ABSTRACTION, CONSOLIDATION, AND EXPLICITNESS IN SPATIO-TEMPORAL VISUAL STATISTICAL LEARNING

DOMINIK GARBER

SUBMITTED TO
CENTRAL EUROPEAN UNIVERSITY
DEPARTMENT OF COGNITIVE SCIENCE

IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY IN COGNITIVE SCIENCE

JÓZSEF FISER
PRIMARY SUPERVISOR

MÁTÉ LENGYEL
SECONDARY SUPERVISOR

VIENNA, AUSTRIA
2024

Declaration of Authorship

I hereby declare that this submission is my own work, and to the best of my knowledge it contains no materials previously published or written by another person or which have been accepted for the award of any other degree or diploma at Central European University or any other educational institution, except where due acknowledgment is made in the form of bibliographical reference.

Dominik Garber

Abstract

Visual statistical learning (VSL) describes how humans automatically and implicitly become sensitive to the statistics of visual input in the absence of supervision or reinforcement. Research on VSL usually focuses on learning either temporal or spatial regularities and almost always excludes the influence of prior knowledge. In this dissertation, I present a reconceptualization of VSL as part of a larger human unsupervised learning system operating by combining lower-level spatio-temporal co-occurrence statistics and higher-level top-down biases. I identified three types of higher-level biases affecting statistical learning: (1) pre-existing biases independent of properties of the experiment, (2) biases formed based on the abstraction of learned low-level statistics, and (3) biases based on observed higher-level features of the input. Furthermore, I identified important moderators of this hierarchical learning system: explicitness and consolidation of knowledge.

Extending the classical spatial VSL paradigm to a transfer learning paradigm, I found that while participants with explicit knowledge could immediately abstract from their acquired representations and generalize to novel input, participants with implicit knowledge showed a structural novelty effect in immediate transfer. This means they were better at learning novel input that was not aligned with what they had learned before. However, after a period of asleep consolidation, participants with implicit knowledge switched their behavior and showed generalization, as the participants with explicit knowledge did before. Using control experiments, I confirmed that this effect is specific to sleep and could not be explained simply by time passing or a time-of-day effect. Furthermore, using matched sample analysis, I demonstrated that differences in the strength of initial learning cannot explain the qualitative differences found between participants with explicit and implicit knowledge.

In order to combine the previously disjoint lines of spatial and temporal VSL, I developed a novel spatio-temporal visual statistical learning paradigm. There, spatially defined patterns were unfolding to the observer over time. I demonstrated that implicit learning is possible for spatio-temporal input and provided experimental evidence that the temporal statistics of the input were used for the implicit acquisition of spatial patterns. Furthermore, I showed that when confronting participants with the complexity of spatio-temporal input, top-down, bottom-up interactions naturally emerged, linking this line of research with the VSL transfer learning paradigm described above. I found that both the overall motion direction and the overall arrangement of shapes can bias participants learning and their beliefs about what types of structures are present in the input. Furthermore, by combining the spatio-temporal VSL paradigm with a prediction task, I found that participants with explicit knowledge but not participants with implicit knowledge can use it for prediction, adding to the findings on differences between explicit and implicit representations described above.

Overall, this dissertation demonstrates that the narrow limitation and control that enabled the initial success of VSL research need to be carefully and incrementally overcome to understand the role of VSL in the overall human cognitive system. It does so by introducing two new VSL paradigms that enable novel, systematic ways of investigating the human unsupervised learning system.

Acknowledgments

I want to mention some of the people who helped me during my studies and dissertation research.

First, I owe thanks to all the members of the Vision Lab at CEU. Without your support and feedback, this would not have been possible. Especially, I would like to thank my supervisor and head of the lab, József Fiser, for his help, support, and guidance during the past five years. I would also like to thank my secondary supervisor Máté Lengyel, for his feedback along the way.

Furthermore, I would like to thank the members of the Department of Cognitive Science at CEU for invaluable feedback and learning opportunities. I owe special thanks to the members of my cohort who made my move to Budapest a great experience, both at our office and locked away at home.

I would like to thank the LLAMB Lab at the Haskins Laboratories and Yale for the opportunities to work with and learn from them and for the warm welcome I received there. Especially, I would like to thank Dick Aslin.

Last but not least, I thank my family and friends for their support throughout the years and for finally stopping to ask what exactly it is that I am doing.

Table of Contents

List of Figures

CHAPTER 1

Introduction

Statistical learning (SL) is a form of unsupervised implicit learning that automatically allows humans and other animals to become sensitive to regularities in perceptual input. Due to the vast amount of perceptual input that people experience and the common absence of obvious reinforcement or labeling, such learning is critical. SL is often treated as a bottom-up process for identifying and storing reappearing patterns in perceptual input based on elementwise co-occurrence statistics. Most research in SL tried to eliminate top-down influences to isolate the effects of varying input statistics while eliminating the effects of prior knowledge. However, there is mounting evidence that this simplification does not hold up to the ecological role of SL. The first objective of this dissertation was a reconceptualization of *visual statistical learning* (VSL), putting bottom-up and top-down influences on equal standing and showing that what is canonically called VSL is part of a larger, hierarchically structured system of human *unsupervised learning*. This aims to combine previously separated approaches and concepts from outside and within the statistical learning literature, showing how our understanding of the underlying phenomena can be advanced this way.

Artificial separations also exist within the SL literature. Almost all of the VSL literature focused either on spatial or temporal regularities, ignoring that real-world visual input is always a combination of both. Due to the nature of ecologically relevant visual input, fully understanding the role and mechanism of VSL requires us to understand how it deals with spatio-temporal regularities. Therefore, the second objective of this dissertation was the development of a joint spatio-temporal VSL paradigm to enable systematic investigations of how the temporal and spatial statistics of the input interact in unsupervised learning. As a corollary of this step towards real-world complexity, we saw that bottom-up, top-down interactions naturally come into sight. Overall, the findings presented in this dissertation paint a picture of VSL as an interaction of lower-level spatio-temporal co-occurrence statistics and higher-level top-down biases. Therein, the multitude of available features across levels are flexible combined to achieve a congruent and comprehensive yet parsimonious interpretation of the world.

The current chapter defines key terminology and summarizes the dissertation's main argument, justifying and contextualizing it with previous statistical learning research and foreshadowing the main results of the dissertation. I show that while statistical learning paradigms were initially designed to focus on low-level co-occurrence statistics and exclude any higher-level knowledge (1.1.1), it was later shown that SL is critically influenced by higher-level knowledge (1.1.2), and propose a view of statistical learning as part of a larger, hierarchically structured, unsupervised learning system (1.1.3). I furthermore argue that the artificial separation of visual statistical learning research into a spatial and a temporal literature is detrimental to understanding how humans' unsupervised learning of visual input can operate in ecologically relevant settings, introducing a novel spatio-temporal VSL paradigm (1.1.4).

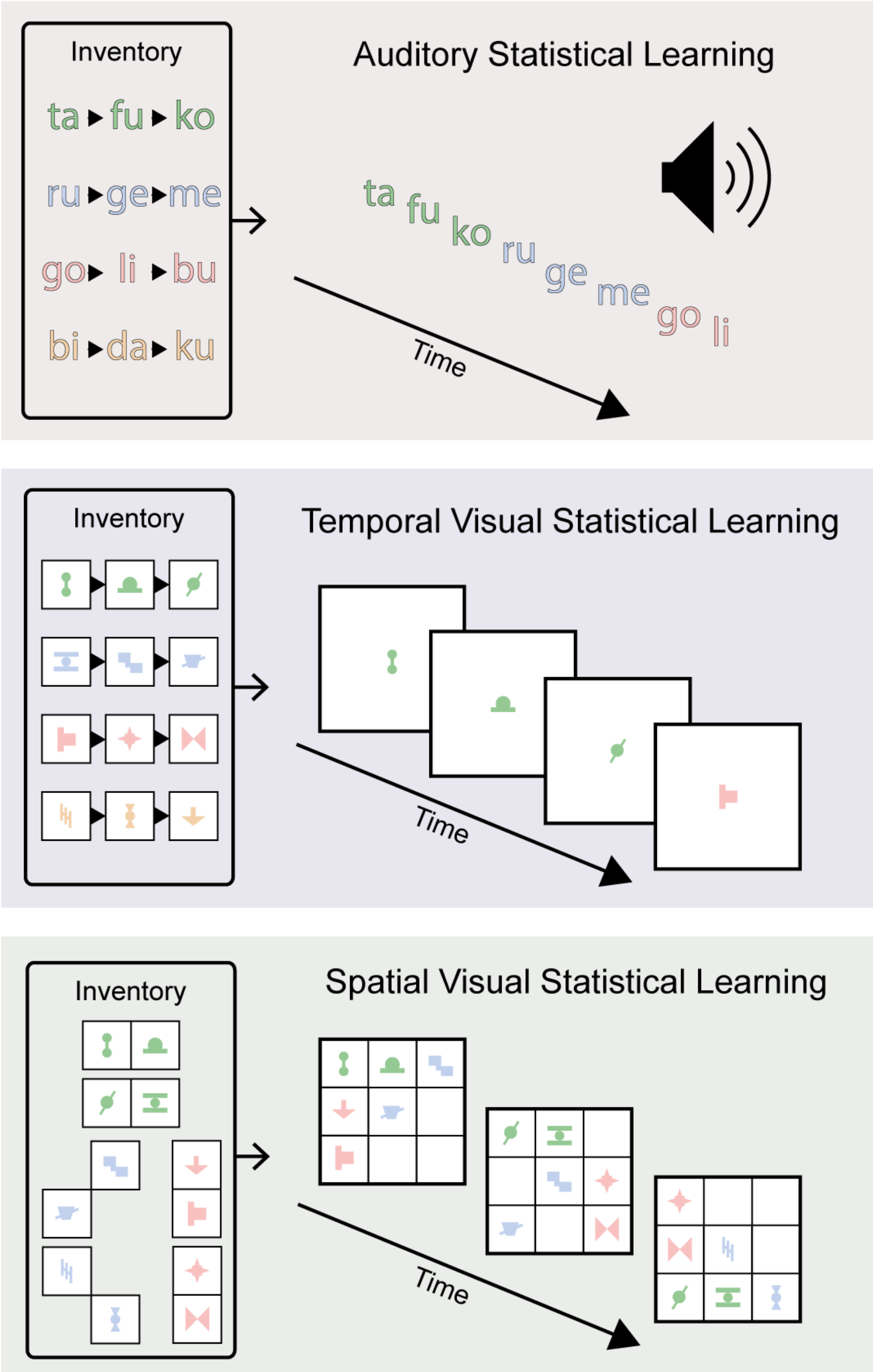

Figure 1.1 Statistical Learning Paradigms. **Top panel.** The original auditory statistical learning setup. An inventory of words is created by random combination of three syllables each. Participants are presented with a continuous stream of these words, with transitional probabilities as the only segmentation cue between words. M**iddle panel.** The temporal visual statistical learning setup. A direct translation of the auditory setup into the visual domain, where fixed combinations of shapes are presented visually. **Bottom panel**. The spatial visual statistical learning setup. This translates the original idea into the spatial domain, by creating scenes as combinations of fixed spatial shape pairs. There are again no segmentation cues between the chunks. All colors are only for illustration purposes. Participants see shapes as black-and-white.

