly significantly different from bad performers, and good performers were closer to the prior mean (Bootstrap_mean_good_performers = 12.066, SD = 6.196; Bootstrap_mean_bad_performers = 2.584, SD = 4.896; t(1998) = 37.966, p < 0.001, Cohen's d = 1.698, BF > 100 000). Interestingly, regarding the bootstrap standard deviation, bad performers were closer to the standard deviation of the prior distribution and the responses of good and bad performers were significantly different (Bootstrap_mean_good_performers = 50.962, SD = 2.779; Bootstrap_mean_bad_performers = 39.8, SD = 2.797; t(1998) = 89.526, p < 0.001, Cohen's d = 4.004, BF = Infinite).

In Experiment 4/c (90 SINGLE BUMPED), regarding the bootstrap means, there was a very mild difference between good and bad performers, with bad performers closer to the prior mean (Bootstrap_mean_good_performers = 20.364, SD = 3.426; Bootstrap_mean_bad_performers = 20.731, SD = 3.481; t(1998) = 2.377, p = 0.018, Cohen's d = 0.106, BF = 0.826). In terms of bootstrap standard deviation, there was a little but significant difference between good and bad performers, with good performers closer to the prior standard deviation (Bootstrap_mean_good_performers = 46.015, SD = 2.091;

Bootstrap_mean_bad_performers = 46.725, SD = 2.237; $t(1998) = 7.328$, $p < 0.001$, Cohen's $d = 0.328$, BF > 100).

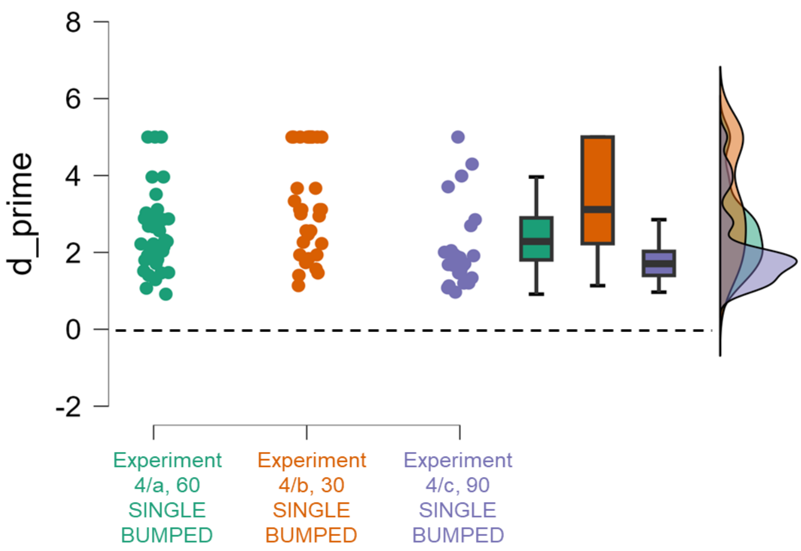

Figure 4.3 Object identity performance in Experiment 4/a, b, c. The dashed line indicates chance level performance. Although performance was significantly above chance throughout, modulating set size effected even identity recognition.

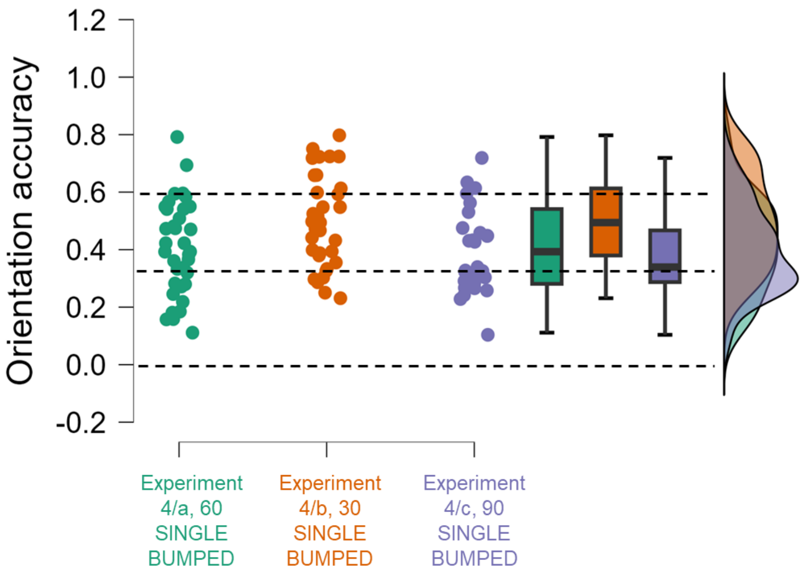

Figure 4.4 Overall orientation recall performance in Experiment 4/a, b, c. The three dashed lines indicate the episodic (0) and the two semantic baselines (~0.35, ~0.6). Overall orientation recall performance decreased with an increase in set size and stayed between the lazy and full semantic baselines throughout the experiments.

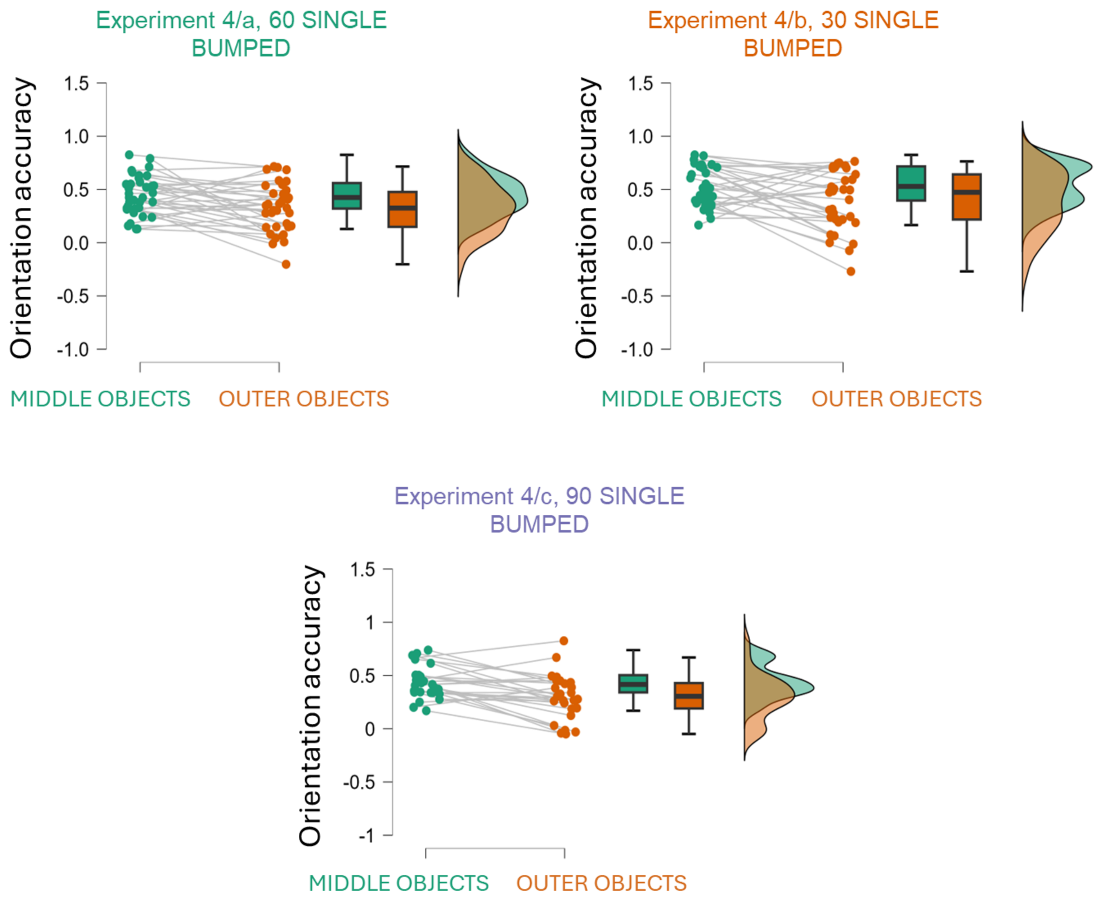

Figure 4.5 Orientation recall performance for middle and outer objects in Experiment 4/a, b, c. In all three experiments, middle objects were recalled with higher accuracy than outer objects, which is one indication of learning the input regularity.

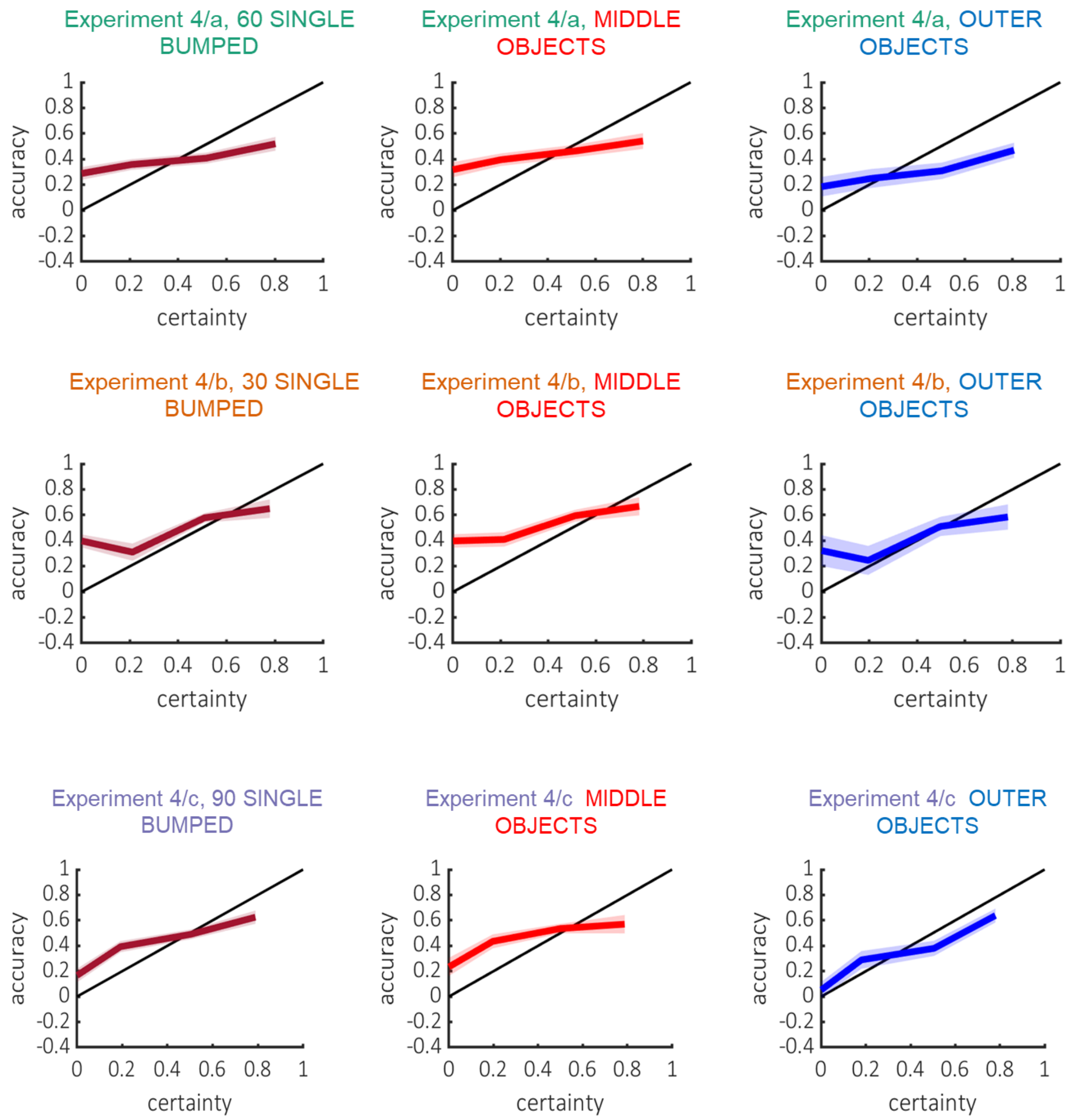

Figure 4.6 Calibratedness of orientation responses in Experiment 4/a, b, c. The first column shows all responses (for presented objects) from the experiments. The second and third column separates the responses for middle and outer objects.

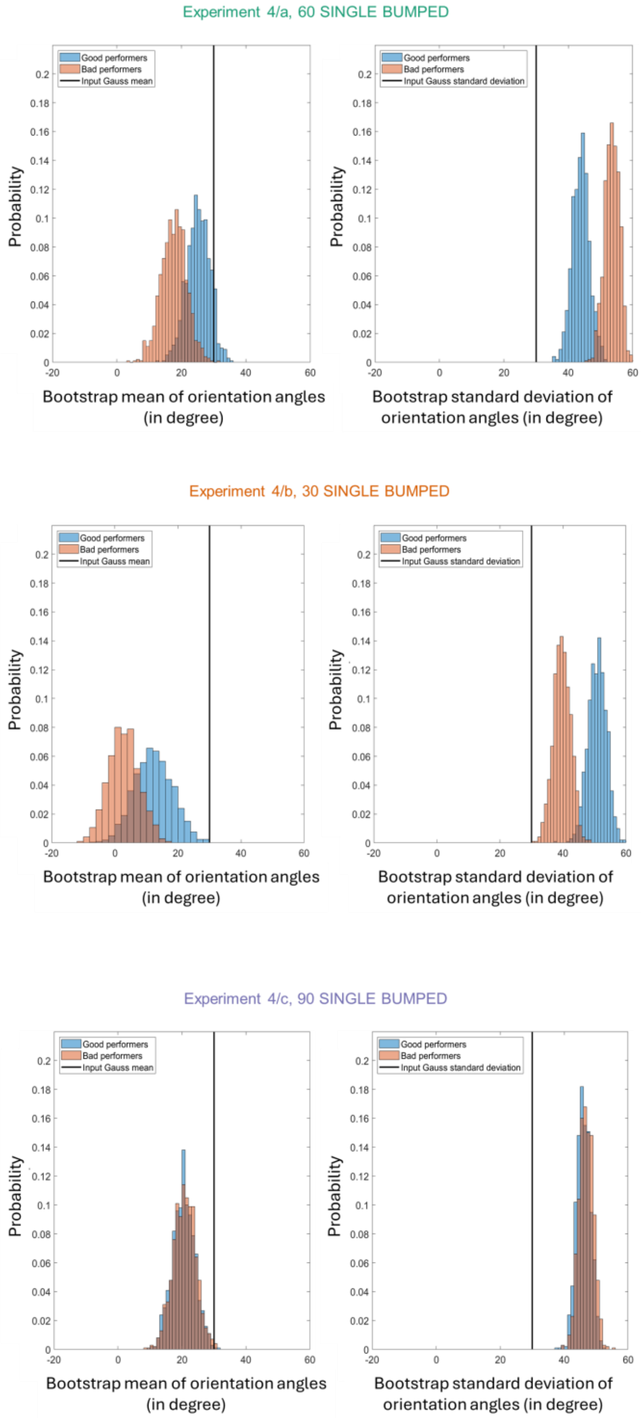

Figure 4.7 Response orientation angles at low certainty levels compared to the input Gaussian distribution (prior) in Experiment 4/a, b, c. Based on overall orientation recall performance, subjects were divided into top 50% (good performers) and bottom 50% (bad performers). Then the distribution of the response orientation angles for outer objects at low certainty levels was resampled 1000 times for both groups. Then, the means and standard deviations of the resampled distributions were compared against each other and in relation to the prior mean and standard deviation (input Gaussian with mean and standard deviation = 30 degrees).

4.3 Summary

Chapter 4 investigated the effect of overarching statistical, semantic regularities on recall precision and the representation of uncertainty. Results from experiments 4/a, b, and c suggest that participants learnt some of the summary statistics of the input and utilized it when giving orientation responses for individual items. However, there were major individual differences. It seems that a significant portion of the participants only partly or did not at all learn the input regularity. Results show that participants who performed better overall are first and foremost good semantic learners, meaning that they learn the regularity of the input more efficiently than bad performers. Interestingly, the results also show that despite the semantic regularity, episodic learning is not abandoned, what is more, better overall performers are better episodic learners as well. It is interesting because it is theoretically possible to generate all responses from the learnt regularity (e.g., Nagy et al., 2020) without storing anything episodically, and in this experiment, it would have led to satisfactory task performance. In addition, it seems that the category (implicit/no learner, explicit learner) that participants fall into based on the after-experiment questionnaire (Experiment 4/a), does not predict how well they perform in the actual memory test or whether they are efficient semantic learners. Lastly, considering the representation of uncertainty, results show that responses remained calibrated (except for outer objects in Experiment 4/b) and item based. However, calibratedness patterns (both quantitatively and qualitatively) indicate a mixture of episode and semantic features, which are not fully integrated in the memory representation.

CHAPTER 5

THE EFFECT OF NOISE AND ATTENTION ON THE REPRESENTATION OF UNCERTAINTY AND RECALL EFFICIENCY IN EPISODIC AND SEMANTIC MEMORY

5.1 Introduction

Previous experiments in Chapter 4 investigated episodic and semantic learning in a relatively simple way, namely, that at a given time one object was presented with a blank background throughout the experiment. I demonstrated that participants learn semantically (the statistics of the input regularity), which is especially apparent with efficient overall performers. At the same time, episodic memories were not redundant, as efficient learners also remembered individual items better. However, Chapter 4 did not examine if and how more external factors such as more complex stimuli or varying levels of attention modify people's representations.

It is known from previous studies that attention plays a significant role in how strongly memory items are encoded (Ramey et al., 2020). In addition, attention can also be investigated along with other topics such as episodic and semantic memory or gist-based and individual, more detailed representations (Allred et al., 2016; Bates et al., 2019; Greene & Naveh-Benjamin, 2022). In fact, from previous studies it seems that attention, working memory, semantic and episodic memory form a complex intertwined system (e.g.: Henderson et al., 1999; Bornstein & Norman, 2017; Theeuwes et al., 2022). Primarily attention decides what gets encoded in working memory as an episode, however, the content and capacity of working memory is also influenced by previously learnt semantic schemas (Brady et al., 2009; Bates et al., 2019), as a consequence, long-term memory representations are also a product of attention, working memory, semanticity and episodic representations. What is more, previous research has established that with divided attention, not only episodic memory but gist-based, semantic memory also becomes impaired (Greene & Naveh-Benjamin, 2022). The current chapter preliminarily investigates the effects of attention on encoding strategies in terms of episodic and semantic memories, namely, which kind of learning (episodic vs. semantic) is prioritized when attention is compromised. In addition, experiments in Chapter 5 examine other encoding contexts that can modify people's attention and hence long-term memory representations. Besides, Chapter 5 examines the possible changes that varying levels of attention can have on the representation of uncertainty in long-term memory.

Experiment 5/a uses the same overarching semantic regularity as experiments in Chapter 4 but also embeds the presented objects in a natural background. It shows that most probably due to the backgrounds, the subject's representation becomes slightly noisier, but the overarching regularity is learnt with the same precision as before. Experiment 5/b with direct divided attention demonstrates that under this encoding condition semantic learning remains intact but individual episodes are forgotten to a larger extent relative to semantic representations. Lastly, in Experiment 5/c participants implicitly needed to encode more than one object dimension compared to previous experiments, where it was enough to encode the object category and orientation to perform well in the task. As a consequence, in Experiment 5/c, attention levels at encoding were indirectly increased and directed to more than one object dimension. Experiment 5/c shows that despite supposedly paying attention to several object dimensions, people can automatically learn the semantic regularity for one particular dimension (here: orientation). As for the representation of uncertainty, results in Chapter 5 are reminiscent of Chapter 4 and also of previous chapters, in the dissertation. Namely, objects are represented along with their item-based uncertainty which becomes biased due to learning the input regularity. All this suggests that episodic memories with semantic regularity are represented in a probabilistic manner which external factors such as noise or varying levels of attention can further modify but not fundamentally change.

5.2.1 Experiment 5/a: The Effect of Background Embedding on Episodic and Semantic Learning

Experiment 5/a investigates whether people can still sufficiently extract the overarching semantic input regularity when there is more noise and complexity in the input. Experiment 5/a examines this question by keeping the input regularity from Chapter 4 (Gaussian distribution for object orientation) but also embedding the oriented objects in natural backgrounds.

Participants

Twenty-seven people participated in Experiment 5/a (60 SINGLE BUMPED BACKGROUND) They were recruited through the Hungarian MADS student organization and were mostly Hungarian university students. They received hourly compensation of 3000 HUF (~7.5$) for their participation. The experiment lasted for about an hour. The study was conducted in person, in the visionlab of CEU in Budapest. Every participant had normal or corrected to normal vision. The study was approved by the Hungarian United

Ethical Review Committee for Research in Psychology (EPKEB), and all participants provided informed consent.

Materials and Procedure

The design of Experiment 5/a was similar to experiments presented in previous chapters. It consisted of three practice sessions and one test session. The first two practice sessions were the usual line drawing and orientation practice sessions. First, in 34 trials, participants learnt the function between line length and wedge width on the tablet. Then, in 50 trials, in a perceptual decision-making task, they reported the orientation of Gabor patches along with their subjective certainty.

The long-term memory part was very similar to Experiment 4/a, the base structured experiment of the dissertation. It consisted of a practice and a test session, both including a study and a recall phase. In the study phase, participants saw 60 (5 during practice) individually presented, oriented objects that were embedded in a background. For this experiment, the same set of 190 objects were used as before along with 90 natural backgrounds. The orientation of the objects was sampled from a Gaussian distribution with mean = 30 degrees, and standard deviation = 30 degrees. The backgrounds were predominantly natural scenes of various kinds, such as city scenes, buildings, landscapes. The object and the background were not selected to be semantically related. Although the selection of objects and backgrounds for every scene was random, accidental semantic connections might have occurred. The objects and the backgrounds were displayed for 1 second with a 2 second gap between the trials. During the 2 second gap, a grey circle/mask appeared in the middle of the screen with a fixation dot in the middle, to keep participants' attention and gaze from wandering, just like in previous experiments.

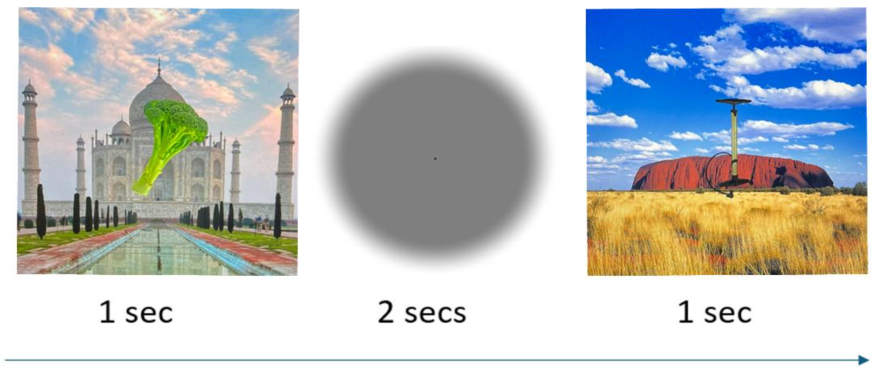

Figure 5.1 Study phase of Experiment 5/a. The design of the experiment is identical to Experiment 4/a, except objects are embedded in a natural background.

In the recall phase, participants gave responses to 120 objects (10 during practice), which meant that the experiment was done in the usual 50% old, 50% new design. The first response was whether they had seen the given object in the study phase. This response was given on the tablet, using a horizontal scale and a slider, exactly as before. Then, as a second response, participants had to recall the orientation of the objects from the study phase. This response was also identical to the design in previous chapters. In each trial, participants drew a line on the tablet, which was their orientation estimate, and the length of the drawn line corresponded to their subjective certainty.

Results

Regarding object identity recognition performance, there was no significant difference between Experiment 5/a (60 SINGLE BUMPED BACKGROUND) and 4/a (60 SINGLE BUMPED) (M_exp4/a = 2.519, SD = 1.067; M_exp5/a = 2.091, SD = 0.914; $t(58) = 1.625$, $p = 0.110$, Cohen's d = 0.425, BF = 0.794). The subjective confidence ratings for object identity were high across participants in Experiment 5/a (mean = 0.87, SD = 0.09).

As for object orientation performance, there was no significant decrease compared to Experiment 4/a (M_exp4/a = 0.407, SD = 0.164; M_exp5/a = 0.365, SD = 0.175; $t(58) = 0.956$, $p = 0.343$, Cohen's d = 0.250, BF = 0.389). Further, orientation recall performance was not significantly above the full semantic baseline (~0.35) ($t(24) = 0.417$, $p = 0.680$, Cohen's d = 0.083, BF = 0.228). And it was significantly below the lazy semantic baseline (~0.6) ($t(24) = 6.729$, $p < 0.001$, Cohen's d = 1.346, BF > 1000) in Experiment 5/a. Further, in Experiment 5/a, there was a significant difference in orientation accuracy between middle and outer objects (M_middle = 0.417, SD = 0.215; M_outer = 0.246, SD = 0.221; $t(24) = 3.067$, $p = 0.005$, Cohen's d = 0.613, BF = 8.203).

As for calibratedness, the correlation between orientation accuracy and subjective certainty across subjects was significant in Experiment 5/a (mean_pearson_r = 0.103, SD = 0.199; $t(24) = 2.586$, $p = 0.016$, Cohen's d = 0.517, BF = 3.194). The difference in correlation between middle and outer objects was not significant (M_middle = 0.083, SD = 0.244; M_outer = 0.127, SD = 0.319; $t(24) = 0.553$, $p = 0.585$, Cohen's d = 0.111, BF = 0.242). Experiment 5/a was the second example in the dissertation, where there was no significant calibratedness, when the responses were separated to middle and outer, although in the case of outer objects it approached significance. Middle: $t(24) = 1.708$, $p = 0.1$, Cohen's d = 0.342, BF = 0.748. Outer: $t(24) = 1.990$, $p = 0.058$, Cohen's d = 0.398, BF = 1.139.

Two subjects were excluded from the analyses from Experiment 5/a for below chance level orientation accuracy.

Discussion

Considering general trends, Experiment 4/a and Experiment 5/a echoed each other. There has been a difference between middle and outer recall performance in Experiment 5/a, just like in Experiment 4/a, although the magnitude of the difference was slightly smaller in the latter. In Experiment 5/a, the difference in correlation between middle and outer objects was not significant, but outer objects were better calibrated than middle objects, similar to Experiment 4/a. Importantly, in Experiment 5/a, when calibratedness was analyzed separately for middle and outer objects, neither of them reached significance, which was especially apparent with middle objects where the calibration curve was almost flat. This suggests that when objects were shown in a background, participants' general learning strategy did not change, they still learned the input regularity to some extent along with individual episodes. At the same time, subjects' representation generally seemed to become noisier due to the background, resulting in a mild overall decrease in performance. In addition, the noisier representation was more apparent in the representation of uncertainty, and the loss of significant calibratedness when middle and outer objects were analyzed separately. It seems that the presence of the background further disturbed participants uncertainty representation, even though it was possibly neglected during encoding because it was not task relevant. It shows that even though something (background) is not encoded from the input, the knowledge that it was present can slightly alter the overall structure of the representation.

5.2.2 Experiment 5/b: The Effect of Divided Attention on Episodic and Semantic Learning

Experiment 5/a implied that embedding objects in a background does not significantly affect people's ability to extract semantic input regularity. Experiment 5/b investigates participants' learning strategies and the connection between episodic and semantic memories further by introducing an extra task during the study phase of the long-term memory part of the experiment. It asks whether doing a supplementary task during encoding affects the amount and quality of episodic and semantic learning in any major way.

Participants

Twenty-seven people participated in Experiment 5/b (60 SINGLE BUMPED EXTRA TASK). They were recruited through the Hungarian MADS student organization and were mostly Hungarian university students. They received hourly compensation of 3000 HUF (~7.5$) for their participation. The experiment lasted for about an hour. The study was conducted in person, in the visionlab of CEU in Budapest. Every participant had normal or corrected to normal vision. The study was approved by the Hungarian United Ethical Review Committee for Research in Psychology (EPKEB), and all participants provided informed consent.

Materials and Procedure

The materials and procedure used in Experiment 5/b was almost identical to Experiment 4/a (60 SINGLE BUMPED). The difference in Experiment 5/b was that in the long-term memory part of the experiment, during the study phase, in each trial, either a red or a blue dot appeared above the presented objects. If the dot was red, participants were instructed to press the left arrow on the keyboard, if it was blue, the right arrow. The dots always appeared vertically above the objects. The object and the dot appeared on the screen for only 1 second as before, however, the next dot and object were presented only when the participant pressed either the left or right arrow on the keyboard. Until the keyboard press, the grey circle with the fixation dot was present on the screen. . Everything else, including the recall phase was identical to Experiment 4/a, and Experiment 5/a.

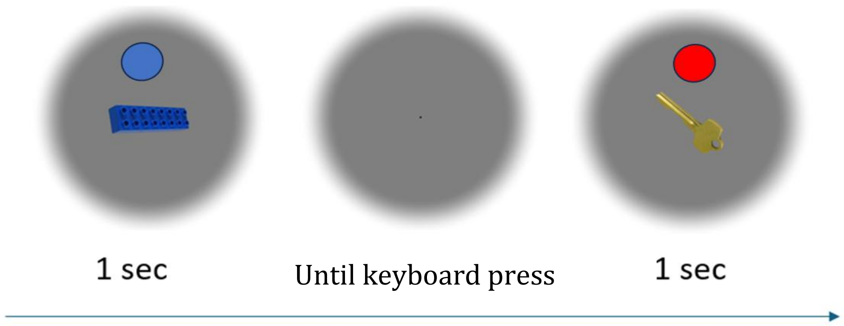

Figure 5.2 Study phase of Experiment 5/b. During the study phase, there is an extra task in Experiment 5/b. If the dot above the object is blue, they have to press the right arrow on the keyboard, if red, the left arrow.