1.1 General Definitions

1.1.1 Statistical Learning

Aslin and Newport (2012) defined *statistical learning* in the following way:

> […] a rapid and robust mechanism that enables adults and infants to extract patterns embedded in both language and visual domains. Statistical learning operates implicitly, without instruction, through mere exposure to a set of input stimuli (p. 170).

This definition of SL contains a number of key features that, in combination, distinguish SL from other forms of learning. First, SL is a form of *unsupervised learning* (Barlow, 1989) rather than *supervised* or *reinforcement learning* (Jordan & Mitchell, 2015), suggesting that it is based on discovering patterns in the input data rather than learning prespecified input-output mappings or reward contingencies. This is simultaneously in line with the vast majority of humans' ecologically relevant learning situations and in contrast with the bulk of psychological research on learning (Barlow, 1989; G. Hinton, 2014; G. E. Hinton, 2010). Second, SL is an implicit form of learning, suggesting that it builds knowledge without building awareness of the knowledge (Zoltán Dienes & Berry, 1997). This again contrasts with the bulk of research on learning, which focuses more on explicit learning (Reber, 1989). Third, SL discovers patterns in unsegmented input, where the item co-occurrence statistics reveal how the input can be segmented into its constituting elements. Again, this property sets SL apart from the usually highly segmented input used in learning research. Fourth, SL allows for rapid learning, where participants show long-lasting familiarization with an input stream after only a few minutes (Kim et al., 2009). Lastly, SL is not seen as modality-specific but operates at least on auditory (Saffran et al., 1996), visual (Fiser & Aslin, 2001), and haptic (Lengyel et al., 2019) input (Figure 1.1).

1.1.2 Structure Learning

The almost exclusive focus of research on statistical learning is learning specific reappearing patterns in unsegmented perceptual input. This can be contrasted with the learning of more abstract properties of the input, which can be achieved by abstracting from several observed or learned instances, usually studied for more segmented input under labels such as *rules* (Geambașu et al., 2023; Marcus et al., 1999), *concepts* (Bruner et al., 1956), *schemas* (Bartlett, 1932), and *grammars* (Reber, 1967). Which term is used depends on the specific experimental paradigm and carries different historical and empirical connotations. As an overarching term, I will use *structure learning* to refer to learning the abstract, latent structure underlying perceptual input. Therefore, for the purposes of this thesis, I use the term statistical learning to refer to the extraction of reappearing chunks from unsegmented input, while I use the term structure learning to refer to abstracting what is shared between multiple chunks. I consider this to be on a higher level of *abstraction* than statistical learning, as abstraction critically builds on the extraction of commonalities between several instances (Blackburn, 2008). In this case, the instances are the chunk representations built during statistical learning.

1.1.3 Explicit and Implicit: Learning, Tasks, and Instructions

The definition of statistical learning quoted in section 1.1.1 points out that statistical learning operates *implicitly*. However, what does this term and the related term *explicitly* mean? As this is a key concept in the SL literature at large, and for this dissertation specifically, in the current section, I will give an overview of what those terms can mean and how they are used here. As with many key terms in psychological research, the explicit-implicit distinction has not been used consistently to describe one specific aspect or phenomenon. There are at least three major ways in which these terms are used:

1. Talking about explicit and implicit *knowledge* or *representations*, these terms seem to be a replacement for the unfashionable terms conscious and unconscious, specifically in the sense of access consciousness (Block, 1995). In this sense, *implicit knowledge*

describes that a person has knowledge of something but is not aware of having it, suggesting that the utilization of this knowledge is not under rational control.

2. Talking about explicit and implicit *tasks* or *tests*, these terms describe if a participant in an experiment is aware of the fact that they have to apply previously acquired knowledge, i.e., they have to make an *explicit judgment* (Turk-Browne et al., 2005). It is, therefore, not referring to a person's knowledge about their mental states or representations but their knowledge about the experimental task at hand or how their knowledge relates to that task.
3. Talking about explicit or implicit instructions, in the context of SL research, these terms describe whether or not participants are told about the existence of regularities to be learned embedded in the input stream (also referred to as *intentional* vs. *incidental* conditions, (Arciuli et al., 2014)). It is, therefore, about what type of prior information participants receive about what is to be learned.

Unfortunately, researchers do not always clearly indicate which meaning of explicit/implicit they are referring to. To further complicate things, in some cases, these different meanings have been equated or confounded, assuming that an explicit task is a measure of only explicit knowledge and an implicit task is a measure of only implicit knowledge (Baker et al., 2004; Kim et al., 2009). It is not clear what theoretical conceptualization of explicit and implicit knowledge underlies this equation. For the purpose of this dissertation, I follow the approach of Dienes (Zoltán Dienes, 2007; Zoltán Dienes & Berry, 1997), building on the idea that implicit knowledge exists if a person knows something without being able to report that knowledge. This is operationalized by relating objective performance measures with subjective reports. The key idea is that if participants perform above chance on the objective measure while the subjective report does not indicate access to knowledge, implicit knowledge is present and has been used. For my studies, the subjective report utilized is open-ended questions at the end of the experiments, probing whether participants have any knowledge about the underlying structure of the used visual scenes. In this dissertation, I used the terms *participants with explicit knowledge* or *participants with explicit representations* to refer to participants that are able to report relevant structures in subjective reports. In contrast, I use the terms *participants with implicit knowledge* or *participants with implicit representations* to refer to participants that are not able to do so. In the spirit of readability, I also use the shortcuts *explicit participants* and *implicit participants* to refer to the same. When using the terms explicit/implicit in any other sense (points 2 and 3 from above) this is clearly pointed out.

1.2 Statistical Learning Based on Low-Level Statistics

The origins of research on statistical learning lie in developmental language research, demonstrating that infants can acquire basic building blocks of unsegmented linguistic input in an unsupervised learning setup (Saffran et al., 1996). It has quickly been extended to other domains, showing how adults and infants can similarly deal with unsegmented visual input (Fiser & Aslin, 2001, 2002b), termed visual statistical learning (VSL).

Early work in both auditory (Aslin et al., 1998; Saffran et al., 1996) and visual (Fiser & Aslin, 2001, 2002b) statistical learning demonstrated that infants and adults can learn both the *joint* and *conditional probabilities* between elements in the input. The conditional probability can be seen as a measure of predictability, demonstrating that statistical learning is more than merely counting item co-occurrence. Going beyond that, for the visual domain, it was also demonstrated that adult participants evade the challenge of the *combinatorial explosion*, which would make learning item-item statistics in situations with real-world complexity practically impossible, by parsimoniously learning a set of item chunks that is sufficient to explain the input (Fiser & Aslin, 2005; Orbán et al., 2008). Taken together, these studies demonstrate that

SL is more than simply keeping track of frequencies; instead, it builds representations that are as simple as possible while being able to explain the input based on low-level statistics.

A variety of computational models have been suggested to explain human statistical learning (Endress & Johnson, 2021; Mareschal & French, 2017; Orbán et al., 2008; Thiessen, 2017). Although they differ widely in grand computational principles and detailed implementation, they are all based exclusively on learning based on current low-level input statistics, ignoring potential abstract and hierarchical representations of the input. This can be understood in light of the fundamental debates in the statistical learning literature and the experimental paradigms providing the data to be modeled.

1.3 Statistical Learning Based on Higher-Level Biases

Experimental paradigms in statistical learning are usually designed to eliminate the influence of prior knowledge by using randomized combinations of arbitrary stimuli and carefully controlled procedures. The idea is that all that is left to learn are the item-item co-occurrence statistics, unbiased by existing associations. The initial efforts in the SL literature to exclude the effects of any prior knowledge on learning were based on debates in the linguistics literature on how much prior knowledge is necessary for language learning (Chomsky, 1959). In this sense, the original SL paradigm (Saffran et al., 1996) was purposefully extreme in its limitation of relevant prior knowledge to achieve a proof-of-concept demonstration. This was followed by numerous studies treating SL as a purely bottom-up process and manipulating various aspects of the input statistics to study the properties of this bottom-up information processing. Of course, this does not necessarily suggest a serious theoretical conviction that the process of statistical learning is not practically influenced by prior knowledge in many real-world situations.

Several recent studies directly demonstrated the effect of prior knowledge on statistical learning. It was shown that linguistic knowledge acquired over a lifetime influences what is learned in an SL paradigm (Gómez Varela et al., 2024; Stärk et al., 2022), that prior associations about stimuli used in an SL setup influence future processing and learning involving these stimuli (Antovich & Graf Estes, 2023; Chough & Zinszer, 2022; Kóbor et al., 2020), that facilitating learning of specific structures guides future learning (Zettersten et al., 2020), and that knowledge about the structure of the input provided explicitly can boost statistical learning under some conditions (Arciuli et al., 2014; Bertels et al., 2015). Taken together, these studies demonstrate different ways in which prior knowledge has an impact on statistical learning. This also highlights how a paradigm originally designed to exclude any effects of prior knowledge, can itself be a useful tool to study exactly these influences. By starting from such a design, researchers can systematically introduce different types of prior knowledge to study their effects in isolation.

1.4 Statistical Learning Based on Low-Level – High-Level Interactions

We can see that although SL paradigms were initially designed to exclude the influence of prior knowledge, statistical learning will interact with such knowledge if we leave the most narrowly controlled setups. In this dissertation, I take the position that what is commonly referred to as statistical learning is part of a larger, hierarchically structured, human unsupervised learning system, which is based on interactions of low-level co-occurrence statistics and higher-lever structural knowledge. The general idea here is that in this system, there are ongoing interactions between lower and higher levels of abstraction, where what is learned on the lower level is constrained by what is learned on the higher levels while simultaneously constraining what is learned on these higher levels. This realizes a reciprocal constraining of learning at different levels, intending to reach a hierarchical model of the process in the world that causes the sensory input. This idea was formulated before for unsupervised learning of more abstract conceptual knowledge using *hierarchical Bayesian models* (Kemp & Tenenbaum, 2008; Tenenbaum et al., 2011; Ullman et al., 2018) and is here applied to the unsupervised learning of patterns in perceptual input. The aim of this thesis is not to reapply these existing computational models to the experimental paradigms found in SL research. Instead, the goal is to empirically demonstrate that interactions across levels of abstraction arise in human unsupervised learning and identify some critical moderators that would not be obviously following from computational models. Specifically, I empirically show that phases of consolidation and the explicitness of knowledge are major moderators of unsupervised learning across levels of abstraction.