Results

Considering object identity recognition performance, there was a highly significant difference between Experiment 4/a and Experiment 5/b (M_exp4/a = 2.519, SD = 1.067; M_exp5/b = 1.490, SD = 1.056; $t(59) = 3.743$, $p < 0.001$, Cohen's d = 0.969, BF = 65.1). Participants' subjective confidence ratings, however, remained high in Experiment 5/b for object identity (mean = 0.81, SD = 0.11). Further, the difference between Experiment 5/a and Experiment 5/b was mildly significant (M_exp5/a = 2.091, SD = 0.914; M_exp5/b = 1.490, SD = 1.056; $t(49) = 2.170$, $p = 0.035$, Cohen's d = 0.608, BF = 1.862) in identity performance.

In terms of orientation recall accuracy, the difference between Experiment 4/a and Experiment 5/b was approaching significance (M_exp4/a = 0.407, SD = 0.164; M_exp5/b

= 0.319, SD = 0.181; t(59) = 1.974, p = 0.053, Cohen's d = 0.511, BF = 1.317). Between Experiment 5/a and Experiment 5/b the difference was not significant in orientation accuracy (M_exp5/a = 0.365, SD = 0.175; M_exp5/b = 0.319, SD = 0.181; t(49) = 0.909, p = 0.368, Cohen's d = 0.255, BF = 0.394). Overall orientation accuracy in Experiment 5/b was not significantly different from the full semantic baseline (~0.35) (t(25) = 0.865, p = 0.395, Cohen's d = 0.170, BF = 0.291). In addition, it was significantly below the lazy semantic baseline (~0.6) (t(25) = 7.905, p < 0.001, Cohen's d = 1.550, BF > 1000). Considering middle and outer objects in Experiment 5/b, the difference was highly significant (M_middle = 0.408, SD = 0.205; M_outer = 0.127, SD = 0.221; t(25) = 6.592, p < 0.001, Cohen's d = 1.293, BF > 1000).

The correlation between orientation accuracy and subjective certainty was significant in Experiment 5/b across subjects (mean_pearson_r = 0.129, SD = 0.180; t(25) = 3.649, p = 0.001, Cohen's d = 0.716, BF = 29.045). Lastly, the difference in correlation between middle and outer objects was not significant (M_middle = 0.121, SD = 0.207; M_outer = 0.122, SD = 0.281; t(25) = 0.019, p = 0.985, Cohen's d = 0.004, BF = 0.207).

One subject was excluded from the analyses, from Experiment 5/b for below chance level orientation accuracy.

Discussion

Both object identity performance and orientation recall performance dropped further in Experiment 5/b compared to Experiment 4/a and Experiment 5/a. The drop in orientation recall was partly due to poor performance for outer objects which is also shown by the massive difference between middle and outer orientation recall performance in Experiment 5/b. This can indicate that episodic encoding was more impaired in Experiment 5/b, due to the extra task. This is only the case if recall performance for outer objects can be taken as episodic memory. However, there is a good reason for that, as in the case of outer objects, participants cannot rely on the summary statistics, or much less than in the case of middle objects. Other patterns were similar in Experiment 5/b to 4/a and 5/a. The calibratedness of middle objects was biased supposedly due to semantic learning, shown by the tilted calibration curve. Overall, the extra task during encoding seemed to have pushed participants' representation to be more semantic, which was based on the general mean and summary statistics of the input regularity.

5.2.3 Experiment 5/c: The Effect of Encoding Multiple Object Details on Semantic and Episodic Learning

All previous experiments so far made participants focus on object orientation more than other object details. Experiment 5/c investigates whether subjects can extract and utilize the input regularity of one object detail (here: orientation) when other object dimensions/details are also important for completing the task with higher accuracy. In Experiment 5/c participants must explicitly pay attention to several object dimensions apart from orientation, as upon recall they should distinguish objects that differ only in small details from the ones that they had seen during the study phase.

Participants

Forty people participated in Experiment 5/c (60 SINGLE BUMPED OBJECT DETAILS) They were recruited through the Hungarian MADS student organization and were mostly Hungarian university students. They received hourly compensation of 3000 HUF (~7.5$) for their participation. The experiment lasted for about an hour. The study was conducted in person, in the visionlab of CEU in Budapest. Every participant had normal or corrected to normal vision. The study was approved by the Hungarian United Ethical Review Committee for Research in Psychology (EPKEB), and all participants provided informed consent.

Materials and Procedure

The design of Experiment 5/c was almost identical to Experiment 4/a. The difference was that in Experiment 5/c, in the long-term memory part, in the recall phase, not all previously presented objects were tested. Of the original 60 objects that were presented in the study phase, 30 were tested in the recall phase, and an additional 30 new objects were added and tested. The new items differed in small details from the old ones that were not tested, from the study phase. As a consequence, the recall phase was done in the usual 50% old, 50% new design, however, participants gave responses to only 60 objects altogether. The objects for this experiment were selected from the same large pool of ~800 objects. But now similar pairs, for instance 2 trumpets that differed in minor details from each other, were specifically constructed (see figure 5.3).

It is inherent from the nature of the design that there is a fraction of the participants (called "structured" participants) for whom the orientation distribution of the tested old objects did follow the overall input Gaussian with mean and standard deviation = 30 degrees, and there are participants for whom the tested old objects deviate from the overall input Gaussian (called "non-structured" participants). It is because unlike in previous experiments, not all the presented objects are tested from the study phase, only 50% of them. Consequently, subjects and the analyses in Experiment 5/c are divided into the aforementioned two parts. For 13 subjects out of the 40, the distribution of the tested old items followed the input

Gaussian, and for 14 subjects it did not. Whenever the mean and the standard deviation of the tested old objects' orientation was within 5 degrees of the mean and standard deviation of the input Gauss (for instance, mean = 27 degrees and standard deviation = 28 degrees for the tested old items) then that particular subject was classified as a structured participant. When the mean orientation of the tested old objects was more than 30 degrees away from the mean of the input Gaussian then the given subject was classified as a non-structured participant. This also meant that the orientation distribution of the tested old items in the case of non-structured participants was close to random. In the case of 10 subjects, a clear division was not possible to either the structured or the non-structured group, consequently, those subjects were not included in the majority of the analyses. An additional 3 subjects were excluded from the analysis for below chance orientation accuracy in Experiment 5/c.

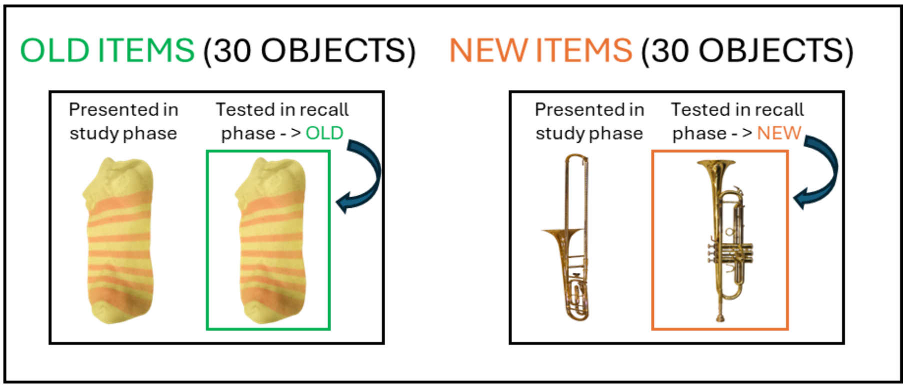

Figure 5.3 Recall phase of Experiment 5/c. In the recall phase of Experiment 5/c, 30 out of the presented 60 objects are tested exactly as they were presented in the study phase. The other 30 objects are technically new items, however, they only differ in minor details from the ones (remaining 30 objects) presented in the study phase.

Results

Regarding object identity recognition performance, all subjects from Experiment 5/c could be included in the analysis, as for the object identity dimension, there was no structural difference between the two kinds of participants. In object identity recognition, the difference between Experiment 4/a and Experiment 5/c was highly significant (M_exp4/a = 2.519, SD = 1.067; M_exp5/c = 1.264, SD = 0.822; t(70) = 5.610, $p < 0.001$, Cohen's d = 1.323, BF > 1000). Nevertheless, the corresponding subjective confidence ratings were high across subjects in Experiment 5/c with mean = 0.87, SD = 0.11. The difference in identity recognition was significant between Experiment 5/a and Experiment 5/c as well (M_exp5/a = 2.091, SD = 0.914; M_exp5/c = 1.264, SD = 0.822; t(60) = 3.714, $p < 0.001$, Cohen's d = 0.962, BF = 60.7). Lastly, there was no significant difference between Experiment 5/b and Experiment 5/c in object identity recognition (M_exp5/b = 1.490, SD = 1.056; M_exp5/c = 1.264, SD = 0.822; t(61) = 0.955, p = 0.343, Cohen's d = 0.244, BF = 0.382). Regarding the difference between structured and non-structured participants from Experiment 5/c, the difference between the two groups approached significance (M_exp5/c_structured = 0.928, SD = 0.557; M_exp5/c_non_structured = 1.323, SD = 0.534; t(25) = 1.883, p = 0.071, Cohen's d = 0.725, BF = 1.280).

In terms of orientation recall accuracy it is more sensible to compare structured participants to Experiment 4/a (60 SINGLE BUMPED) and non-structured participants to Experiment 2/a (60 SINGLE RANDOM). Between Experiment 4/a and the structured participants from Experiment 5/c, there was no significant difference (M_exp4/a = 0.407, SD = 0.164; M_exp5/c_structured = 0.398, SD = 0.213; t(46) = 0.158, p = 0.875, Cohen's d = 0.051, BF = 0.318). As for the semantic baselines, orientation accuracy of structured participants from Experiment 5/c was not significantly different from the full semantic baseline (~0.35)

($t(12) = 0.805$, $p = 0.436$, Cohen's $d = 0.223$, $BF = 0.367$), and it was significantly below the lazy semantic baseline (~0.6) ($t(12) = 3.422$, $p = 0.005$, Cohen's $d = 0.949$, $BF = 10.136$). Comparing orientation accuracy for middle and outer objects, there was significant difference in the case of structured participants from Experiment 5/c ($M_{middle} = 0.437$, $SD = 0.207$; $M_{outer} = 0.308$, $SD = 0.272$; $t(13) = 2.450$, $p = 0.031$, Cohen's $d = 0.680$, $BF = 2.364$). Further, the difference in orientation recall accuracy between Experiment 2/a and the non-structured participants from Experiment 5/c was approaching significance ($M_{exp2/a} = 0.290$, $SD = 0.168$; $M_{exp5/c\_non\_structured} = 0.399$, $SD = 0.181$; $t(38) = 1.908$, $p = 0.064$, Cohen's $d = 0.633$, $BF = 1.290$). Lastly, there was no significant difference in orientation recall accuracy between structured and non-structured participants from Experiment 5/c ($M_{exp5/c\_structured} = 0.398$, $SD = 0.213$; $M_{exp5/c\_non\_structured} = 0.399$, $SD = 0.181$; $t(25) = 0.018$, $p = 0.985$, Cohen's $d = 0.007$, $BF = 0.358$).

The correlation between orientation accuracy and subjective certainty was significant in Experiment 5/c, considering all participants ($mean\_pearson\_r = 0.212$, $SD = 0.213$; $t(36) = 6.049$, $p < 0.001$, Cohen's $d = 0.994$, $BF > 1000$). The correlation was significant both in the case of structured ($mean\_pearson\_r = 0.199$, $SD = 0.171$; $t(12) = 4.2$, $p = 0.001$, Cohen's $d = 1.165$, $BF = 33.2$) and non-structured ($mean\_pearson\_r = 0.222$, $SD = 0.262$; $t(13) = 3.170$, $p = 0.007$, Cohen's $d = 0.847$, $BF = 7.2$) participants separately. Considering only structured participants, there was no significant difference between middle and outer objects in terms of correlation ($M_{middle} = 0.149$, $SD = 0.151$; $M_{outer} = 0.311$, $SD = 0.313$; $t(12) = 1.920$, $p = 0.079$, Cohen's $d = 0.532$, $BF = 1.143$). Lastly, there was no significant difference in correlation between the structured and the non-structured participants, from Experiment 5/c ($M_{exp5/c\_structured} = 0.199$, $SD = 0.171$;

M_exp5/c_non_structured = 0.222, SD = 0.262; t(25) = 0.271, p = 0.789, Cohen's d = 0.104, BF = 0.368).

Discussion

In Experiment 5/c there has been a significant drop in object identity recognition even compared to Experiment 5/a. It is not surprising in the sense that the object recognition task became more difficult, as now small details had to be distinguished. On the other hand, participants knew about this feature in the experiment. In that sense, object recognition performance is surprisingly low. One further surprising result from Experiment 5/c is that object orientation recall performance was not significantly different from Experiment 4/a (neither in structured, nor in non-structured participants) despite the fact that presumably subjects paid more attention to other object features apart from orientation. What is more, the orientation performance of non-structured participants (who could only learn the tested old items episodically, in an item-based manner) was almost significantly higher than subjects' performance in Experiment 2/a. This pattern in learning is especially interesting in the light of the poor object identity performance. Besides, structured participants showed a pattern which was similar to previous experiments, with overarching semantic regularity, there was a significant difference in middle and outer orientation recall performance.

Further results, analyses, Experiment 5/a, b, c altogether

The following set of analyses, just like in Chapter 4, targeted the amount of semantic/input structure and episodic learning that good vs. bad performers had. The division of participants into good and bad performers was done by the same method as before, top 50%, in overall orientation performance, were classified as good performers and bottom 50% as bad performers. First, in each experiment the general trend between good and bad performers, regarding middle and outer objects were analyzed. Then, with the same method as in Chapter 4, the drawn orientation responses for the outer objects at low certainty of each group were resampled 1000 times and the means and standard deviations of the resampled distributions were compared against each other and in terms of which one is closer to the target distribution (input Gaussian) regarding the mean and standard deviation.

The logic of the analysis was similar yet different in the case of Experiment 5/c (60 SINGLE BUMPED OBJECT DETAILS). In that experiment the orientation of the objects that were presented in the study phase and then later tested in the recall phase did not follow the shape of a Gaussian distribution with mean = 30 degrees and standard deviation = 30 degrees *for every subject*. This was not intentional in the design, simply the result that old items that were tested in the recall phase did not always get orientations that ultimately resembled a Gaussian with the above parameters. Therefore, this analysis used the exact same division, which was described before, into structured and non-structured participants. This inherent "design flaw" of Experiment 5/c gave an opportunity to test semantic learning from another angle. Namely, the drawn angles of structured participants should be significantly closer to the input Gauss than non-structured participants, if there is semantic learning.

First, regarding the difference in middle object accuracy between good and bad performers in Experiment 5/a (60 SINGLE BUMPED BACKGROUND), there was a highly significant difference between good and bad performers ($M_{good} = 0.561$, $SD = 0.162$; $M_{bad} = 0.260$, $SD = 0.144$; $t(23) = 4.896$, $p < 0.001$, Cohen's $d = 1.960$, $BF > 100$). In Experiment 5/b (60 SINGLE BUMPED EXTRA TASK) the difference in middle object accuracy between good and bad performers was highly significant ($M_{good} = 0.570$, $SD =$

0.104; M_bad = 0.246, SD = 0.142; t(24) = 6.617, p < 0.001, Cohen's d = 2.596, BF > 1000).

Further, considering the difference in outer object accuracy, in Experiment 5/a, there was a significant difference between good and bad performers (M_good = 0.343, SD = 0.188; M_bad = 0.142, SD = 0.213; t(23) = 2.518, p = 0.019, Cohen's d = 1.008, BF = 3.197). In Experiment 5/b, similarly, there was a significant difference between good and bad performers considering outer objects (M_good = 0.237, SD = 0.211; M_bad = 0.016, SD = 0.175; t(24) = 2.908, p = 0.008, Cohen's d = 1.141, BF = 6.235). In Experiment 5/c (60 SINGLE BUMPED OBJECT DETAILS), these comparisons would not be meaningful, as there was no middle and outer in the case of non-structured participants.

In Experiment 5/a (60 SINGLE BUMPED BACKGROUND), considering the bootstrap resampled means, the drawn responses of good performers were highly significantly closer to the mean of the prior distribution (Bootstrap_mean_good_performers = 26.035, SD = 3.913; Bootstrap_mean_bad_performers = 19.092, SD = 4.153; t(1998) = 38.481, p < 0.001, Cohen's d = 1.721, BF > 100 000). In terms of the bootstrap resampled standard deviations, again good performers were highly significantly closer to the prior standard deviation (Bootstrap_mean_good_performers = 43.262, SD = 2.903; Bootstrap_mean_bad_performers = 46.322, SD = 2.587; t(1998) = 24.889, p < 0.001, Cohen's d = 1.113, BF > 100 000).

In Experiment 5/b (60 SINGLE BUMPED EXTRA TASK), regarding the bootstrap means, good performers were highly significantly closer to the prior mean than bad performers (Bootstrap_mean_good_performers = 24.835, SD = 2.849; Bootstrap_mean_bad_performers = 17.716, SD = 4.395; t(1998) = 42.985, p < 0.001, Cohen's d = 1.922, BF > 100 000). As for the bootstrap standard deviation, good performers were likewise significantly closer to the prior standard deviation than bad performers (Bootstrap_mean_good_performers = 35.191, SD = 2.334; Bootstrap_mean_bad_performers = 48.925, SD = 2.248; $t(1998) = 134.022$, $p < 0.001$, Cohen's d = 5.994, BF = Infinite).

In Experiment 5/c (60 SINGLE OBJECT DETAILS) very similar patterns to Experiment 5/a and 5/b can be seen from the data. In this analysis all the drawn orientation responses of both structured and non-structured participants for presented objects were compared. First, considering the bootstrap means, structured participants were highly significantly closer to the prior mean than non-structured participants (Bootstrap_mean_structured_participants = 27.412, SD = 3.243; Bootstrap_mean_non_structured_participants = -1.241, SD = 3.717; $t(1998) = 183.677$, $p < 0.001$, Cohen's d = 8.214, BF = Infinite). Considering the bootstrap standard deviation, structured participants were highly significantly closer to the target distribution than non-structured participants (Bootstrap_mean_structured_participants = 44.22, SD = 2.439; Bootstrap_mean_non_structured_participants = 49.376, SD = 1.826; $t(1998) = 53.526$, $p < 0.001$, Cohen's d = 2.394, BF = Infinite).

Another set of analysis compared the drawn low certainty orientation responses for only the outer objects in the case of structured participants to the same set of responses as before in the case of non-structured participants (low certainty drawn orientation responses for all presented objects). The results were very similar to before. The bootstrap mean was highly significantly different between structured participants and non-structured participants (Bootstrap_mean_structured_participants = 25.861, SD = 6.216; Bootstrap_mean_non_structured_participants = -1.083, SD = 3.758; $t(1998) = 117.310$, p

< 0.001, Cohen’s $d = 5.246$, BF = Infinite). Regarding the bootstrap standard deviation, again structured participants were highly significantly closer to the prior standard deviation than non-structured participants (Bootstrap_mean_structured_participants = 47.290, SD = 3.745; Bootstrap_mean_non_structured_participants = 49.376, SD = 1.819; $t(1998) = 15.841$, $p < 0.001$, Cohen’s $d = 0.708$, BF $> 100\,000$).

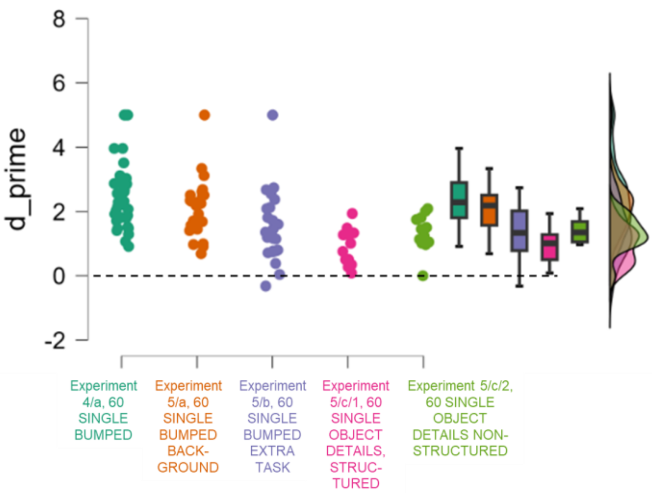

Figure 5.4 Object identity recognition performance in Experiment 4/a, 5/a, b, c. There has been a gradual decrease in performance as noise, divided attention, and more object details were introduced into the input.

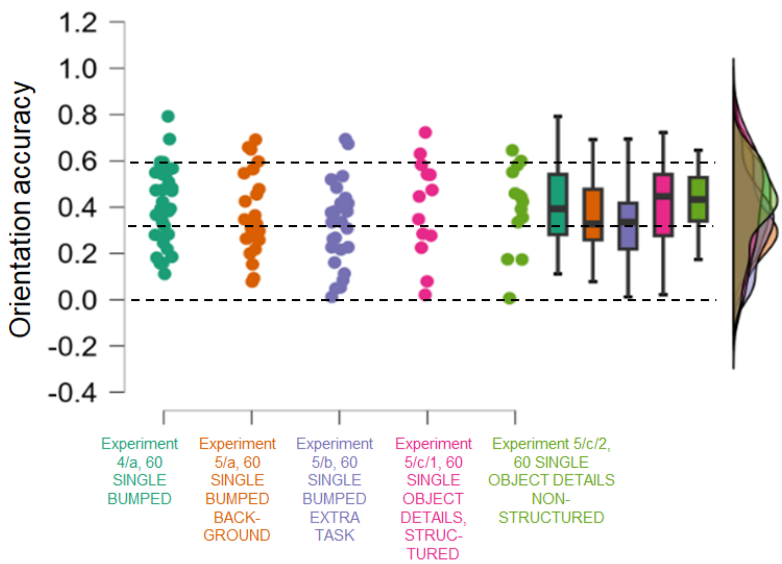

Figure 5.5 Orientation recall performance in Experiment 4/a, 5/a, b, c. In both Experiment 5/a and 5/b there has been a decrease in performance, compared to the base experiment (Exp. 4), however, in Experiment 5/c subjects (both structured and non-structured) performed as well as in Exp. 4. The black dashed lines indicate episodic chance level (0), the full semantic baseline (~0.35), and the lazy semantic baseline (~0.6).

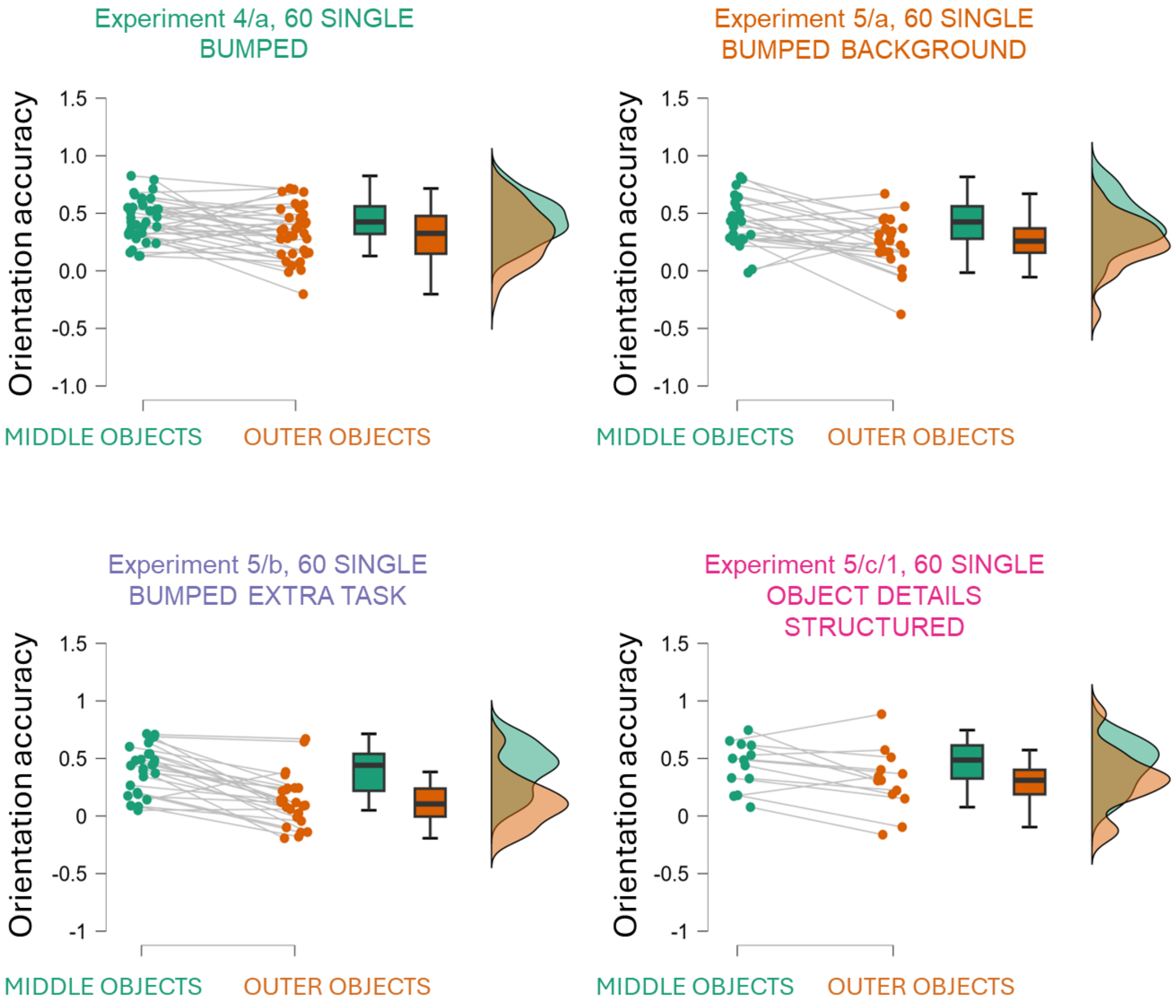

Figure 5.6 Orientation recall performance for middle and outer objects in Experiment 4/a, 5/a, b, c/1. In Experiment 5/a and 5/b, and in the case of structured participants, from Experiment 5/c, middle objects were recalled with significantly higher accuracy than outer objects (just like in Experiment 4/a), which is one indication of learning the input regularity.

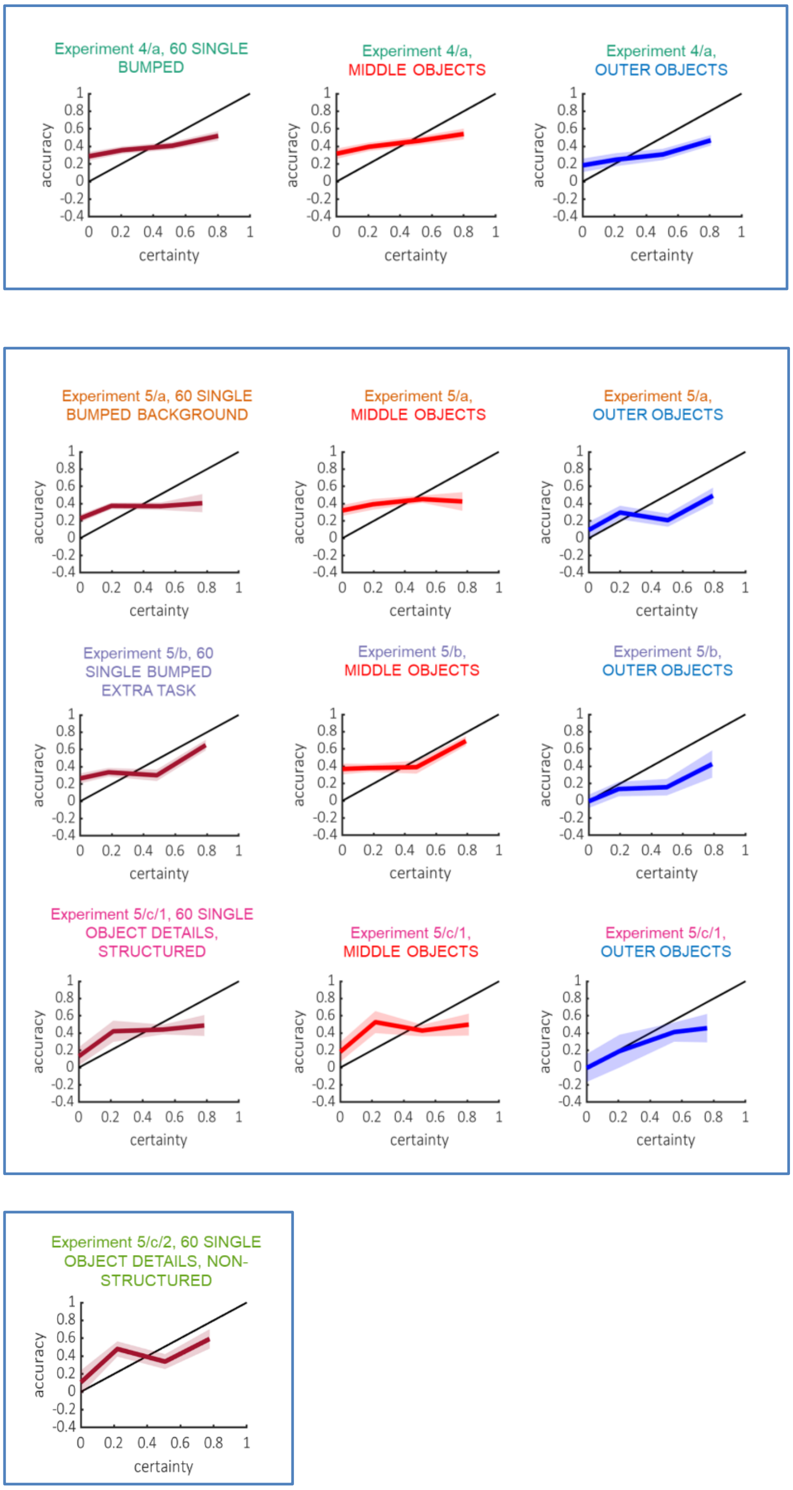

Figure 5.7 Calibratedness in Experiment 4/a, 5/a, b, c. In the case of middle objects in 4/a, 5/a, 5/b, the semantic bias is apparent in the shape of the calibration curve. Even at the lowest certainty level, orientation accuracy is above chance. This pattern was slightly different in Experiment 5/c where the calibration curves echoed the ones that were previously seen in pure episodic experiments (especially in 5/c/2). In the case of outer objects, in every experiment, responses were better calibrated, the shape of the curves echoed the shape of the ones in the pure episodic memory experiments.