Empirically, this is realized by extending the existing spatial visual statistical learning paradigm to an unsupervised transfer learning paradigm. This way, we can see, under direct experimental control, how what was previously learned interacts with the acquisition of new representations and biases them in specific ways. In Chapter 2, I present a series of five experiments utilizing this paradigm to show that participants with explicit knowledge can abstract and generalize the types of structure underlying a statistical learning task immediately, while participants with implicit knowledge show a *structural novelty effect* in immediate transfer and application of knowledge. In Chapter 3, I present a series of three experiments that utilize the same principal paradigm but introduce phases of consolidation of different quality and duration. The results show that participants with implicit knowledge show abstraction and generalization after a phase of asleep but not awake consolidation. Overall, the findings of Chapters 2 and 3 demonstrate that unsupervised structure learning critically interacts with both the explicitness of knowledge and phases of consolidation.

1.5 Spatio-temporal Visual Statistical Learning

As discussed in section 1.1.1, the bulk of research in statistical learning is concerned with item-item co-occurrence statistics. While studies in the auditory domain naturally focus on the temporal statistics of elements in the input (usually syllables), in the visual domain studies focus either on temporal (Fiser & Aslin, 2002a; Kirkham et al., 2002) or spatial (Fiser & Aslin, 2001; Yu & Zhao, 2018) statistics of elements in the input (objects or abstract shapes) (see Figure 1.1). However, real-world visual input is not either spatial or temporal but always both. Spatial patterns unfold to the observer over time and the temporal order of spatial patterns is not arbitrary. Although it is useful to be able to separate spatial and temporal regularities and study them independently, the interaction of the two during natural vision can only be understood by studying them together. Therefore, if the goal is to understand how the human unsupervised learning system can deal with real-world input, it is insufficient to study the learning of regularities in these two dimensions in isolation. To address this, I developed a novel *spatio-temporal VSL* paradigm in which spatial patterns are presented over time. In Chapter 4, I present a series of five experiments that demonstrate that implicit learning is possible in a spatio-temporal setup and that temporal regularities are used in the implicit learning of spatial patterns. This therefore supports both the feasibility and the necessity of such an approach in VSL research. In Chapter 5, this new spatio-temporal VSL setup is connected to the previous chapters by demonstrating that higher-level spatial and temporal features of the input, transcending low-level co-occurrence statistics, are used in implicit learning, and that there are qualitative differences in how explicit and implicit knowledge can be utilized. Specifically, in two experiments I show that features such as an overall perceived motion direction or overall orientation of shape arrangement bias participants beliefs about what types of structure are present. This complements the findings of Chapters 2 and 3, by showing that biases about types of structures cannot only be abstracted from prior learning of specific patterns, but can also be induced by higher-level features of the input. In an additional experiment I show that while participants with explicit knowledge can use their knowledge for a prediction task, participants with implicit knowledge cannot. This again complements findings from previous chapters, demonstrating differences in the utilization of explicit and implicit knowledge.

1.6 Summary of Results

Overall, the findings presented in this dissertation paint a picture of VSL as a combination of lower-level spatio-temporal co-occurrence statistics (Chapter 4, Experiments 4a-b, 5a-c) and higher-level top-down biases (Chapters 2 and 3, Experiments 1a-c, 2a-b, 3a-b), where the multitude of available features across levels are flexibly combined to achieve a congruent and comprehensive, yet parsimonious interpretation of the input (Chapter 5, Experiments 6a-b). I identified three different types of higher-level biases: pre-existing biases that are independent of properties of the experiment (Chapter 2, Experiment 2b), biases formed based on the abstraction of learned low-level statistics (Chapters 2 and 3, Experiments 1a-c, 2a-b, 3a-b), and biases based on observed higher-level features of the input (Chapter 5, Experiments 6a-b). Additionally, two critical moderators of this hierarchical learning system are identified: explicitness and consolidation. The explicitness of knowledge crucially influences how it can be applied to new input for generalization and prediction (Chapters 2 and 5, Experiments 1a-c, 2a-b, 7). Consolidation of knowledge enables generalization of the structure of the input even in the absence of explicitness (Chapter 3, Experiment 3a, 3c). Taking all these findings together, this dissertation demonstrates that the narrow limitation and control that enabled the initial success of SL research need to be carefully and incrementally overcome to understand the role of SL in the overall human cognitive system. It does so by introducing two new VSL paradigms that enable novel, systematic ways of investigating the human unsupervised learning system.

CHAPTER 2

Structural Knowledge in Visual Statistical Learning

The study presented in this chapter relates what is usually called statistical learning to the learning of more abstract, structural features and transferring such structural knowledge to novel learning or decision-making. I present a series of five experiments utilizing a transfer learning version of the classical statistical learning paradigm to show that participants with explicit knowledge can abstract and generalize the types of structure underlying a statistical learning task immediately, while participants with implicit knowledge show a structural novelty effect in immediate transfer and utilization of knowledge.

2.1 Abstraction and Structural Transfer

Humans store and use knowledge about the world at different levels of abstraction. For example, we can have specific knowledge about a particular table that we have encountered before and remember its approximate height, material, and other features. On the other hand, we can also have conceptual or categorical knowledge about tables in general; the range of measurements, materials, and configurations of parts of the objects we would consider tables. It has been empirically demonstrated that visual statistical learning can operate either at the level of individual items or at the categorical level (Jun & Chong, 2016, 2018; Jung et al., 2021; Rogers et al., 2021; Sherman et al., 2023), and perceptual and contextual factors have been identified that influence which regularities are preferably learned (Aslin & Newport, 2012; Emberson & Rubinstein, 2016). Although these investigations are useful in their own right, this compartmentalized approach isolates the learning at different levels of abstraction and conceptualizes them as independent tasks, even though acknowledging that they are realized via the same statistical learning mechanism. However, since abstraction critically builds on the extraction of commonalities between several instances (Blackburn, 2008), a more fruitful approach might be to ask how learning the statistical regularities of specific objects leads to the abstraction of their shared underlying structure to form conceptual/structural knowledge. In this chapter, I will use the term statistical learning to refer to the extraction of re-appearing chunks from unsegmented input, while by structure learning, I will refer to the process of abstracting what is shared between multiple chunks. The term structural transfer then refers to generalizing what has been abstracted to novel input.

There exists a vast literature tackling abstract learning of underlying structures such as *rules* (Geambașu et al., 2023; Marcus et al., 1999), *concepts* (Ashby & Maddox, 2011; Bruner et al., 1956), *schemas* (Bartlett, 1932; Gilboa & Marlatte, 2017), and *grammars* (Pothos, 2007; Reber, 1967). However, due to their focus, such studies usually start from a point where the major challenge of statistical learning is already assumed to be completed: they work with either familiar or at least highly segmented input, in which the basic units from which a concept needs to be abstracted are readily available. In contrast, participants in typical statistical learning studies need to find patterns without available segmentation cues, but these patterns usually do not have more abstract regularities over multiple different patterns. However, these two types of learning cannot be isolated from each other as the ongoing interaction with the real world is iterative and does not adhere to encapsulated phases where the different processes work in isolation, as in psychological experiments. The key idea underlying the work presented in this chapter is that since these two mechanisms cannot be isolated in ecological settings and they always co-function, understanding them also requires understanding how they interact. Therefore, the overall objective of this chapter is to empirically investigate unsupervised learning that starts from unsegmented input and spans multiple levels of abstraction.

As previewed in the Introduction (1.1.3), the last decades saw remarkable progress in theoretical and computational models that can build representational hierarchies at increasing levels of abstraction without the need for supervision (Heald et al., 2021; Kemp & Tenenbaum, 2008; Lake et al., 2015; Zhuang et al., 2021). Much of the advances in our computational understanding of the unsupervised learning of hierarchies of abstraction are based on a class of computational models called hierarchical Bayesian models (Kemp & Tenenbaum, 2008), which critically builds on the interaction and mutual constraining of representations over different levels of abstraction. This means that whatever is learned at the lower levels will constrain what can be learned on the higher level, and reciprocally and simultaneously, it is also constrained by acquired knowledge at the higher level. The levels of the hierarchy interact in order to reach an overall best-fitting hierarchical description of the process generating the input (technically, a best-fitting probability distribution over parameters of the hierarchical model). This view establishes a natural connection between themes introduced in the current and the previous chapter as such hierarchical Bayesian models realize an iterative formation of representation, which uses the existing prior knowledge, realize abstraction by discovering shared structure, and in turn, uses discovered structure to constrain learning of specific instances.

Despite these computational advances, the situation is much more mixed on the experimental side. Several lines of research focus on learning hierarchically structured representations based on explicit supervision or reinforcement and describe fundamental features of such learning (Behrens et al., 2018; Eckstein & Collins, 2020; Lewis & Durrant, 2011). Markedly less empirical knowledge has been accumulated about whether and how humans and animals develop hierarchical internal representations without explicit guidance, even though such learning constitutes the vast majority of knowledge acquisition in natural settings (Barlow, 1989). It is precisely this gap that the current study starts to address.

Based on previous research in various domains, I hypothesized that explicitness might be an important moderator in the process described above. First, statistical learning is usually defined as a form of implicit learning leading to implicit knowledge (Aslin & Newport, 2012), although it was demonstrated that during typical SL setups, explicit knowledge can arise (Bertels et al., 2012; Goujon et al., 2014; Kim et al., 2009; H. Liu et al., 2023). Conversely, learning more abstract features has been routinely studied under assumptions of explicit knowledge but is also demonstrated for implicit knowledge (Reber, 1967). Second, it has previously been shown that explicit and implicit forms of learning can follow different rules in various domains (Ball et al., 2021; Bloch et al., 2016; Robertson et al., 2004; Yang & Li, 2012). Third, although explicitness of knowledge is conceptually broader than just the ability to verbalize knowledge, the two are related (Dienes & Perner, 1999), and verbalizability is my primary measure of explicitness in this study. A large body of research has argued for a role of language-based processing for abstraction (Connell, 2018; Davis & Yee, 2018; Jiang et al., 2019; Lupyan & Lewis, 2017; Sloutsky & Deng, 2017). Taken together, these findings suggest that the explicitness of knowledge might be an important moderator of unsupervised learning over the levels of abstraction investigated in this chapter. Therefore, the analyses for all experiments in this study focus on the interaction of abstraction, as demonstrated by structural transfer, and the explicitness of knowledge, as demonstrated by the ability to verbalize knowledge. In order to experimentally combine statistical learning with the learning of more abstract features, I extended the classic spatial visual statistical learning setup (Fiser & Aslin, 2001) to an unsupervised transfer learning paradigm. In this chapter, I present a series of five experiments utilizing this paradigm to show that participants with explicit knowledge can abstract and generalize the types of structure underlying a statistical learning task immediately, while participants with implicit knowledge show a structural novelty effect in immediate transfer (2.2) and application (2.3) of knowledge.

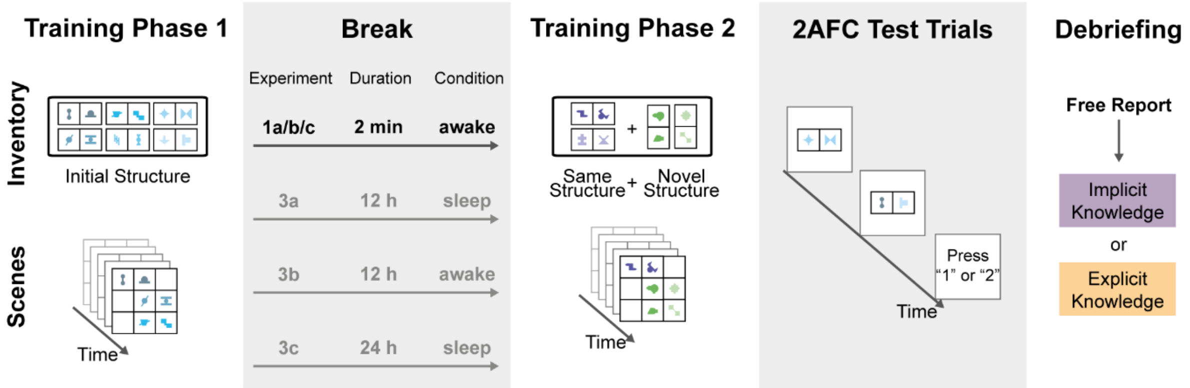

Figure 2.1 Statistical Learning Paradigm - 1 **Training Phase 1** The first training phase follows the classical spatial VSL paradigm, presenting scenes made from shape pairs without segmentation cues. All colours are only for illustration; for participants, everything is black and white. For this learning phase, all pairs have the same orientation (horizontal or vertical), counterbalanced between participants. **Break** The break after the first training phase varies between Experiments as described in the figure. **Training Phase 2** The second training phase consists of novel scenes, made from novel shapes. In this phase all participants see horizontal and vertical pairs. **2AFC Test Trials** In all 2AFC test trials, participants are presented with a real pair from the training phases and a foil pair made by combining shapes of two real pairs. They need to decide which of the two shape pairs seems more familiar. **Debriefing** After the experiment participants answer a set of open questions which are used to assess whether they have explicit knowledge of the presence of pairs in the input.

2.2 Learning Based on Induced Biases

2.2.1 Experiment 1a: The Structural Novelty Effect in Implicit Transfer Learning

Experiment 1a was designed to deliver a first proof of concept that the shared structure underlying several chunks learned during classical spatial visual statistical learning influences the subsequent learning of new chunks. The experimental setup builds on the standard spatial VSL paradigm (Fiser & Aslin, 2001), extending it to a transfer learning paradigm (Figure 2.1). This paradigm is the ideal candidate for investigating the unsupervised learning of higher-order structures as it uses unsegmented input and investigates the learning of specific patterns of fixed combinations of shapes (*chunks*), which can easily be created from different underlying structures. The underlying structure used in the current series of experiments is the orientation of shape pairs, i.e., whether the shapes of a pair are arranged horizontally or vertically. Critically, due to the nature of the spatial VSL setup, the orientation is not an obvious property of the scenes per se and cannot be directly observed as there are no segmentation cues between chunks. The orientation is a property of the association between shapes and can, therefore, only be learned in conjunction with learning these associations. This transfer learning version of the VSL paradigm goes beyond simply asking if a previously used structure is recognized and instead measures how exposure to one type of structure differentially influences the acquisition of multiple types of structures later on. By also probing the explicitness of participants' knowledge, we can then look at the interaction of explicitness of knowledge and the type of structures learned.

Participants

251 participants (92 female, mean age = 28.0, SD = 9.5) were recruited via prolific.co. The hourly compensation was £ 6.3. All participants had normal or corrected-to-normal vision. The sample size was chosen to achieve 80% power for expected small to medium effect sizes (d = 0.4) in paired t-tests, accounting for a high number of multiple comparisons (therefore: α = .005). The power analysis conducted with the *pwr* R package (Champely et al., 2017) suggested a needed sample of 168.38, which I generously bolstered to account for expected high exclusions for an online study and the subset of explicit participants. 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 were taken from Fiser and Aslin (2001) and consisted of 20 abstract black shapes on a white background (see Figure 2.1). The shapes were grouped to form six pairs of the same orientation (horizontal or vertical; randomly assigned to participants) for the first learning phase and four pairs, two horizontal and two vertical, for the second learning phase. The assignment of shapes to pairs was randomized for each participant, leading to superficially different scenes for each participant. Scenes were created by placing three pairs together on a 3x3 grid without segmentation cues. 160 scenes were created for the first and 48 for the second learning phase. In the second learning phase, each scene was used twice for a total of 96 presented scenes.

As this was an online study, participants conducted it on their own computers using Google Chrome, Safari, or Opera browser. Only desktop and laptop computers were admissible, and no smartphones or tablets. Stimuli were presented using custom JavaScript code built on the *jsPsych* library (version 6.1.0) (Leeuw, 2015). As participants used different devices (screen size and resolution), the visual angle of the shapes was not the exact same for all participants. Instead, the 3x3 grid extended over 600x600 pixels and was centered in the middle of the screen. The remaining screen outside the grid was empty (white).

Procedure

Participants passively observed 160 scenes in the first training phase. For half of the participants, these scenes contained only horizontal pairs (horizontal condition), and for the other half, they contained only vertical pairs (vertical condition). Note that from the makeup of the scene, it was impossible to distinguish whether a scene was made up of horizontal or vertical pairs. Each scene was presented for 2 s with a 1 s interstimulus interval (ISI). After a two-minute passive break, participants passively observed 96 scenes in the second training phase. Participants were not told about the presence of any structure in the scenes and were instructed to simply be attentive so that they could later answer simple questions. After half of each training phase, an attention check appeared, asking participants to press the spacebar to continue. Response time for the attention check was recorded to detect inattentive participants. After the second training phase, participants had another two-minute passive break. Following this, pair learning was tested with a two-alternative forced choice task (2AFC). In each trial, participants saw a real shape pair from one of the training phases and a foil pair created by combining shapes from two different pairs of the same training phase. Overall, all real and foil pairs were used the same number of times during the test phase to ensure no learning effects within the test phase. Real and foil pairs were presented after each other in the 3x3 grid for 2 s with a 1 s ISI. The order of real and foil pairs was randomized. Participants were asked to indicate which of the two pairs was more familiar by pressing "1" or "2" on their keyboard. Participants first completed 16 trials using pairs from the second training phase, followed by 24 trials using pairs from the first training phase. Finally, participants answered five open questions about their beliefs about the experiment and their knowledge of pair structure (see Appendix A for the detailed questionnaire).

Results

Based on pilot data, I chose 20 seconds combined response time for both attention checks as the cut-off value for inclusion. 19 participants were rejected for failing this criterion. *Response bias* was defined as the proportion with which participants used one of the two response options ("1" and "2"), and participants who were 2.5 SD away from the mean were excluded. 3 participants were excluded for failing this criterion. This left us with 229 participants after exclusions. Based on the open responses at the end of the experiment, participants were categorized into one of three groups. Participants who reported no knowledge of pairs were counted as implicit (n = 192), participants who reported knowledge of the presence of pairs were counted as explicit (n = 34), and participants who also reported the underlying horizontal/vertical structure were excluded from analysis (n = 3) as they were too few for meaningful analysis (whether or not these 3 particiapnts were included in the group of explicit participants did not significantly change the results).

Participants' responses were summarized in three scores. Performance in the first training phase is the proportion correct for trials using pairs of that phase. Performance for the *same structure* pairs is the proportion correct for trials of the second training phase using pairs with the same orientation as the pairs of the first training phase (this can be horizontal or vertical, depending on the assigned condition). Performance for *novel structure* pairs is the proportion correct for trials of the second training phase using pairs with a different orientation than the pairs of the first training phase. Explorative analysis of the trials of the second training phase revealed a strong negative correlation between trials of the novel and the *same structure* (r = -.432, p < .001). There are two potential explanations for this. First, individual participants only learn pairs of one or the other structure. Second, as the foil pairs used for one type of structure were created by recombining shapes of pairs of the other structure, this could be a type of *consistency effect* where participants tend to choose the same shapes, independent of pair knowledge. If the second interpretation is correct, we expect to see an increase in this negative correlation as participants complete more test trials. Indeed, we find that for the first eight trials, this correlation is r = -.209., while it increases to r = -.435 in the second 8 trials. Based on this exploration, I included only the first eight trials in all the following analyses to minimize this consistency effect. The following experiments use the same number of test trials to keep comparability, but again, I only analyze the first 8.

The data was collapsed over vertical and horizontal conditions for all further analysis, as a 3x2 mixed ANOVA with *test type* (levels: *training phase 1*, *same structure*, *novel structure*) as within-subject factor and *condition* (levels: horizontal, vertical) as between-subject factor showed no significant main effect of *condition* (F(1, 224) = 0.776, p = .379, BF = 0.08, η_p^2 = .003) and no significant *test type* - *condition* interaction (F(2, 448) = 0.199, p = .820, BF = 0.04, η_p^2 = .001). *Bayes Factors* (BF) reported for the ANOVA and t-tests here and throughout the dissertation are based on the *BayesFactor* R package, realizing Bayesian tests with models analogous to the frequentist counterpart, and employing a JZS (Jeffrezs, Zellner, Siow) prior, unless specified otherwise (Rouder et al., 2012). P-values reported throughout the text were subject to experiment-wise correction for multiple comparisons using the Holm-Bonferroni method (Holm, 1979).

The results (Figure 2.2 top panel) show above chance performance for the first training phase for both explicit (M = 67.9, SE = 4.6, d = 0.67, t(33) = 3.93, p = .002, BF = 70.8) and implicit (M = 55.0, SE = 1.1, d = 0.32, t(191) = 4.46, p < .001, BF = 919) participants. However, the performance was significantly higher for explicit participants (d = 0.73, t(224) = 2.75, p = .009, BF = 195). For the second learning phase, implicit participants show above chance performance for pairs of the *novel structure* (M = 57.8, SE = 1.9, d = 0.30, t(191) = 4.16, p < .001, BF = 281), and chance performance for pairs of the *same structure* as before (M = 49.1, SE = 2.1, d = 0.03, t(191) = -0.44, p = .659, BF = 0.09). The performance for these types of pairs is significantly different from each other (d = 0.20, t(382) = 2.78, p = .030, BF = 3.38). Explicit participants show the opposite pattern with above-chance performance for pairs of the *same structure* as before (M = 67.6, SE = 6.3, d = 0.48, t(33) = 2.81, p = .033, BF = 5.0), and chance performance for pairs of the *novel structure* (M = 58.1, SE = 5.6, d = 0.25, t(33) = 1.46, p = .465, BF = 0.48). However, for them, this difference is not significant (d = 0.19, t(33) = 1.12, p = .539, BF = 0.33). We do, however, see a significant, medium to large correlation between learning in the first training phase and learning pairs of the *same structure* in the second training phase (r = .45, p = .008) for the explicit participants. Such a correlation is not observed between learning in the first learning phase and learning pairs of the *novel structure* in the second learning phase (r = -.01, p = .947). The different patterns of transfer behavior for explicit and implicit participants are also confirmed when entering the data into a 2x2 mixed ANOVA with factors *participant type* (explicit or implicit) and *structure type* (novel or same). The results show a significant main effect of *participant type* (F(1, 224) = 8.03, p = .005, BF = 2.1, η_p^2 = .03)) and of *structure type* ($F(1, 224) = 4.05$, $p = .045$, $BF = 1.2$, $\eta_p^2 = .02$)) as well as a significant interaction of both ($F(1, 224) = 4.89$, $p = .028$, $BF = 31.6$, $\eta_p^2 = .02$).