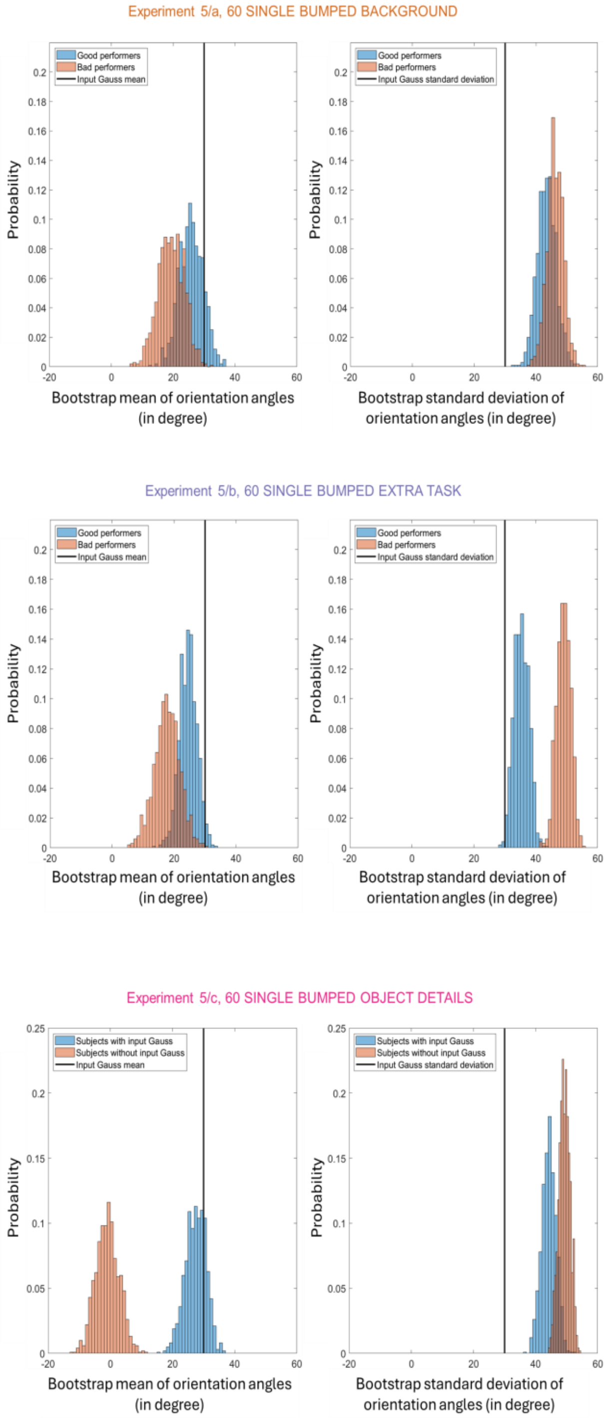

Figure 5.8 Drawn orientation angles compared to the input Gaussian distribution (prior), at low certainty levels, in Experiment 5/a, b, c. Just like in Chapter 4, in experiments 5/a and 5/b, subjects were divided into top 50% (good performers) and bottom 50% (bad performers) based on overall orientation performance. Then their drawn orientation responses were analyzed with the same method as in Chapter 4. In Experiment 5/c, the same method was used as well except subjects were not grouped based on performance but based on whether their input was structured or not considering the orientation of the tested old objects.

5.3 Summary

Besides the overarching semantic regularity that was introduced in Chapter 4, Chapter 5 embedded objects in a natural background, divided attention at encoding and increased attention to more than one object detail. Experiments 5/a and 5/b demonstrated that when the input is noisier or when attention is divided, correspondingly semantic (5/a) or episodic (5/b) recall performance will be slightly more impaired, nevertheless, the quality of semantic learning of good performers tended to remain intact, in both experiments. Surprisingly, results from Experiment 5/c suggested that when participants must encode several object dimensions, recall precision was slightly higher for the tested dimension as well (compared to the corresponding base experiments), although not significantly. The pattern in the representation of uncertainty remained very similar to Chapter 4, namely, item-based, biased by the learnt semantic regularity, but significantly calibrated to accuracy in almost all cases.

CHAPTER 6

GENERAL DISCUSSION

6.1 Massivity and the Probabilistic Nature of Representation in Pure Episodic Memories

In Chapter 2, I argued that the concept of massivity in connection with long-term visual memories is vague depending on the context of the experiment and the researchers' interpretation. Upon closer examination, visual long-term memory appears to be rather poor than massive when it comes to recalling details. While in Brady et al. (2008) study, participants saw 2500 images/objects altogether in one experiment, encoding and recall was divided into blocks of 300 images that were tested, with a 2AFC method. With this setup, participants performed at above 90% accuracy even when memories for small object details were probed. However, familiarity tested with a 2AFC method and free recall are distinct retrieval processes (e.g., Yonelinas, 2001) with better recognition being possible based on a weaker memory signal in the case of 2AFC. Nevertheless, Exp 2/d in Chapter 2 could not demonstrate an accuracy improvement with a 2AFC method compared to free recall. It shows that the amount of vivid detail that is stored in memory about episodic input without any regularity, is very limited.

Wolfe et al. (2023) employed a different kind of learning and testing design that is more similar to those used in Chapter 2, of the dissertation. First, their participants had to tell in an old/new setup whether they had seen particular objects before. Second, they also had to remember the locations of the objects on a grid and at a later point in time they had to locate those objects. Wolfe et al. found that location or memory for object detail was not as good as object recognition but several of their participants could recall the location of about 100 objects (out of 300) with high precision. While some might argue that this indicates a massive memory, this fidelity is not at the same level as the one reported in Brady et al. (2008), for example. I also reported in Chapter 2 that participants can achieve only ~0.3 level of accuracy on a [0, 1] scale when tested in free recall, while object identity recognition remains higher. This is an aspect of our results that is in line with Wolfe et al. (2023).

A further important aspect affecting the massivity of episodic memory is regularity and associations within the input. Persaud & Hemmer (2024) showed that as background and meaningful semantic associations are systematically removed from scenes, episodic memory accuracy decreases. My results in later chapters support this observation, both within and across scene regularity was key for high overall (although not exclusively episodic) memory accuracy, as humans learned input statistics and regularities and built internal models based on those effectively. Consequently, rather than asking whether or not short or long-term visual memory is massive, future research should consider the various contexts in which details of visual memories should be massive in order to solve real life problems efficiently, and see whether people can, indeed, perform in those cases as expected in terms of memory capacity.

In Chapter 2, I also explored whether real-world objects are represented as bound units or by object parts that can have their separate encoding and representation in long-term memory. To reiterate earlier results, while there is evidence for a holistic recollection of episodic memories (Horner et al., 2015), there is also proof, mostly from Brady and his colleagues, that objects parts have their independent storage and representation even in working memory (Brady et al., 2013; Utochkin & Brady, 2020; Markov et al., 2021). Moreover, combining both the holistic and independent representation idea, Kuhbandner (2020) has argued that depending on requirements in a given context, objects can be represented either as a set of independent elements or as bound units. Experiment 2/d showed that in about 28% of all trials, it is possible that either the orientation of an object or the identity of an object is not correctly recalled (miss response) while the other part is correctly recalled (hit response). Given these results, objects might be represented in an unbound way since parts of the representation can be forgotten while others are still remembered. On the other hand, analyses of subjective confidence and certainty reports in Experiment 2/d showed that whenever an object part is not recalled correctly, participants' subjective confidence/certainty in the other part also decreases significantly. Specifically, when comparing hit-miss responses (when the identity of the object was correctly recognized, but the orientation was not) to hit-hit responses (when both identity and orientation were correctly recognized), there was a marginally significant drop in orientation certainty. This by itself is not surprising as the orientation of the object was not correctly recognized in the hit-miss condition. However, there was a significant drop in object identity confidence when comparing hit-hit and hit-miss responses. This indicates that once vividness for one part of a memory weakens, it affects the subjective confidence for the other part of the memory as well even though that other part was recognized correctly. Consequently, taking subjective uncertainty into account, the default representation of an object seems to be holistic. This pattern was even more apparent in miss-hit responses (when the identity of the object was not recognized correctly but the orientation of the object was). The subjective confidence in identity responses in miss-hit compared to hit-hit decreased significantly. At the same time, subjective certainty in orientation responses also decreased highly significantly in miss-hit compared to hit-hit, demonstrating again that if performance for one part of the memory degrades, then confidence for the memory of the other part decreases, too. When only the orientation is recalled correctly, it may be the case that participants' subjective certainty in the orientation decreases because they know they did not remember the identity. There, it may be the case that this phenomenon is an indication that confidence relies on all information instead of an indication of holistic representation. However, in the case when only the identity was recalled correctly, there was a drop in confidence as well. And in that case the first response could not bias the second one, because the identity was the first response. All in all, similarly to the case with massivity, the strategy of encoding episodes into bound or unbound representations might depend on the context as Kuhbandner (2020) argued. To clarify this issue, the most valuable direction of research would be investigating the various cases and scenarios in which object representation in long-term memory is either bound or unbound.

Importantly, in Chapter 2 I investigated subjective (un)certainty both in working memory (perceptual decision-making) and in long-term visual episodic memory tasks. Previous research in perceptual decision-making showed that people utilize subjective certainty in their decisions (e.g., Körding & Wolpert, 2004; Ma & Jazayeri, 2014). In contrast, long-term memory has been investigated much less in this regard, and comparisons between the two memory systems have been missing. In addition, the methods used for assessing subjective certainty varied substantially. Many studies in long-term memory research used discrete rather than continuous scales for confidence measures and a separate step with an independent subtask to assess certainty. Arguably, our tablet-based method is a more instantaneous and cognitively less tainted measure of people's uncertainty representations than most methods that other studies have used. The instantaneous motor response reflects participants' true representations as much as possible, without potential higher-level biases (such as language) that could originate, for instance, from using a discrete, language-based scale. In addition, accuracy and certainty can be on the same scale in our design allowing for a fine-grained calibratedness measure.

Through the comparison between experiments 1 and 2/a, I showed that the levels and patterns of calibratedness are almost identical (well-calibrated) in working and long-term visual episodic memory indicating that the same underlying probabilistic representation is used upon recall. These results are important not only because there is a lack of studies investigating uncertainty in long-term episodic memory, but also because as Rahnev et al. (2020) showed, there are fewer studies that use a continuous scale for confidence ratings in long-term memory than in perception. Consequently, it is difficult to compare the two memory systems and decide whether metacognitive representations are domain-specific or domain-general. The highly similar calibratedness levels and patterns in Experiment 1 and Experiment 2/a suggest that uncertainty representations are probabilistic and domain general. Indirectly, these results are also in line with the idea put forth by Schurgin et al. (2020) that memories are ‘continuous in strength, not all or none’. In their study, they demonstrated that upon encoding the color of a stimulus, many other similar colors are encoded and represented. Chapter 2 presents a similar idea in that people recall a whole distribution of orientation values with their corresponding and meaningful probabilities.

6.2 Episodic Memories Embedded in Increasingly Complex Scenes

In the stimulus set of Chapter 3, individual objects were organized into simple scenes with particular connections between the objects. Experiment 3/a is the corresponding pure episodic memory experiment to the one presented in Chapter 2, as there were no perceptual or semantic connections between the objects in the scenes. Participants had an increased overall time to inspect the objects compared to Experiment 2/a (the pure episodic memory experiment in Chapter 2). In addition, while they had arguably more opportunity to make associations between objects due to the simultaneous presentation of multiple objects, their orientation recall accuracy did not significantly increase in Experiment 3/a compared to 2/a. This comparison between single vs. multiple element-based scenes echoes the study of Brady & Störmer (2022) using a working memory task in which case, objects but not simple stimuli such as color benefited from a sequential presentation. Brady & Störmer argued that with sequential presentation, the objects are processed in a deeper way than when they are presented simultaneously. The patterns of results in experiments 3/a and 2/a can be interpreted in the lights of Brady & Störmer (2020) since, in my design, when objects were presented simultaneously, participants had more time to inspect each object. Thus, with equal inspection time (per object) between sequential and simultaneous presentation, performance in the sequential case would have shown an advantage in long-term recall similarly to the case in working memory tasks.

On the other hand, a 6 second scene inspection time could be enough to study the individual objects in a complex scene in more detail and to make associations between them. In principle, these associations, in Experiment 3/a, can be only arbitrary since the objects in the scenes are completely unrelated. However, imaginary episodes are also possible to construct. Wimmer et al. (2020) showed that people can create connections between elements of an episode and that efficient learners recall them in a clustered, simultaneous manner, suggesting a build-up of a strong connection between the elements. Indeed, parts of an episode in the Wimmer et al. study could be linked more meaningfully together than in Experiment 3/a. This suggests (what Experiment 3/d also demonstrates) that efficient/structured encoding of a higher-level episode, such as a scene or story, is mainly (or exclusively) possible with meaningful semanticity or narrative in the input. This argument is strengthened further by the study of Cohn–Sheehy et al. (2022) also concluding that a storytelling structure plays a crucial role in how well separate events are recalled. All these studies support the notion that separate elements must be intelligibly connected in order to be recalled with higher precision.

Results in Experiment 3/b showed that surface-level scene regularity (perceptual gluing or collinearity) did not affect overall orientation recall accuracy. Although no post-experiment questionnaire was used in our study, it is highly unlikely that subjects did not realize the salient feature of collinearity in the scenes. Since scene gist is extracted rapidly and simultaneously on multiple levels (Oliva, 2005), collinearity is most probably processed, either implicitly or explicitly, within the allowed 6 second inspection time. A possible explanation of the puzzling lack of gain in accuracy due to perceptual structure comes from the study of Oliva & Torralba (2006) in which they argued that after initial scene recognition, the individual objects and their parts are further processed. Carrying this argument further, one possibility is that, in Experiment 3/b, a lack of initial high-level conceptual idea about the scenes (made up by three random objects), as a whole, disrupted subjects' ability to utilize the salient perceptual regularity. They kept looking for overall meaning in the scene to which collinearity could be hooked on, and not finding one they failed to fully utilize the perceptual structure for later more precise recall.