To test for a possible time-of-day effect (Tandoc et al., 2021) in learning or generalization, I correlated test performance with the hour of the day at which participants completed the experiment. There were no significant correlations for pairs of the *same structure* (explicit participants: $r = -.03$, $p = .882$; implicit participants: $r = .01$, $p = .893$) or pairs of the *novel structure* (explicit participants: $r = .13$, $p = .477$; implicit participants: $r = -.05$, $p = .458$). Additionally, I looked separately at groups of participants completing the experiment early in the day (7-11 am) and late in the day (7-11 pm). For implicit participants, there was no significant difference between participants that participated early (n = 23) or late (n = 22) as a 2x2 mixed ANOVA with *hour-of-day* and *test type* as factors showed no significant main effect of *hour-of-day* ($F(1, 43) = 0.019$, $p = .892$, $BF = 0.26$) and no significant *hour-of-day – test type* interaction ($F(1, 43) = 0.095$, $p = .759$, $BF = 0.31$).

As reported above, explicit participants show higher average learning in the first learning phase, which could be what enables the generalization of the learned structure. To test this idea, I conducted a matched sample analysis (Ho et al., 2007). The general idea of this analysis is to create a sub-sample of the implicit participants that perform like the explicit participants for the pre-training trials (see Appendix C for details). The question, then, is how this sub-sample performs for the same and novel structure trials. In a first step, I ran six applicable matching algorithms implemented in the *MatchIt* R package (Ho et al., 2011). The six created matched implicit samples were then compared to the original explicit sample according to four metrics: *standardized mean difference*, *variance ratio*, *mean of the empirical cumulative density function*, *and maximum of the empirical cumulative density function*. All comparisons can be seen in Supplementary Table 1 in Appendix C. The overall best fit was the *nearest neighbor matching with replacement* using *propensity scores*. I found that this sample showed a similar pattern for learning in the second training phase as the original full sample (see Figure 2.2 bottom panel). This is captured by a 2x2 ANOVA using the *novel* and *same structure* pairs for the original explicit and the matched implicit data showing a significant interaction ($F(1, 87) = 8.53$, $p = .004$, $BF = 10.7$, $\eta_p^2 = .09$) and post-hoc comparisons show a significant difference between *novel* and *same structure* trials for the synthetic implicit data ($p = .012$; $BF = 3.6$). This analysis suggests that the difference between the two groups is not merely based on different strengths of learning in the first training phase.

Discussion

The results of Experiment 1a show that the structure underlying multiple chunks acquired during unsupervised learning biases subsequent learning. Critically, this structural bias points in opposite directions for explicit and implicit participants. Participants with explicit knowledge show *structural transfer*; they are able to generalize and, therefore, learn more new pairs of the same structure as before. However, participants with implicit knowledge show a *structural novelty effect*, learning more new pairs of a structure different from before. Importantly, using a matched sample analysis, it was shown that the differences in transfer behavior between explicit and implicit learners could not be explained by the explicit learners' overall higher learning outcomes in the first training phase. The type of transfer behavior is not predicted by the quantity of knowledge but by the quality of the knowledge - its explicitness.

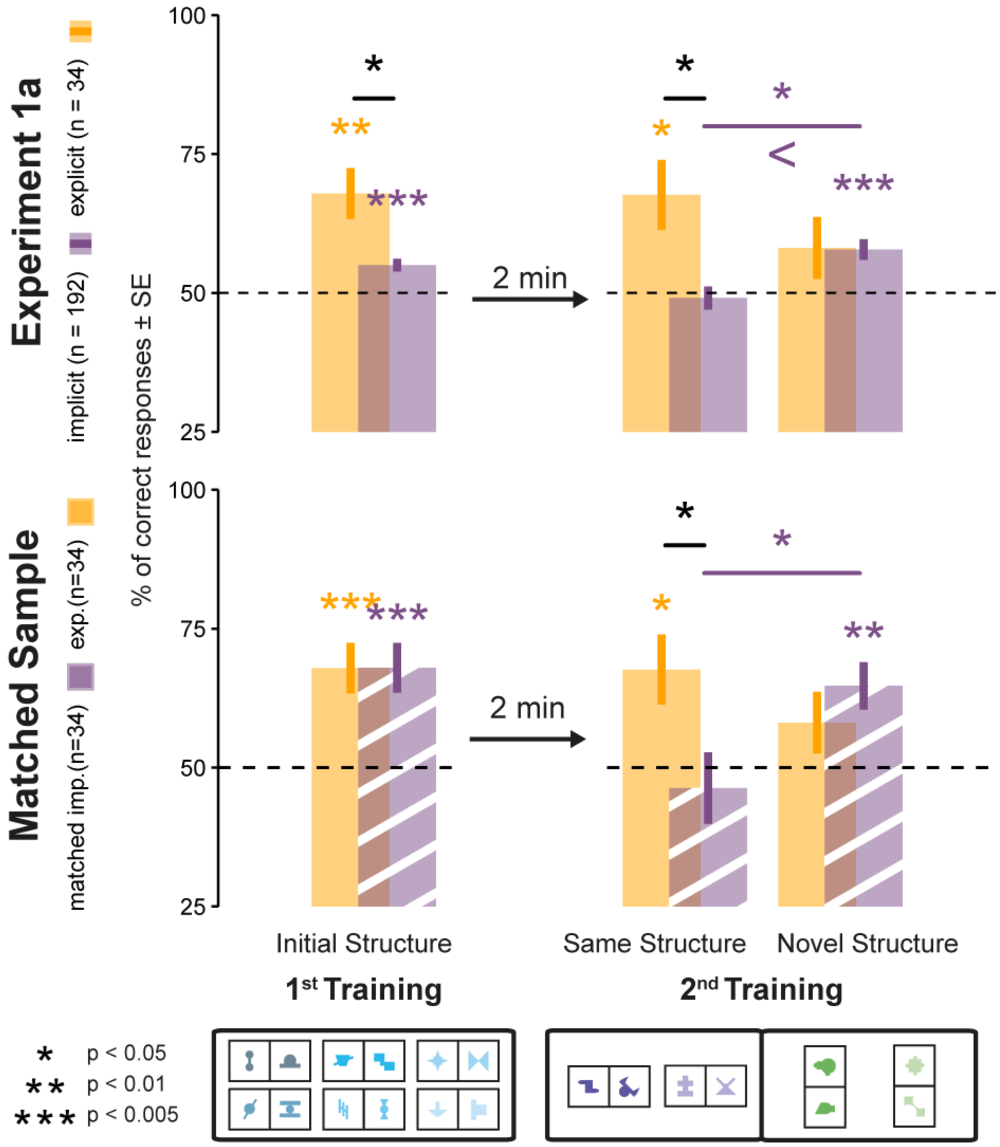

Figure 2.2 Experiment 1a Results. Results of 2AFC familiarity tests in Experiment 1a. Test trials are grouped along the x-axis according to which training phase they appeared in, and for the second training phase, according to whether they follow the same or a different structure (horizontal or vertical) than the pairs of phase one. The y-axis represents the proportion of correct responses in the 2AFC test trials. Bars represent the standard error. Color coding indicates implicit and explicit subgroups of the participants. The horizontal dotted line at 50% denotes chance performance. Asterisks above bars denote significance levels from chance, while above lines, significance level comparing two conditions below the tips of the line. The legend of significance levels is shown in the lower left corner. **Top Panel** Performance in Experiment 1a. **Bottom Panel** Performance for matched sample in Experiment 1a (see methods).

2.2.2 Experiment 1b: Replication of the Structural Novelty Effect

As the structural novelty effect found in Experiment 1a is novel and surprising, I conducted a close conceptual replication in order to ensure the robustness of the finding. Experiment 1b (Figure 2.3 top panel) follows the procedure of Experiment 1a but uses two types of diagonal pairs instead of horizontal and vertical pairs. These two types of diagonal pairs are orthogonal to each other, leading to the same logic of *same* and *novel structure* pairs as in Experiment 1a.

Participants

243 participants (128 female, mean age = 30.1, SD = 11.7) were recruited via prolific.co. The hourly compensation was £ 6.3. All participants had normal or corrected-to-normal vision. The sample size was chosen to match that of Experiment 1a. The study was approved by the Psychological Research Ethics Board of the Central European University, and all participants provided informed consent.

Materials

This experiment used the same materials as Experiment 1a. However, the shapes were not grouped into horizontal and vertical pairs but into two orthogonal groups of oblique pairs. For one group of pairs, the second shape was always in the top right grid cell from the first shape; for the other group of pairs, the second shape was always in the top left grid cell. As such pairs allow for fewer unique combinations within a 3x3 grid, the scenes for this Experiment were set in a 5x5 grid. The shapes mainly occupied the central 3x3 sub-grid, with one shape per scene being in the outer cells. All shapes appeared in the outer cells an equal number of times over all scenes. In order to ensure that the whole grid is visible on the participants' devices, the size was not set to a fixed pixel value. Instead, the size of one grid cell was set to 1/7 of the pixel height of the participant's screen. Therefore, the whole grid filled 5/7 of the screen height.

Procedure

The procedure was identical to Experiment 1a.

Results

The exclusion criteria were identical to those used in Experiment 1a. This led to 15 exclusions for failed attention checks and four exclusions for response bias. This left us with 224 participants after exclusions. Participants were categorized as explicit or implicit in the same way as in Experiment 1a. This led to 12 participants being categorized as explicit and 212 as implicit. Bayes Factors from Bayesian t-tests for implicit participants reported for Experiment 1b used an r-scale parameter of .5 instead of the default √2/2, therefore changing the prior to reflect that Experiment 1a found small effect sizes for this group.