Supporting this argument, Experiment 3/c revealed that as soon as there were semantic connections between objects, their orientation was remembered significantly better overall compared to results with scenes made up by randomly chosen objects, such as in Experiment 3/a. Several studies showed the importance of semantic connections, both within and across scenes, in guiding attention, improving performance, and extracting regularities (Brady & Oliva, 2008; Wu et al., 2014; Hu & Jacobs, 2021). The results of Experiment 3/c are consistent with this account. As soon as subjects had the opportunity to process the scenes more conceptually via semantic connections, memory performance improved for perceptual details as well. Furthermore, the advantage of unglued objects in Experiment 3/c suggests that participants flexibly altered their encoding strategy by recognizing the connection between the glued objects but encoding the unglued, odd-one-out object with higher precision. This aligns with Bates et al. (2019), who reported that during statistical and categorical learning, humans adaptively allocate their encoding capacity based on expectations of schemas or regularities in upcoming trials formed from previous ones.

Experiment 3/d echoed the results of Experiment 3/c in that overall orientation performance was not significantly different between the two, while both were significantly higher than in Experiments 3/a. This result confirmed that semantic connections between objects was the main contributor to the increased overall orientation performance. The synergistic effect of perceptual and semantic gluing was evident in Experiment 3/d, where the glued objects showed a significant advantage (compared to unglueds) in orientation recall performance in contrast to Experiment 3/c, where the unglued ones were recalled better. While this effect is most likely due to participants forming more structured or chunked representations of the scenes, it may also reflect a shift in encoding strategy. Specifically, items that are easier to encode in a given scene, primarily due to inherent scene regularities, tend to attract more attention, ultimately resulting in more vivid memories driven by the structure of the input. Future research using eye tracking could explore these questions further complementing the present behavioral experiments. One outstanding question is whether participants spend most of their time inspecting the part and structures of the scene that they later represent and recall with higher accuracy or, each scene segment (object) in the scene gets roughly equal amount of inspection time, and the semantic structure makes certain elements ultimately more memorable.

Regarding the representation of uncertainty, the correlation between accuracy and uncertainty in Experiment 3/a was as high as in Experiment 1. This suggests that the representation of uncertainty in Experiment 3/a remained individual/object based in a well-calibrated manner. A competing possibility would be to represent the scene-level, overall uncertainty which can bias or decrease the well-calibratedness of the individual items. However, the results indicate otherwise, providing further evidence for the similarity between working and long-term memory in terms of the fundamental probabilistic representation of stimuli. The well-calibrated representation of uncertainty in Experiment 3/b, for both glued and unglued objects, revealed that the strategy for representing uncertainty is similar to what has been shown in previous studies, namely, item-based. The slightly reduced correlation between accuracy and certainty in Experiment 3/b suggests that adding perceptual gluing made the representation noisier, although not fundamentally different.

Experiments 3/c and 3/d were the first in which the calibration of either glued (Exp. 3/c) or unglued (Exp. 3/d) items noticeably decreased, although both remained significantly above chance. In Experiment 3/d, it is possible that the orientations of the unglued objects were recalled partly in relation to the glued ones. At the same time, certainty reports remained item-based, which explains the decreased correlation and the altered calibration curve compared to previous experiments. This result, the input regularity being utilized for recall but not for uncertainty reports, is even more pronounced in Chapters 4 and 5, where added input regularity becomes more dominant.

Finally, the 'chunking' analysis across all experiments in Chapter 3 indicated that, whenever possible due to the input regularity, participants began forming chunks in their representations that mirrored the structure of the input. In Experiment 3/a, where there was no regularity in the input, none of the objects within the scenes showed significant correlations with each other in terms of orientation recall accuracy. This suggests that participants were not effective at forming arbitrary associations between object pairs in the scenes. In Experiment 3/b, with perceptual gluing, the glued objects already formed a moderately strong chunk, suggesting that this connection was present in participants' representations but not strong enough to dominate scene recall accuracy. In Experiment 3/c, participants' chunking behavior aligned with the accuracy results: the unglued objects were recalled with higher accuracy than the glued ones and did not form a significant chunk with them. In addition, the glued objects did not form a significant chunk with each other either. This outcome aligns with expectations if the encoding strategy prioritizes the odd-one-out object, simply because it is easier to memorize (in terms of orientation) compared to the glued objects. The only highly significant chunk involving the glued objects appeared in Experiment 3/d, where the glued objects were strongly correlated in recall accuracy, and therefore likely in participants' representations as well. As expected, the unglued object did not form a significant chunk with the glued objects. Nevertheless, as previously mentioned, a thorough eye-tracking investigation would likely be valuable in uncovering the actual process by which people map out and recall simple scenes.

6.3 Semantic Learning and Representation of Uncertainty with Overarching Input Regularity

In Chapter 4, a type of semantic regularity spanning all stimuli was introduced, arguably the closest approximation to what long-term memory research refers to as semantic structure. Experiment 4/a did not show improved object identity performance compared to Experiment 2/a from Chapter 2, which is not surprising, as the semantic regularity applied only to the orientation level of the input. On the other hand, reducing memory load in one object dimension can potentially enhance the encoding precision of other features, especially if objects are encoded more holistically. Interestingly, however, this was not the case in Experiment 4/a. Furthermore, there was a significant drop in identity performance from Experiment 4/b to Experiment 4/c, most likely due to a simple numerosity effect. Importantly, Experiment 4/a showed a significant improvement in orientation-level performance compared to Experiment 2/a. This finding aligns with previous research, as several studies have shown that ensembles can enhance and bias recall performance for individual items (e.g., Hemmer & Steyvers, 2009; Utochkin & Brady, 2020). What has been less explored is whether this improvement results from participants learning the summary statistics of the input (e.g., mean and/or standard deviation), or whether the input regularity supports more effective episodic encoding of individual items.

Three lines of evidence from the experiments in Chapter 4 suggest that participants learn typical items not in isolation, but through regularity in the input, a pattern especially evident among those with higher overall memory performance. For these investigations, in all three experiments, the objects presented during the study phase were categorized as either middle or outer objects, depending on their position within the input Gaussian distribution. The first line of evidence is that across all experiments, middle objects were recalled with higher accuracy than outer objects. If recall performance for middle objects reflects semantic learning, and performance for outer objects reflects episodic learning, then this accuracy difference provides preliminary evidence for semantic learning across participants. To further analyze performance differences, subjects were divided into high and low performers based on overall orientation recall accuracy. In all four experiments, high performers (top 50%) outperformed low performers (bottom 50%) on both middle and outer object recall, revealing a consistent pattern.

Second, across all four experiments, the magnitude of the performance difference was greater for middle objects than for outer objects. This suggests that high performers are, above all, effective semantic learners, individuals who efficiently extract and represent input regularities. Nonetheless, these participants consistently demonstrated superior episodic recall as well. One possible explanation is that once the input regularity is accurately and explicitly learned, it becomes easier to perceive outliers as surprising, thereby promoting more faithful encoding of those items. In contrast, "average" objects may be treated as components of a semantic structure and therefore not encoded with high individual precision. These results align somewhat with previous studies attempting to uncover the dynamics of the semantic bias for typical and atypical items (Huttenlocher, Hedges, & Duncan, 1991; Huttenlocher et al., 2000; Duffy et al., 2010; Tompary & Thompson-Schill, 2021; Xu, Hemmer, & Zhang, 2024). However, those results and the results of Chapter 4 complement each other, as those studies showed the magnitude of the bias for typical and atypical items. On the other hand, my results targeted whether learning the input summary statistics also facilitate enhanced memory for outliers.

The third line of evidence that high performers are primarily effective semantic learners comes from an analysis of the orientation angles these participants drew under low-certainty conditions. In Experiment 4/a, the low-certainty responses of high performers were notably closer to the prior (i.e., the input Gaussian distribution) in terms of both mean and standard deviation than the responses of low performers. In Experiment 4/b, high performers' responses were closer to the prior mean, whereas low performers' responses were closer to the prior standard deviation. However, it is worth noting that even high performers' mean responses were still quite far from the prior mean (by about 18 degrees) suggesting that with 30 objects, encoding and recall remained largely episodic. Participants may not have had sufficient time to fully detect and exploit the input regularity.

In Experiment 4/c, there was little difference between high and low performers in the drawn angles under low certainty, both in terms of mean and standard deviation. The bootstrap means for both groups were around 20 degrees (i.e., a 10-degree distance from the prior mean), suggesting that both groups learned the input statistics to some extent. However, the high standard deviation of the drawn response angles (approximately 46 degrees) in both groups indicates that their representations were noisy. The fact that even high performers did not learn the input regularity more precisely with 90 objects is puzzling. One possibility is that the larger set size included more outliers, which may have caused confusion. Future research should further investigate the relationship between set size and the balance of semantic and episodic learning in long-term memory. Overall, the results suggest that presenting 60 objects may be a "sweet spot" for distinguishing high and low performers in terms of semantic learning. With 30 objects, neither group demonstrated sufficient learning of the prior. With 60, only high performers did. And with 90, both groups exhibited noisy but detectable learning of the prior. It should be noted that, in Experiment 4/a, a post-experiment questionnaire showed that classification into “implicit/non-learners” and “explicit learners” did not predict overall orientation performance or the extent of semantic learning. Future research should explore more deeply the relationship between statistical semantic learning and post-experiment verbal reports.

Regarding the representation of uncertainty in orientation responses, the overall quantitative pattern resembled that of purely episodic experiments in that there was a significant correlation between orientation accuracy and certainty. However, the calibration curve was flatter, particularly for middle objects across all three experiments. This suggests a probabilistic encoding strategy in which participants represent and use the summary statistics for orientation, yet their certainty reports remain anchored to individual items. As a result, a bias appears in the overall calibration curve.

This is notable because, when relying on summary statistics, participants should expect their orientation estimates to be reasonably accurate even with weak memory for specific items and might therefore be more confident overall. Yet their certainty reports do not reflect this statistical learning, implying that item-level certainty and semantic knowledge of input regularities are not fully integrated in memory representations. Future research should explore whether there are conditions under which uncertainty representations align more closely with acquired semantic knowledge.

6.4 Semantic Learning and Representation of Uncertainty with Overarching Input Regularity, Added Background, and Divided Attention

Chapter 5 aimed to investigate episodic and semantic learning within a more comprehensive framework by embedding objects in a background, modulating attention levels, and directing attention to multiple object details. Experiment 5/a showed that embedding objects in a background slightly reduced participants' overall performance, although neither object identity performance nor orientation recall accuracy decreased significantly compared to Experiment 4/a. Additionally, the general recall accuracy pattern, that middle objects are recalled with significantly higher accuracy than outer objects, remained very similar in Experiment 5/a to that in Experiment 4/a. On the one hand, the slight decrease is not surprising, given that processing a background requires extra cognitive resources, especially since the images were meaningful natural scenes. On the other hand, participants could make associations between the background and objects, which arguably made encoding the stimuli easier. Or at minimum, it is reasonable to assume that the fact that the objects are in a natural scene can provide a more naturalistic encoding environment than plain white backgrounds, and lead to better recall. In one experiment, Persaud and Hemmer (2024) investigated the effect of background removal from scenes. Their scenes were meaningful, coherent wholes, such as a kitchen. They found that removing the background—while preserving the layout of other elements like tables—impaired later recall at short presentation times (2 seconds). They argued that the background is important for rapid scene interpretation, aligning with Oliva's (2004, 2006) claim that gist is a quick understanding of the entire input.

In our case, Experiment 5/a used a similarly short presentation time (1 second), but unlike Persaud and Hemmer's study, the background did not facilitate more accurate recall, even for outliers (outer objects). One possible explanation is that although the background was a meaningful natural image, there was no deliberate semantic connection between the object and background (e.g., a fruit placed against the Taj Mahal). This further supports the idea that quick gist extraction and effective scene encoding require a meaningful overall layout and structure. Furthermore, in Experiment 5/a, the amount and quality of semantic learning for both good and bad performers were very similar to Experiment 4/a, suggesting that the overarching semantic regularity dominated encoding, especially since participants were instructed to remember the objects and their orientations. In this context, the background was simply part of the overall input layout that added noise rather than aiding more precise episodic or semantic encoding. The slight decrease in overall identity and orientation accuracy compared to Experiment 4/a supports the view that the background acted primarily as noise. Importantly, our reasoning suggests that people can allocate encoding resources in a task-dependent manner, effectively ignoring salient but irrelevant elements in the input.

In Experiment 5/b, the pattern that middle objects are recalled with significantly higher accuracy than outer objects became stronger. The effect size increased from medium (in Experiments 4/a and 5/a) to large in Experiment 5/b. This provides preliminary evidence that when people must attend to multiple tasks, the overall 'big picture' or gist of the input is prioritized, although episodic details are not completely lost. This aligns with the study of Allred et al. (2016) that shows that with high cognitive load the central tendency bias increases compared to low cognitive load. This idea was reinforced by additional analysis of semantic learning, which showed that the amount and quality of semantic learning for both good and bad performers were comparable to Experiments 4/a and 5/a. Together, Experiments 4/a, 5/a, and 5/b demonstrate that when task-relevant overarching semantic regularity is present in the input, it is prioritized and learned effectively, regardless of input manipulations such as background presence or divided attention. These results contradict findings by Greene and Naveh-Benjamin (2022, 2023), who showed that gist-based representations are also susceptible to forgetting under divided attention, though less so than episodic details. However, they argued in their 2022 study that younger people are more vulnerable to overall memory impairment under divided attention, and in their 2023 study they demonstrated gist forgetting after a 24-hour delay. In contrast, semantic learning in this dissertation appears to be robust.

Results from Experiment 5/c are puzzling. First, object identity recognition was significantly worse than in the baseline experiment, 4/a. This is not surprising in itself, as the task became harder due to the need to distinguish small object details. However, it is very surprising given that overall orientation recall accuracy was not significantly different from Experiment 4/a, neither for structured nor for non-structured participants.

Furthermore, the increase in orientation performance is surprising in the case of non-structured participants because they essentially could only answer the orientation question episodically, similar to Experiment 2/a. For non-structured participants, exclusive episodic learning, at least for the tested old items, is clearly shown by the analysis of drawn orientation responses. Their response angles were very far from the input Gaussian (about 30 degrees away). Yet, despite the lack of semantic learning, there was a borderline significant increase in orientation performance in Experiment 5/c (non-structured participants) compared to Experiment 2/a. It seems that knowing they needed to pay attention to small object details gave participants a boost in attention to the main feature to be encoded, which was orientation.

What is puzzling, then, is why other object features, such as color and texture, were poorly encoded, as shown by the decrease in identity performance. One possible explanation is that for object features other than orientation, participants did not know what would be tested during recall—they only knew that objects would differ in small details compared to the study phase. In contrast, the instruction for orientation was clear from the beginning that it must be recalled specifically later. Still, future research should further investigate the recall precision of specific object features depending on whether they are task-relevant or not, and under episodic versus semantic learning conditions.