Overall, the data shows the same pattern as in Experiment 1a. The results for the implicit participants (n = 212) closely follow the results of Experiment 1a. They perform above chance for pairs of the first training phase (M = 53.9, SE = 0.8, d = 0.34, t(211) = 4.99, p < .001, $BF = 1.0*10^4$) and for pairs of a *novel structure* (M = 57.1, SE = 1.7, d = 0.28, t(211) = 4.06, p < .001, BF = 245) but not pairs of the *same structure* (M = 50.1, SE = 1.8, d = 0.03, t(211) = 0.46, p = .999, BF = 0.12) in the second training phase. The performance for pairs of the *same* and *novel structure* is again significantly different (d = 0.18, t(211) = 2.67, p = .049, BF = 3.2). The results for explicit participants (n = 12) show the same qualitative pattern as in Experiment 1a, however, without reaching a significant difference from chance: first training phase(M = 59.4, SE = 4.8, d = 0.56, t(11) = 1.95, p = .387, BF = 1.2), *novel structure* (M = 52.1, SE = 10.4, d = 0.06, t(11) = 0.20, p = .999, BF = 0.29), *same structure* (M = 60.4, SE = 7.8, d = 0.39, t(11) = 1.33, p = .084, BF = 0.59). This can be explained by the significantly smaller number of participants acquiring explicitness in Experiment 1b compared to 1a (X^2 = 10.47, df = 1, p = .001, BF = 37.0), resulting in diminished power for these tests.

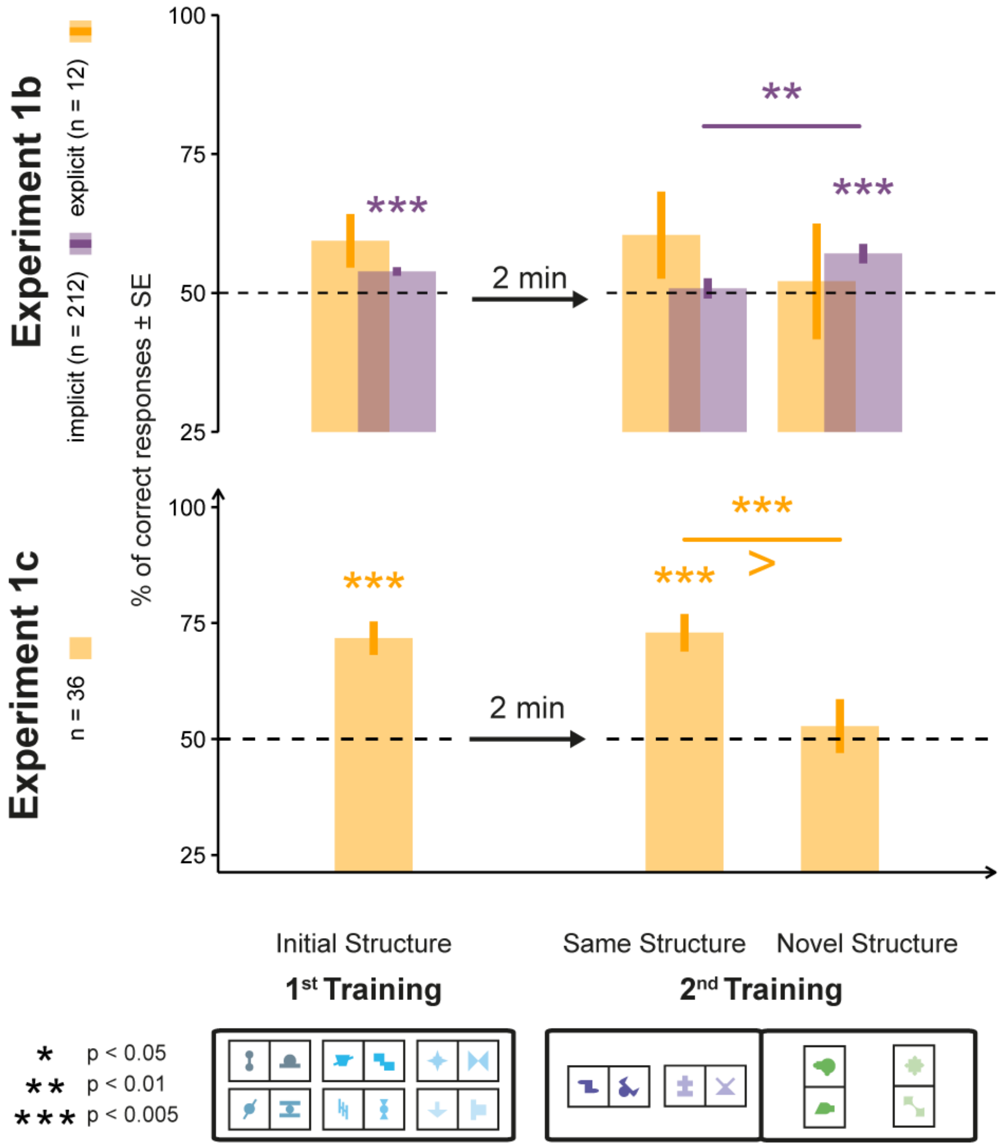

Figure 2.3 Experiments 1b and 1c Results. Results of 2AFC familiarity tests in Experiment 1a. Test trials are grouped along the x-axis according to which training phase they appeared in, and for the second training phase, according to whether they follow the same or a different structure (horizontal or vertical) than the pairs of phase one. The y-axis represents the proportion of correct responses in the 2AFC test trials. Bars represent the standard error. Color coding indicates implicit and explicit subgroups of the participants. The horizontal dotted line at 50% denotes chance performance. The asterisks above bars denote significance levels from chance, while above lines they denote significance levels comparing two conditions. The legend of significance levels is shown in the lower left corner. **Top Panel** Performance in Experiment 1b. **Bottom Panel** Performance in Experiment 1c.

Discussion

Experiment 1b replicated the structural novelty effect found for implicit participants in Experiment 1a, therefore demonstrating the robustness of that finding. The pattern for explicit participants could only descriptively be replicated, as the low rate of explicitness led to underpowered tests.

2.2.3 Experiment 1c: Investigating the Role of Explicitness in Structural Transfer

The previously reported explicit-implicit distinction was quasi-experimental rather than true experimental since the groups were formed naturally. Therefore, it is unclear whether it is indeed explicitness that enables generalization or whether the two groups of participants differ in other important ways (e.g., task engagement or attentional processes). Experiment 1c was designed to answer this question by inducing explicitness and, therefore, testing if everyone can, in principle, show the type of behavior that participants who happened to attain explicit knowledge in Experiment 1a showed.

Participants

40 participants (18 female, mean age = 28.4, SD = 11.8) were recruited via prolific.co. All participants had normal or corrected-to-normal vision. The sample size was chosen to approximately match the number of explicit participants in Experiment 1a after an expected exclusion rate of 10%. The hourly compensation was £ 6.3. 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 identical to Experiment 1a.

Procedure

The procedure was identical to Experiment 1a, apart from the instructions. In this experiment, participants were told about the pair structure before the beginning of the experiment. Participants were told that whenever a specific shape appears in the grid a second specific shape will appear in a fixed position near it.

Results

The exclusion criteria were identical to those used in Experiment 1a. This led to 4 exclusions for failing the attention checks and 0 exclusions for response bias. This left us with 36 participants after exclusions.

Overall, the results (Figure 2.3 bottom panel) show the same pattern as the explicit participants in Experiment 1a, with above-chance performance for pairs of the first training phase ($M = 71.8$, $SE = 3.6$, $d = 1.01$, $t(35) = 6.07$, $p < .001$, $BF = 2.6*10^4$) and for pairs of the *same structure* ($M = 72.9$, $SE = 4.0$, $d = 0.95$, $\mathrm{i}(35) = 5.69$, $p < .001$, $BF = 9{,}058$) but not pairs of a *novel structure* ($M = 52.8$, $SE = 5.8$, $d = 0.08$, $t(35) = .48$, $p = .634$, $BF = 0.20$) in the second training phase. The performance for the *same* and *novel structure* pairs is significantly different ($d = 0.64$, $t(35) = 3.83$, $p = .001$, $BF = 58.0$).

Discussion

The results of Experiment 1c are very similar to those of participants with explicit knowledge in Experiment 1a. Experiment 1c, therefore, demonstrated that verbal instructions can easily induce the type of explicitness studied here and that it is, indeed, this explicit knowledge that drives the observed pattern of generalization.

2.3 Direct Utilization of Induced Biases

All experiments reported so far are based on transferring structural knowledge from one learning phase to a second one. An open question is how participants behave when they need to apply the acquired structural knowledge directly to test trials instead of to new unsegmented input. That is, how can they apply their structural knowledge in decision-making rather than learning? Experiment 2a tested this by using only the pre-training phase of Experiment 1a, biasing participants with either vertical or only horizontal pairs and following it up with a series of new test trials, probing different aspects of the participants' representations. Critically, I tested participants' knowledge of the orientation of the learned pairs and their willingness to transfer orientation to novel pairs. However, this was not transfer to learning novel scenes, as in Experiment 1a, but to direct utilization in test trials providing horizontal and vertical choices.

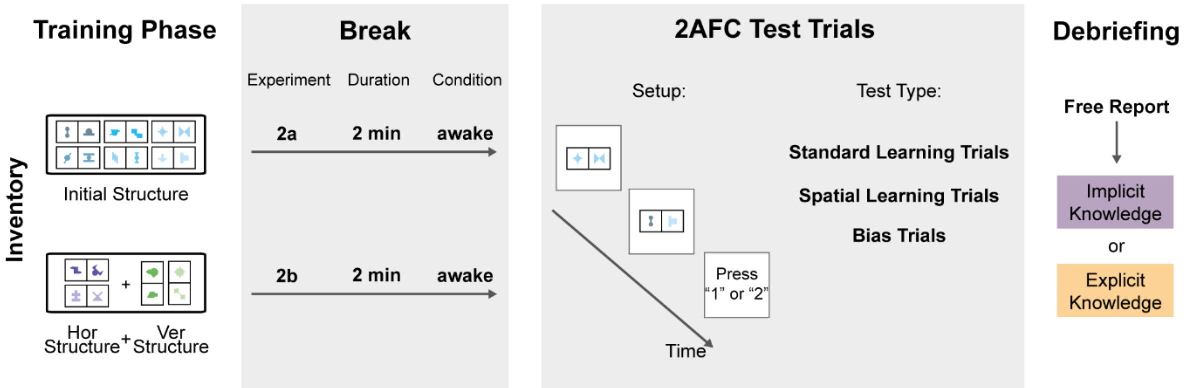

Figure 2.4 Direct Utilization of Induced Biases in Statistical Learning. **Training Phase** The training phase follows the classical spatial VSL paradigm, presenting scenes made from shape pairs without segmentation cues. All colors are only for illustration; for participants, everything is black and white. Participants in Experiment 2a see only horizontal or vertical pairs, as in Experiments 1a-c. Experiment 2b is a control condition using horizontal and vertical pairs for all participants. **Break** The break after the first training phase is identical to Experiments 1a-c. **2AFC Test Trials** In all 2AFC test trials, participants need to decide which of two shape pairs seems more familiar. In *Standard Learning Trials* trials, participants need to decide between a pair of the training phase and a random recombination of shapes from two pairs. In *Spatial Learning Trials* trials, participants see a pair used in the training phase twice, once in its correct orientation and once rotated by 90 degrees. In *Bias Trials* trials, participants see the same pair twice in different orientations. However, for these trials, the pair is either a foil pair from the *Standard Learning* trials or a combination of completely novel shapes. For these trials, there is no correct answer. **Debriefing** After the experiment, participants answer a set of open questions used to assess whether they have explicit knowledge of the presence of pairs in the input.