Finally, in terms of calibratedness, experiments 5/a and 5/b, as well as structured participants from Experiment 5/c, showed very similar patterns to the base experiments containing overarching regularity (experiments 4/a, 4/b, and 4/c). Especially for middle objects, the learning of semantic regularity was apparent in the calibration curve. In the case of outer objects, the calibration plots resembled pure episodic memory experiments from Chapter 2. For non-structured participants in Experiment 5/c, both the correlation level and the calibration curve reflected episodic-like learning from previous experiments without semantic input structure, further confirming that their recall was episodic. In the case of Experiment 5/a, when a background and semantic regularity was simultaneously present in the input, it became more difficult for participants to integrate all this into their uncertainty representation. Future research should investigate the integration of multiple factors (such as semantic knowledge and background embedding) into people's representation of uncertainty.

6.5 Conclusions and Future Directions

This dissertation set out to achieve three major goals. First, to thoroughly investigate the representation of uncertainty in episodic memories under varying types and degrees of semantic regularity. Second, to examine the amount and quality of episodic and semantic learning given different kinds of semantic input regularities, while also accounting for individual differences in learning. Third, to explore episodic and semantic learning under more complex input modulations, such as background embedding and varying levels of attention.

To achieve the first goal, I used a fine-grained continuous subjective uncertainty measurement method for perceptual decision-making and long-term memory. Through this method, I demonstrated not only that subjective uncertainty is represented during both perceptual decision-making and the formation of long-term, pure episodic memories, but also that this representation is well-calibrated, as uncertainty reliably predicted objective performance across all levels. This is a significant result, as it supports the idea that the encoding and recall of episodic memory are carried out in a probabilistic manner, resulting in a recall representation as a full probability distribution over certain variables. Furthermore, I provided evidence that the fundamentally probabilistic nature of episodic memory representations remains unchanged, even when semantic regularity is imposed on the input - although, in these cases, uncertainty is not as predictive of objective performance as it is with pure episodic memories. This is because observers, on the one hand, learn and implicitly apply some of the acquired semantic regularities during their decisions, while on the other hand, they base their uncertainty judgments on how well they episodically remember the current item. Nevertheless, several open questions remain. First, my dissertation did not explore uncertainty representations at higher levels of memory, such as whole events or episodic stories. Therefore, future research should investigate the representation of uncertainty associated with episodic and semantic memories under various contexts and input complexities. Second, it would be beneficial to compare various uncertainty measurement methods across different encoding contexts to gain a comprehensive understanding of their differences and effectiveness. This exploration could be crucial, as it would likely shed light on what is actually meant by "uncertainty" in both short-term and long-term memory. Finally, a key open question regarding the representation of uncertainty in episodic memories with imposed semantic regularity is whether people are able to fully incorporate their semantic knowledge (or knowledge of other high-level input structure such as background embedding) into their uncertainty estimates, or the representation of uncertainty remains item-based across all contexts.

Regarding the second major goal of my thesis, I demonstrated that various types of semantic regularity improved overall long-term recall performance. Importantly, I showed that this improvement was due to genuine semantic learning: people responded based solely on their semantic knowledge when their certainty about the exact episode was low. I also found that the structure of the learned semantic knowledge mirrored the structure of the input regularity. This suggests that participants learned the semantic regularity directly from the input and applied it on a given trial, rather than relying on high-level approximations, such as point estimates. Additionally, I demonstrated substantial individual differences in semantic learning, as only about half of the participants learned the input regularity to a degree sufficient to influence their decisions. Future research could explore what exactly people learn under specific task requirements. While I provided evidence that participants learn more than just the mean of a distribution, this may not always be the case.

A further question for future research is the pervasiveness of long-term semantic learning. In this dissertation, the semantic input regularity was fairly straightforward to notice, yet around 50% of participants continued to rely on an episodic encoding strategy. Related to this, future research should explore the dynamics between episodic and semantic learning. While I found that better semantic learners were also better episodic learners, it remains unclear whether this increased episodic learning is a result of effective semantic learning, or simply reflects better overall encoding and memory ability among high-performing individuals.

Finally, as the third major goal, I investigated the effects of background embedding and varying levels of attention on the amount of episodic and semantic learning, as well as on the representation of uncertainty in both. I found that semantic learning was just as robust under conditions of background embedding, divided attention, and when participants had to attend to multiple object features, as it was when attention was not modulated. In addition, the proportion of participants who successfully learned the semantic input regularity remained around 50%, regardless of embedding or attention manipulations. Meanwhile, I also showed that the relative amount of episodic learning, compared to semantic learning, did change with attention modulations. Clearly, this only scratches the surface of a complex phenomenon, and further investigation is needed to clarify the vulnerability of different types of semantic learning under varying levels of attentional demand.

All in all, many questions and layers of detail remain to be uncovered in our effort to understand the fine-grained representational form and dynamics of episodic and semantic memories. Hopefully, our work has contributed to this broad endeavor by investigating several aspects of long-term memory together, in a detailed and integrated way.

REFERENCES

Allred, S.R., Crawford, L.E., Duffy, S. *et al.* Working memory and spatial judgments: Cognitive load increases the central tendency bias. *Psychon Bull Rev* **23**, 1825–1831 (2016). https://doi.org/10.3758/s13423-016-1039-0

Baddeley A. (2000). The episodic buffer: a new component of working memory?. *Trends in cognitive sciences*, *4*(11), 417–423. https://doi.org/10.1016/s1364-6613(00)01538-2

Baddeley A. (2003). Working memory and language: an overview. *Journal of communication disorders*, *36*(3), 189–208. https://doi.org/10.1016/s0021-9924(03)00019-4

Bae, G.-Y., Olkkonen, M., Allred, S. R., & Flombaum, J. I. (2015). Why some colors appear more memorable than others: A model combining categories and particulars in color working memory. *Journal of Experimental Psychology: General, 144*(4), 744–763. https://doi.org/10.1037/xge0000076

Balaban, H., Assaf, D., Arad Meir, M., & Luria, R. (2020). Different features of real-world objects are represented in a dependent manner in long-term memory. *Journal of experimental psychology. General*, *149*(7), 1275–1293. https://doi.org/10.1037/xge0000716

Bates, C. J., Lerch, R. A., Sims, C. R., & Jacobs, R. A. (2019). Adaptive allocation of human visual working memory capacity during statistical and categorical learning. *Journal of vision*, *19*(2), 11. https://doi.org/10.1167/19.2.11

Bona, S., & Silvanto, J. (2014). Accuracy and confidence of visual short-term memory do not go hand-in-hand: behavioral and neural dissociations. *PloS one*, *9*(3), e90808. https://doi.org/10.1371/journal.pone.0090808

Bornstein, A. M., & Norman, K. A. (2017). Reinstated episodic context guides sampling-based decisions for reward. *Nature neuroscience*, *20*(7), 997–1003. https://doi.org/10.1038/nn.4573

Brady, T. F., & Oliva, A. (2008). Statistical learning using real-world scenes: extracting categorical regularities without conscious intent. *Psychological science*, *19*(7), 678–685. https://doi.org/10.1111/j.1467-9280.2008.02142.x

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. Proceedings of the National Academy of Sciences of the United States of America, 105(38), 14325–14329. https://doi.org/10.1073/pnas.0803390105

Brady, T. F., Konkle, T., & Alvarez, G. A. (2009). Compression in visual working memory: using statistical regularities to form more efficient memory representations. *Journal of experimental psychology. General*, *138*(4), 487–502. https://doi.org/10.1037/a0016797

Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: ensemble statistics bias memory for individual items. *Psychological science*, *22*(3), 384–392. https://doi.org/10.1177/0956797610397956

Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: Beyond individual items and toward structured representations. *Journal of vision*, *11*(5), 4. https://doi.org/10.1167/11.5.4

Brady, T. F., & Tenenbaum, J. B. (2013). A probabilistic model of visual working memory: Incorporating higher order regularities into working memory capacity estimates. *Psychological review*, *120*(1), 85–109. https://doi.org/10.1037/a0030779

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2013). Real-world objects are not represented as bound units: independent forgetting of different object details from visual memory. *Journal of experimental psychology. General*, *142*(3), 791–808. https://doi.org/10.1037/a0029649

Brady, T. F., Schacter, D. L., & Alvarez, G. (2018). The adaptive nature of false memories is revealed by gist-based distortion of true memories. https://doi.org/10.31234/osf.io/zeg95

Brady, T. F., & Störmer, V. S. (2022). The role of meaning in visual working memory: Real-world objects, but not simple features, benefit from deeper processing. *Journal of experimental psychology. Learning, memory, and cognition*, *48*(7), 942–958. https://doi.org/10.1037/xlm0001014

Brod, G., Werkle-Bergner, M., & Shing, Y. L. (2013). The influence of prior knowledge on memory: a developmental cognitive neuroscience perspective. *Frontiers in behavioral neuroscience*, *7*, 139. https://doi.org/10.3389/fnbeh.2013.00139

Chetverikov, A., Campana, G., & Kristjánsson, Á. (2016). Building ensemble representations: How the shape of preceding distractor distributions affects visual search. *Cognition*, *153*, 196–210. https://doi.org/10.1016/j.cognition.2016.04.018

Chetverikov, A., Campana, G., & Kristjánsson, Á. (2017). Rapid learning of visual ensembles. *Journal of vision*, *17*(2), 21. https://doi.org/10.1167/17.2.21

Chetverikov, A., Campana, G., & Kristjánsson, Á. (2017). Representing Color Ensembles. *Psychological science*, *28*(10), 1510–1517. https://doi.org/10.1177/0956797617713787

Chetverikov, A., Campana, G., & Kristjánsson, Á. (2020). Probabilistic rejection templates in visual working memory. *Cognition*, *196*, 104075. https://doi.org/10.1016/j.cognition.2019.104075

Cohn-Sheehy, B. I., Delarazan, A. I., Crivelli-Decker, J. E., Reagh, Z. M., Mundada, N. S., Yonelinas, A. P., Zacks, J. M., & Ranganath, C. (2022). Narratives bridge the divide between distant events in episodic memory. *Memory & cognition*, *50*(3), 478–494. https://doi.org/10.3758/s13421-021-01178-x

Cowan N. (2008). What are the differences between long-term, short-term, and working memory?. *Progress in brain research*, *169*, 323–338. https://doi.org/10.1016/S0079-6123(07)00020-9

Cowan, N., Bao, C., Bishop-Chrzanowski, B. M., Costa, A. N., Greene, N. R., Guitard, D., Li, C., Musich, M. L., & Ünal, Z. E. (2024). The Relation Between Attention and Memory. *Annual review of psychology*, *75*, 183–214. https://doi.org/10.1146/annurev-psych-040723-012736

Dayan, P., Hinton, G. E., Neal, R. M., & Zemel, R. S. (1995). The Helmholtz machine. *Neural Computation, 7*(5), 889–904. https://doi.org/10.1162/neco.1995.7.5.889

Dodson, C. S., & Krueger, L. E. (2006). I misremember it well: why older adults are unreliable eyewitnesses. *Psychonomic bulletin & review*, *13*(5), 770–775. https://doi.org/10.3758/bf03193995

Dodson, C. S., Bawa, S., & Krueger, L. E. (2007). Aging, metamemory, and high-confidence errors: a misrecollection account. *Psychology and aging*, *22*(1), 122–133. https://doi.org/10.1037/0882-7974.22.1.122

Duffy, S., Huttenlocher, J., Hedges, L. V., & Crawford, L. E. (2010). Category effects on stimulus estimation: Shifting and skewed frequency distributions. *Psychonomic Bulletin & Review, 17*(2), 224–230. https://doi.org/10.3758/PBR.17.2.224

Fang, J., Rüther, N., Bellebaum, C., Wiskott, L., & Cheng, S. (2018). The Interaction between Semantic Representation and Episodic Memory. *Neural computation*, *30*(2), 293–332. https://doi.org/10.1162/neco_a_01044

Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene?. *Journal of vision*, *7*(1), 10. https://doi.org/10.1167/7.1.10

Fiser, J., Berkes, P., Orbán, G., & Lengyel, M. (2010). Statistically optimal perception and learning: from behavior to neural representations. Trends in cognitive sciences, 14(3), 119–130. https://doi.org/10.1016/j.tics.2010.01.003

Graham, K. S., Simons, J. S., Pratt, K. H., Patterson, K., & Hodges, J. R. (2000). Insights from semantic dementia on the relationship between episodic and semantic memory. *Neuropsychologia*, *38*(3), 313–324. https://doi.org/10.1016/s0028-3932(99)00073-1

Greenberg, D. L., & Verfaellie, M. (2010). Interdependence of episodic and semantic memory: evidence from neuropsychology. *Journal of the International Neuropsychological Society : JINS*, *16*(5), 748–753. https://doi.org/10.1017/S1355617710000676

Greene, N. R., & Naveh-Benjamin, M. (2022). The effects of divided attention at encoding on specific and gist-based associative episodic memory. *Memory & cognition*, *50*(1), 59–76. https://doi.org/10.3758/s13421-021-01196-9

Greene, N. R., Chism, S., & Naveh-Benjamin, M. (2022). Levels of specificity in episodic memory: Insights from response accuracy and subjective confidence ratings in older adults and in younger adults under full or divided attention. *Journal of experimental psychology. General*, *151*(4), 804–819. https://doi.org/10.1037/xge0001113

Greene, N. R., & Naveh-Benjamin, M. (2023). Forgetting of specific and gist visual associative episodic memory representations across time. *Psychonomic bulletin & review*, *30*(4), 1484–1501. https://doi.org/10.3758/s13423-023-02256-8

Griffiths, T. L., & Steyvers, M. (2002). A probabilistic approach to semantic representation. In W. D. Gray & C. D. Schunn (Eds.), *Proceedings of the 24th Annual Conference of the Cognitive Science Society*. 381–386. Cognitive Science Society. https://cocosci.princeton.edu/tom/papers/semrep.pdf

Haberman, J., Brady, T. F., & Alvarez, G. A. (2015). Individual differences in ensemble perception reveal multiple, independent levels of ensemble representation. *Journal of experimental psychology. General*, *144*(2), 432–446. https://doi.org/10.1037/xge0000053

Hassabis, D., & Maguire, E. A. (2007). Deconstructing episodic memory with construction. *Trends in cognitive sciences*, *11*(7), 299–306. https://doi.org/10.1016/j.tics.2007.05.001

Hembacher, E., & Ghetti, S. (2014). Don't look at my answer: subjective uncertainty underlies preschoolers' exclusion of their least accurate memories. *Psychological science*, *25*(9), 1768–1776. https://doi.org/10.1177/0956797614542273

Hemmer, P., & Steyvers, M. (2009). A Bayesian account of reconstructive memory. *Topics in cognitive science*, *1*(1), 189–202. https://doi.org/10.1111/j.1756-8765.2008.01010.x

Hemmer, P., & Steyvers, M. (2009). Integrating episodic memories and prior knowledge at multiple levels of abstraction. Psychonomic bulletin & review, 16(1), 80–87. https://doi.org/10.3758/PBR.16.1.80Hemmer, P., & Persaud, K. (2014). Interaction between categorical knowledge and episodic memory across domains. *Frontiers in psychology*, *5*, 584. https://doi.org/10.3389/fpsyg.2014.00584

Hemmer, P., Tauber, S., & Steyvers, M. (2015). Moving beyond qualitative evaluations of Bayesian models of cognition. *Psychonomic Bulletin & Review, 22*(3), 614–628. https://doi.org/10.3758/s13423-014-0725-z

Henderson, J. M., Weeks, P.A., & Jr., Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *J. Exp. Psychol. Hum. Percept. Perform.* 25(1), 210–228.