2.3.1 Experiment 2a: Direct Utilization of Structural Biases

Participants

120 participants (42 female, mean age = 26.9, SD = 8.5) were recruited via prolific.co. All participants had normal or corrected-to-normal vision. A smaller sample size than for the previous experiments (1a-c) was chosen based on the assumption that this simpler design should produce stronger and more consistent effects. The hourly compensation was £ 6.3. The study was approved by the Psychological Research Ethics Board of the Central European University, and all participants provided informed consent.

Materials

The same materials as in experiment 1a were used.

Procedure

The overall procedure was similar to that of experiment 1a (see Figure 2.4). However, participants only completed the first familiarization phase (using only horizontal or vertical pairs) and then moved on to test trials after a 2-minute break. In this experiment, participants completed four different types of test trials. First, in 24 ***standard learning trials***, participants decided between a real pair and a foil pair, just as in Experiment 1a. Second, in 6 ***spatial learning trials***, participants were presented with shapes of a real pair, once in the correct spatial arrangement and once rotated, e.g., correct horizontal and incorrect vertical arrangement of a pair. This test was used to assess whether participants knew in which orientation the pairs appeared during the training phase. Third, in 6 ***old-token bias trials***, participants again decide between the same shapes in horizontal and vertical arrangement, however this time for foil pairs, i.e., a combination of shapes that did not reliably co-occur during training. Finally, fourth, 8 ***new-token bias trials*** used the same logic but employed never-before-seen shapes to form pairs instead of using shapes from the familiarization phase. There were no correct answers in the two types of bias trials, and they served as a measure of bias toward generalization (same orientation as learned pairs) or toward structural novelty effect (orientation different from same pairs). The order of test trials was as follows. First, the *spatial learning trials* and the *old-token bias trials* were presented in an intermixed manner; this was followed by the *new-token bias trials*, and finally, the *standard learning trials* were shown. This ordering assured a balanced exposure to all shapes and orientations across time and a minimal retraining of the true pairs during the test.

Results

The same exclusion criteria were used as in Experiment 1a. This led to 5 exclusions for failed attention checks and 2 exclusions for response bias, leaving us with 113 participants after exclusions. Participants were categorized as explicit or implicit in the same way as in experiment 1a. This led to 22 participants being categorized as explicit and 91 as implicit.

The results (Figure 2.5 top panel) of Experiment 2a showed that participants with implicit knowledge (n = 91) performed above chance for the *standard learning trials* (M = 55.9, SE = 1.3, d = 0.49, $t(90)$ = 4.66, p < .001, BF = 1,494) but not for the *spatial learning trials* (M = 50.2, SE = 2.7, d = 0.01, $t(90)$ = 0.07, p = .999, BF = 0.12). Therefore, they could correctly identify which shapes co-occurred but not in what arrangement they co-occurred. Furthermore, they show a spatial bias towards the novel structure in the *old-token bias trials* (M = -9.3, SE = 2.3, d = 0.44, $t(90)$ = 4.19, p < .001, BF = 285) but not the *new-token bias trials* (M = -1.5, SE = 1.7, d = 0.09, $t(90)$ = 0.89, p = .999, BF = 0.17). Participants with explicit knowledge (n = 22) show above chance performance for both the *standard learning trials* (M = 84.5, SE = 4.0, d = 1.9, $t(21)$ = 8.67, p < .001, BF = $6.1*10^5$) and the *spatial learning trials* (M = 81.1, SE = 4.6, d = 1.5, $t(21)$ = 6.81, p < .001, BF = $1.8*10^4$). Therefore, participants with explicit knowledge could identify which shapes co-occurred and in what arrangement they co-occurred.

Furthermore, they also show a spatial bias towards the novel structure in the *old-token bias trials* (M = -18.9, SE = 4.8, d = 0.84, t(21) = 3.93 p = .003, BF = 45.3) but not the *new-token bias trials* (M = 1.1, SE = 4.2, d = 0.06, t(21) = 0.27 p = .999, BF = 0.23).

These findings are surprising as for participants with explicit knowledge, the direction of bias is opposite to that observed in Experiment 1a. While in Experiment 1a, they showed generalization of structure, here they showed a structural novelty effect. However, this might be explained by a weakness of experimental design, combined with the high level of learning shown by explicit participants in this experiment. The weakness in experimental design for the *old-token bias trials* is that the foil pairs combine shapes from two different real pairs, keeping the position of shapes within the pair where possible. To create a horizontal foil pair, the left shape of one pair is combined with the right shape of another pair, keeping those relative positions. Therefore, participants with strong knowledge of the real pairs' spatial arrangement know that the respective shapes of the foil pair should never appear in that relation, as this violates their pair knowledge. For example, assuming horizontal pairs AB and CD, shape C cannot be to the left of shape B, as shape A always appeared there during the training. Therefore, the foil pair CB violates the knowledge of participants. This is not the case for the rotated version of the foil pair, as there is no strong expectation about what should be above or below A and B. This means that if participants have a strong understanding of the structure of the pairs, as the explicit participants demonstrate with their high *standard learning* and spatial *learning performance*, they have a choice between a foil pair violating their knowledge and a foil pair not violating their knowledge. Choosing the option that does not violate their knowledge would lead to the effect observed in the *old-token bias trials*.

On the *standard learning trials,* we see higher performance in Experiment 2a than in Experiment 1a for participants with explicit knowledge (d = 0.70, t(53.6) = 2.74, p = .025, BF = 3.7) but not with implicit knowledge (d = 0.06, t(221.83) = 0.53 p = .594, BF = 0.16). This suggests that the explicit participants were subject to retroactive interference: learning pairs in the second phase of Experiment 1a interfered with remembering the previously learned pairs.

Discussion

The results showed the same structural bias for implicit participants that was also found in the transfer setup of Experiment 1a. The results are unclear for explicit participants based on a flaw in the experimental design. Surprisingly, implicit participants show no knowledge of the orientation of specific learned pairs, although it evidently affects making decisions about foil pairs. One potential explanation is that the structural novelty effect and the knowledge of pair orientation point in opposite directions, which, assuming similar magnitudes for both, would predict the observed chance performance. The idea here is that the structural novelty effect demonstrated with the *old-token bias trials* and before in Experiments 1a and 1b, by definition, points in the opposite direction than the orientation of the previously learned pairs. This would suggest that the participants with implicit knowledge are faced with conflicting information in applying their weak knowledge to the *spatial learning trials,* which could explain a chance performance by these two effects canceling out. This interpretation can be tested by removing the structural bias, as we would predict that participants with implicit knowledge then perform above chance for the spatial learning trials and, therefore, demonstrate learning of the pair orientations. This was realized in Experiment 2b.

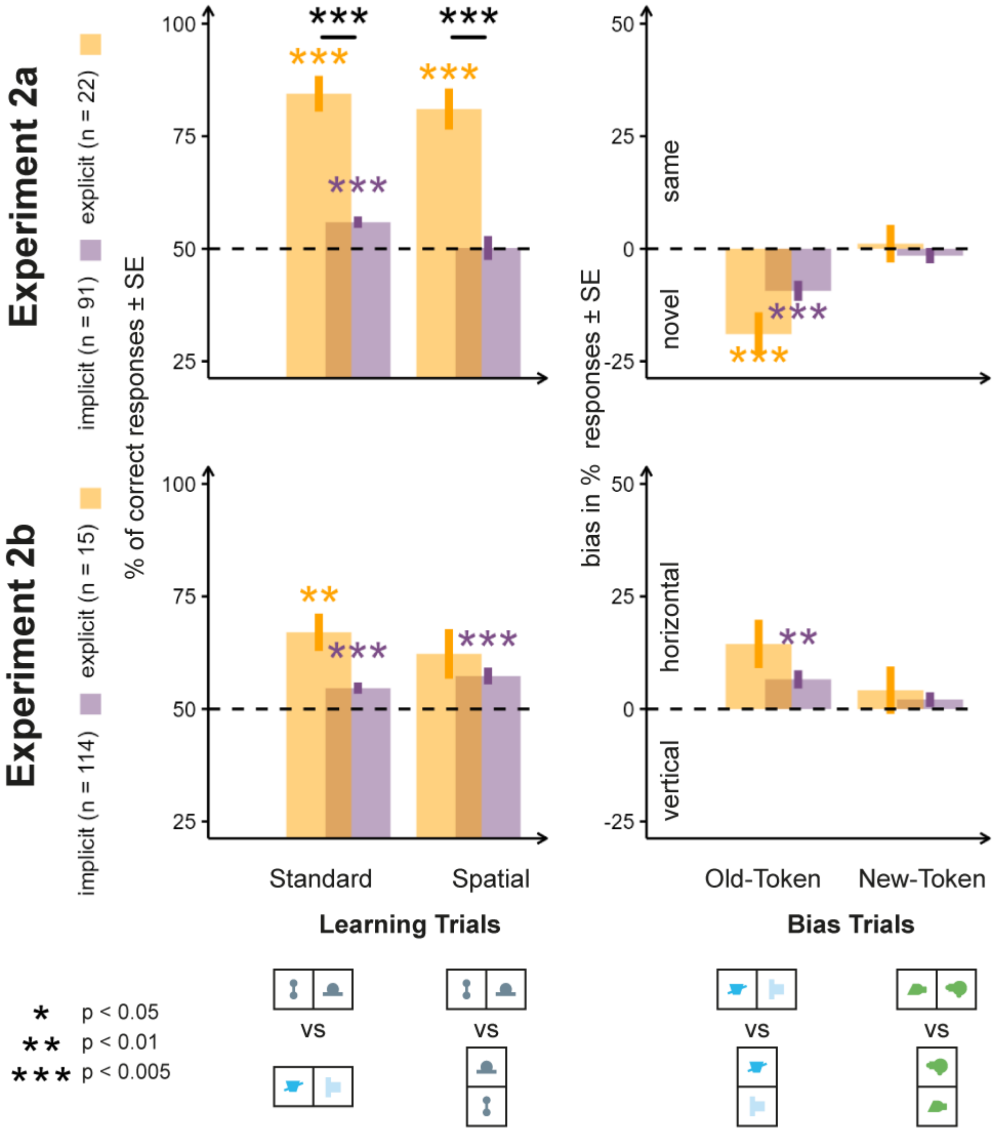

Figure 2.5 Experiments 2a and 2b Results. **Top Panel** Performance in Experiment 2a. **Bottom Panel** Performance in Experiment 2b. Results of 2AFC familiarity tests in Experiment 1a. Test trials are grouped along the x-axis according to test type. **Standard learning trials** signify a decision between a real and a foil pair, identical to the test trials in Experiments 1a-c. **Spatial learning trials** use a real pair, once in its correct orientation and once rotated by 90 degrees. **Old-token bias trials** use a foil pair, once in horizontal and once in vertical orientation. **New-token bias trials** use a pair of shapes that were not used in the training, once in horizontal and once in vertical orientation. For the two types of bias trials, there is no correct answer, and performance is expressed as bias for novel or same structure (Experiment 2a) or bias for horizontal or vertical structure (Experiment 2b). For other trial types, the y-axis represents the proportion of correct responses in the 2AFC test trials. Bars represent the standard error. Color coding indicates implicit and explicit subgroups of the participants. The horizontal dotted line at 50% denotes chance performance. The asterisks above bars denote significance levels from chance, while above lines they denote significance levels comparing two conditions. The legend of significance levels is shown in the lower left corner.