Horner, A. J., Bisby, J. A., Bush, D., Lin, W. J., & Burgess, N. (2015). Evidence for holistic episodic recollection via hippocampal pattern completion. *Nature communications*, *6*, 7462. https://doi.org/10.1038/ncomms8462

Hu, R., & Jacobs, R. A. (2021). Semantic influence on visual working memory of object identity and location. *Cognition*, *217*, 104891. https://doi.org/10.1016/j.cognition.2021.104891

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: prototype effects in estimating spatial location. *Psychological review*, *98*(3), 352–376. https://doi.org/10.1037/0033-295x.98.3.352

Huttenlocher, J., Hedges, L. V., & Vevea, J. L. (2000). Why do categories affect stimulus judgment?. *Journal of experimental psychology. General*, *129*(2), 220–241. https://doi.org/10.1037//0096-3445.129.2.220

Irish, M., & Piguet, O. (2013). The pivotal role of semantic memory in remembering the past and imagining the future. *Frontiers in behavioral neuroscience*, *7*, 27. https://doi.org/10.3389/fnbeh.2013.00027

Kadar, I., & Ben-Shahar, O. (2012). A perceptual paradigm and psychophysical evidence for hierarchy in scene gist processing. *Journal of vision*, *12*(13), 16. https://doi.org/10.1167/12.13.16

Kim, H., & Cabeza, R. (2007). Trusting our memories: dissociating the neural correlates of confidence in veridical versus illusory memories. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *27*(45), 12190–12197. https://doi.org/10.1523/JNEUROSCI.3408-07.2007

Koblinger, Á., Fiser, J., & Lengyel, M. (2021). Representations of uncertainty: where art thou?. *Current opinion in behavioral sciences*, *38*, 150–162. https://doi.org/10.1016/j.cobeha.2021.03.009

Koeritzer, M. A., Rogers, C. S., Van Engen, K. J., & Peelle, J. E. (2018). The Impact of Age, Background Noise, Semantic Ambiguity, and Hearing Loss on Recognition Memory for Spoken Sentences. *Journal of speech, language, and hearing research : JSLHR*, *61*(3), 740–751. https://doi.org/10.1044/2017_JSLHR-H-17-0077

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Scene memory is more detailed than you think: the role of categories in visual long-term memory. Psychological science, 21(11), 1551–1556. https://doi.org/10.1177/0956797610385359

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *Journal of experimental psychology. General*, *139*(3), 558–578. https://doi.org/10.1037/a0019165

Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. Nature, 427(6971), 244–247. https://doi.org/10.1038/nature02169

Kuhbandner C. (2020). Real-World Objects Are Represented in Visual Long-Term Memory Both as Unbound Features and as Bound Objects. *Frontiers in psychology*, *11*, 580667. https://doi.org/10.3389/fpsyg.2020.580667

Lad, M., Holmes, E., Chu, A., & Griffiths, T. D. (2020). Speech-in-noise detection is related to auditory working memory precision for frequency. *Scientific reports*, *10*(1), 13997. https://doi.org/10.1038/s41598-020-70952-9

Leist, L., Lachmann, T., & Klatte, M. (2025). Impact of irrelevant speech and non-speech sounds on serial recall of verbal and spatial items in children and adults. *Scientific reports*, *15*(1), 1951. https://doi.org/10.1038/s41598-025-85855-w

Lengyel, Máté & Koblinger, Ádám & Popović, Marjena & Fiser, Jozsef. (2015). On the role of time in perceptual decision making.

Lew, T. F., Pashler, H. E., & Vul, E. (2016). Fragile associations coexist with robust memories for precise details in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(3), 379–393.

Li, H. H., Sprague, T. C., Yoo, A. H., Ma, W. J., & Curtis, C. E. (2021). Joint representation of working memory and uncertainty in human cortex. Neuron, 109(22), 3699–3712.e6.

Liu, Y., Su, Y., Xu, G., & Pei, M. (2018). When do you know what you know? The emergence of memory monitoring. *Journal of experimental child psychology*, *166*, 34–48. https://doi.org/10.1016/j.jecp.2017.06.014

Ma, W. J., & Jazayeri, M. (2014). Neural coding of uncertainty and probability. Annual review of neuroscience, 37, 205–220. https://doi.org/10.1146/annurev-neuro-071013-014017

Magas, D., Koblinger, Á., Lengyel, M., & Fiser, J. (2022). Reducible and irreducible uncertainty in low-level visual representations. *Journal of Vision, 22*(14), 3877. https://doi.org/10.1167/jov.22.14.3877

Magas, D., & Fiser, J. (2023). Probabilistic encoding and well-calibratedness of long-term episodic memory. *Journal of Vision, 23*(9), 5325. https://doi.org/10.1167/jov.23.9.5325

Markov, Y. A., Utochkin, I. S., & Brady, T. F. (2021). Real-world objects are not stored in holistic representations in visual working memory. *Journal of vision*, *21*(3), 18. https://doi.org/10.1167/jov.21.3.18

Miner, A. E., Schurgin, M. W., & Brady, T. F. (2020). Is working memory inherently more "precise" than long-term memory? Extremely high fidelity visual long-term memories for frequently encountered objects. *Journal of experimental psychology. Human perception and performance*, *46*(8), 813–830. https://doi.org/10.1037/xhp0000748

Nagy, D. G., & Orban, G. (2016). Episodic memory as a prerequisite for online updates of model structure. *38th ann conf cog sci soc*. 2699–2704.

Nagy, D. G., Török, B., & Orbán, G. (2020). Optimal forgetting: Semantic compression of episodic memories. *PLoS computational biology*, *16*(10), e1008367. https://doi.org/10.1371/journal.pcbi.1008367

Oliva, A. (2005). Gist of the scene. In: Itti, L.; Rees, G.; Tsotsos, JK., editors. Neurobiology of attention. San Diego, CA: Elsevier, 251-256.

Oliva, A., & Torralba, A. (2006). Building the gist of a scene: the role of global image features in recognition. *Progress in brain research*, *155*, 23–36. https://doi.org/10.1016/S0079-6123(06)55002-2

Persaud, K., & Hemmer, P. (2016). The dynamics of fidelity over the time course of long-term memory. *Cognitive Psychology, 88,* 1–21. https://doi.org/10.1016/j.cogpsych.2016.05.003

Persaud, K., Hemmer, P. The influence of functional components of natural scenes on episodic memory. *Sci Rep* 14, 30313 (2024). https://doi.org/10.1038/s41598-024-81900-2

Rahnev, D., Desender, K., Lee, A. L. F., Adler, W. T., Aguilar-Lleyda, D., Akdoğan, B., Arbuzova, P., Atlas, L. Y., Balcı, F., Bang, J. W., Bègue, I., Birney, D. P., Brady, T. F., Calder-Travis, J., Chetverikov, A., Clark, T. K., Davranche, K., Denison, R. N., Dildine, T. C., Double, K. S., … Zylberberg, A. (2020). The Confidence Database. *Nature human behaviour*, *4*(3), 317–325. https://doi.org/10.1038/s41562-019-0813-1

Ramey, M. M., Henderson, J. M., & Yonelinas, A. P. (2020). The spatial distribution of attention predicts familiarity strength during encoding and retrieval. *Journal of experimental psychology. General*, *149*(11), 2046–2062. https://doi.org/10.1037/xge0000758

Renoult, L., Davidson, P. S., Palombo, D. J., Moscovitch, M., & Levine, B. (2012). Personal semantics: at the crossroads of semantic and episodic memory. *Trends in cognitive sciences*, *16*(11), 550–558. https://doi.org/10.1016/j.tics.2012.09.003

Reyna, V. F., & Brainerd, C. J. (1998). Fuzzy-trace theory and false memory: new frontiers. *Journal of experimental child psychology*, *71*(2), 194–209. https://doi.org/10.1006/jecp.1998.2472

Richter, F. R., Bays, P. M., Jeyarathnarajah, P., & Simons, J. S. (2019). Flexible updating of dynamic knowledge structures. *Scientific reports*, *9*(1), 2272. https://doi.org/10.1038/s41598-019-39468-9

Rubin, D. C., & Umanath, S. (2015). Event memory: A theory of memory for laboratory, autobiographical, and fictional events. *Psychological review*, *122*(1), 1–23. https://doi.org/10.1037/a0037907

Sadeh, T., Ozubko, J. D., Winocur, G., & Moscovitch, M. (2016). Forgetting Patterns Differentiate Between Two Forms of Memory Representation. *Psychological science*, *27*(6), 810–820. https://doi.org/10.1177/0956797616638307

Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: the prospective brain. *Nature reviews. Neuroscience*, *8*(9), 657–661. https://doi.org/10.1038/nrn2213

Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M., & Norman, K. A. (2017). Complementary learning systems within the hippocampus: a neural network modelling approach to reconciling episodic memory with statistical learning.

Philosophical transactions of the Royal Society of London. Series B, Biological sciences, *372*(1711), 20160049. https://doi.org/10.1098/rstb.2016.0049

Schroeder, S. R., & Marian, V. (2014). Bilingual episodic memory: How speaking two languages influences remembering. In R. R. Heredia & J. Altarriba (Eds.), *Foundations of Bilingual Memory*. 113–132. Springer New York. https://doi.org/10.1007/978-1-4614-9218-4_6

Schurgin, M. W., Wixted, J. T., & Brady, T. F. (2020). Psychophysical scaling reveals a unified theory of visual memory strength. *Nature human behaviour*, *4*(11), 1156–1172. https://doi.org/10.1038/s41562-020-00938-0

Sherman, B. E., & Turk-Browne, N. B. (2020). Statistical prediction of the future impairs episodic encoding of the present. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(37), 22760–22770. https://doi.org/10.1073/pnas.2013291117

Shoval, R., Gronau, N., & Makovski, T. (2023). Massive visual long-term memory is largely dependent on meaning. *Psychonomic bulletin & review*, *30*(2), 666–675. https://doi.org/10.3758/s13423-022-02193-y

Standing L. (1973). Learning 10,000 pictures. The Quarterly journal of experimental psychology, 25(2), 207–222. https://doi.org/10.1080/14640747308400340

Steyvers, M., Griffiths, T. L., & Dennis, S. (2006). Probabilistic inference in human semantic memory. *Trends in cognitive sciences*, *10*(7), 327–334. https://doi.org/10.1016/j.tics.2006.05.005

Steyvers, M., & Hemmer, P. (2012). Reconstruction from memory in naturalistic environments. In B. H. Ross (Ed.), *The psychology of learning and motivation*. 125–144. Elsevier Academic Press. https://doi.org/10.1016/B978-0-12-394393-4.00004-2

Stocker, A. A., & Simoncelli, E. P. (2007). A Bayesian Model of Conditioned Perception. *Advances in neural information processing systems*, *2007*, 1409–1416.

Theeuwes, J., Bogaerts, L., & van Moorselaar, D. (2022). What to expect where and when: how statistical learning drives visual selection. *Trends in cognitive sciences*, *26*(10), 860–872. https://doi.org/10.1016/j.tics.2022.06.001

Tompary, A., & Thompson-Schill, S. L. (2021). Semantic influences on episodic memory distortions. *Journal of Experimental Psychology: General, 150*(9), 1800–1824. https://doi.org/10.1037/xge0001017

Tulving, E. (1993). What Is Episodic Memory? *Current Directions in Psychological Science*, 2(3), 67-70. http://www.jstor.org/stable/20182204

Tulving E. (2002). Episodic memory: from mind to brain. *Annual review of psychology*, *53*, 1–25. https://doi.org/10.1146/annurev.psych.53.100901.135114

Utochkin, I. S., & Brady, T. F. (2020). Independent storage of different features of real-world objects in long-term memory. *Journal of experimental psychology. General*, *149*(3), 530–549. https://doi.org/10.1037/xge0000664

Utochkin, I. S., & Brady, T. F. (2020). Individual representations in visual working memory inherit ensemble properties. *Journal of experimental psychology. Human perception and performance*, *46*(5), 458–473. https://doi.org/10.1037/xhp0000727

Wang, Y., Yang, K., Zhou, S., Zhang, H., Ma, T., Shi, X., & Ma, W. (2025). Bilingual proficiency effects on word recall and recognition. *Behavioral Sciences, 15*(4), Article 437. https://doi.org/10.3390/bs15040437

Wimmer, G. E., Liu, Y., Vehar, N., Behrens, T. E. J., & Dolan, R. J. (2020). Episodic memory retrieval success is associated with rapid replay of episode content. *Nature neuroscience*, *23*(8), 1025–1033. https://doi.org/10.1038/s41593-020-0649-z

Wolfe, J. M., Wick, F. A., Mishra, M., DeGutis, J., & Lyu, W. (2023). Spatial and temporal massive memory in humans. Current biology: CB, 33(2), 405–410.e4. https://doi.org/10.1016/j.cub.2022.12.040

Wood, R., Baxter, P., & Belpaeme, T. (2012). A review of long-term memory in natural and synthetic systems. *Adaptive Behavior, 20*(2), 81–103. https://doi.org/10.1177/1059712311421219

Wu, C. C., Wick, F. A., & Pomplun, M. (2014). Guidance of visual attention by semantic information in real-world scenes. *Frontiers in psychology*, *5*, 54. https://doi.org/10.3389/fpsyg.2014.00054

Xu, Z., Hemmer, P., & Zhang, Q. (2025). Towards a generalized Bayesian model of reconstructive memory. *Computational Brain & Behavior, 8*, 134–146. https://doi.org/10.1007/s42113-024-00222-8

Yonelinas A. P. (2001). Components of episodic memory: the contribution of recollection and familiarity. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *356*(1413), 1363–1374. https://doi.org/10.1098/rstb.2001.0939

Yonelinas, A. P. (2002). The Nature of Recollection and Familiarity. *A Review of 30 Years of Research,* 46(3), 441-517.

Zotow, E., Bisby, J. A., & Burgess, N. (2020). Behavioral evidence for pattern separation in human episodic memory. *Learning & memory (Cold Spring Harbor, N.Y.)*, *27*(8), 301–309. https://doi.org/10.1101/lm.051821.120

APPENDIX

After Experiment Questionnaire (~5 minutes)

NAME:

DATE:

Please answer the following questions and indicate on a scale from 1-10 how sure you are in your answer. Guide for the scale -> 0: I have no idea, 10: I am absolutely sure.

1, Did you notice anything strange in connection with the orientations of the objects?

2, If yes, what was it?

3, If yes, when did you notice it (after 10, 20, … , 50, etc. presentations)?

4, Were the orientation of the objects equally distributed or not, during the presentations?

5, If they were not equally distributed, were there one or two distinguished/main orientations?

6, Roughly what were the distinguished/main orientations (the below figure can help as a guide)?