2.3.2 Experiment 2b: Measuring a priori Structural Biases

Experiment 2b (see Figure 2.4) has been designed to answer two questions. First, do explicit and implicit participants have a priori biases about orientation? I.e., do they preferably choose horizontal or vertical pairs if the learning phase does not bias them in any direction? Second, do implicit participants actually have usable knowledge of the pair orientation, which was only masked by the structural novelty effect in Experiment 2a? In order to test these two questions, Experiment 2b follows the same setup as Experiment 2a, but it uses both horizontal and vertical pairs in the training, therefore eliminating the induction of a structural bias. This setup is very close to that of classical spatial VSL studies (Fiser & Aslin, 2001) while including a larger array of types of test trials to allow for a deeper understanding of the underlying representations.

Participants

138 participants (63 female, mean age = 27.3, SD = 8.3) were recruited via prolific.co. All participants had normal or corrected-to-normal vision. The sample size was chosen to approximately match the one of Experiment 2a. The hourly compensation was £ 6.3. The study was approved by the Psychological Research Ethics Board of the Central European University, and all participants provided informed consent.

Materials

The materials were similar to Experiment 2a; however, the shapes were grouped into three horizontal and three vertical pairs for all participants (instead of 6 horizontal or 6 vertical pairs), therefore eliminating the bias in orientation.

Procedure

The procedure was identical to that of experiment 2a.

Results

The same exclusion criteria as in Experiment 1a were used. This led to 7 exclusions for failed attention checks and 2 exclusions for response bias. This left us with 129 participants after exclusions. Participants were categorized as explicit or implicit in the same way as in Experiment 1a. This led to 15 participants being categorized as explicit and 114 as implicit.

The results (Figure 2.5 bottom panel) show that participants with implicit knowledge (n = 114) perform above chance for both the *standard learning trials* (M = 54.6, SE = 1.2, d = 0.35, $t(113)$ = 3.74 p = .002, BF = 65.2) and the *spatial learning trials* (M = 57.3, SE = 1.9, d = 0.37, $t(113)$ = 3.91, p = .001, BF = 115). For participants with explicit knowledge (n = 15) we observe that while they perform above chance for *standard learning trials* (M = 67.0, SE = 4.1, d = 1.06, $t(14)$ = 4.11, p = .006, BF = 36.8), they perform significantly worse on those trials in experiment 2b as compared to experiment 2b (d = 0.99, $t(32.9)$ = 3.04, p = .019, BF = 7.8). Both participants with explicit and participants with implicit knowledge show a bias toward choosing horizontal over vertical options in the *old-token bias trials* (explicit: M = 14.4, SE = 5.4, d = 0.70, $t(14)$ = 2.69, p = .070, BF = 3.5; implicit: M = 6.6, SE = 2.0, d = 0.31, $t(113)$ = 3.26, p = .007, BF = 14.7) but not the *new-token bias trials* (explicit: M = 4.2, SE = 5.3, d = 0.20, $t(14)$ = .79, p = .442, BF = 0.34; implicit: M = 2.1, SE = 1.6, d = 0.12, $t(113)$ = 1.29, p = .401, BF = 0.23).

Discussion

The results show that both explicit and implicit participants have an a priori bias for horizontal structures. This highlights how we cannot assume our participants to be blank slates even when using abstract, artificial, and seemingly arbitrary stimuli and associations. The results further show that implicit participants have usable knowledge about pair orientation, in line with the interpretation of the results of Experiment 2a given above.

The finding that participants with explicit knowledge perform worse in this experiment than in Experiment 2a might suggest a generalization effect happening within the learning phase in Experiment 2a. This constitutes an even faster generalization than the one reported between learning phases in Experiment 1a. This is the case as the only difference between Experiments 2a and 2b is the presence of only one type of structure in Experiment 2a. Therefore, the better performance likely results from participants immediately re-applying this structure after learning one or a few pairs following that structure.

2.4 General Discussion

Summarizing this chapter, I presented a series of five experiments utilizing a transfer learning version of the classical statistical learning paradigm to show that participants with explicit knowledge can abstract and generalize the types of structure underlying a statistical learning task immediately, while participants with implicit knowledge show a structural novelty effect in immediate transfer and application of knowledge. These findings connect in interesting ways to findings within and outside the statistical learning literature but also leave several open questions.

The combined results of Experiments 2a and 2b highlight how participants use acquired and existing biases in a flexible fashion. The preference for horizontality found in Experiment 2b in the absence of any bias in the input cannot be explained by specific features of this experiment as compared to others. This suggests that this is a pre-existing bias also present in participants in the other experiments. However, in those experiments, it was overshadowed by features of the input (use of only horizontal or vertical pairs) and, therefore, not applied to the test trials. This form of flexible use of biases and the idea of pre-existing biases will be reencountered in Chapters 5 and 4, respectively. Interestingly, a bias for horizontal orientations has previously been reported in other domains, such as visual processing (Lim & Sinnett, 2012), face perception (Balas et al., 2015; Dakin & Watt, 2009), and direction of saccades as captured with eye-tracking (Foulsham et al., 2008; Gilchrist & Harvey, 2006; Tatler & Vincent, 2008; Van Renswoude et al., 2016).

There has been a debate in the statistical learning literature on whether or not explicit task instructions lead to different outcomes for SL. Some authors argued that explicit instructions do not matter (Arciuli et al., 2014) while others argued that it can matter based on context factors (Bertels et al., 2015). In general, this discussion focused on quantitative performance improvements. Contrasting the results of the implicit participants of Experiment 1a with the results of Experiment 1c shows a critical effect of explicit instructions on both quantitative and qualitative behavior. Explicit instructions lead to high learning performance and generalization, compared to low learning performance and a structural novelty effect. The comparison of the results of the explicit participants of Experiment 1a with the results of Experiment 1c furthermore shows that the effects of explicit instructions are very similar to the effect of participants achieving explicitness of knowledge on their own.

An interesting difference between the performance of participants with explicit and implicit knowledge is that the direct comparison between Experiments 1a and 2a shows *retroactive interference* (Dewar et al., 2007; Wixted, 2004) only for participants with explicit knowledge. That is, for explicit participants, learning the pairs in the second learning phase of Experiment 1a interfered with remembering the pairs of the first learning phase. These two experiments are identical in the first learning phase, and the test items used for both of them can, therefore, be directly compared. Interestingly, this retroactive interference for participants with explicit knowledge is in contrast to a potential *proactive interference* for participants with implicit knowledge discussed in the final paragraph of this chapter.

Contrasting the findings of participants with explicit and implicit knowledge over all experiments presented in this chapter demonstrates critical differences between them, both quantitative and qualitative. In contrast to participants with implicit knowledge, participants with explicit knowledge show higher learning performance, generalization in structural transfer, and retroactive interference. This highlights how problematic it can be to ignore explicitness as a covariate, as many studies in statistical learning do. For mere quantitative differences, this can be problematic as averaging over implicit and explicit participants can lead to artificial group means which are not representative of any of the constituting subgroups. This is even more problematic for qualitative differences, as the current example demonstrates. In this case, averaging over explicit and implicit participants would have masked both the generalization and structural novelty effects found within this group. Consistently doing this split of the sample into explicit and implicit learners might, however, be costly for statistical learning research, as it means collecting data from enough participants to be able to make inferences for and between two subgroups and dealing with smaller effects than previously estimated in the literature (for the implicit subgroup). However, the alternative is making inferences based on artificial group means, which do not describe the actually existing subgroups. Interestingly, the significantly different rates of explicitness found between Experiment 1a, using horizontal and vertical pairs, and Experiment 1c, using diagonal pairs, suggests that this issue will differentially affect varying experimental setups.

A critical question about the results presented in this chapter is how the structural novelty effect observed for participants with implicit knowledge relates to the generalization behavior observed for participants with explicit knowledge. In the simplest sense, they can be seen as different cases of the same phenomenon, the transfer of structural properties from one learning phase to another, just with inverted signs. However, are they built on the same type of underlying representation used in different ways or on fundamentally different representations? I would argue that the simplest explanation of the behavior of participants with explicit knowledge is the formation of abstracted knowledge based on the representational overlap of the learned pairs, which consists of their shared orientation. This amounts to a factorized representation, representing the orientation (vertical or horizontal) in its own right, which is later applied in the interpretation of novel input, leading to a bias for the familiar structure, i.e., generalization. In principle, the behavior of participants with implicit knowledge could be explained in the same way, with some additional explanation of why they invert the direction of effect in applying their abstract knowledge to new input. However, a more straightforward explanation for the results of the implicit group might be a *structure-level interference* effect. In this view, implicit participants in my experiment could not utilize the representational overlap for abstraction and, therefore, do not represent the shared pair orientation as an abstract feature of the input. New incoming pairs of the same structure would reactivate the same representational space containing the representation of the previously learned pairs by sheer similarity. This simultaneous activation of old and new pairs could lead to *proactive interference* (Kliegl & Bäuml, 2020), hindering the learning of the new pairs of the same orientation (Experiment 1a+b) and making them less appealing choices in direct utilization to test trials (Experiment 2a). New pairs with a different structure would not be subject to this same interference. This would lead to the pattern of behavior observed in the experiments presented in this chapter. Proactive interference based on similarity is well established in learning (Kliegl & Bäuml, 2020), therefore making my interpretation highly plausible in explaining the transfer learning results of Experiments 1a and 1b. However, this is more speculative for the decision-making required in direct utilization to test items of Experiment 2a, and further studies would be required to support this idea.

If my interpretation of the transfer learning results in Experiments 1a and 1b is correct, a natural next question is: are there circumstances where participants with implicit knowledge can also abstract from the representational overlap of the learned pairs and, therefore, show a generalization behavior? I.e., are there circumstances where participants with implicit knowledge show the same pattern of behavior as the participants with explicit knowledge did in the current chapter? In Chapter 3, consolidation - a phase of offline processing - is investigated as a candidate for this. Using three experiments building on the paradigm developed in the current chapter, I show that asleep but not awake consolidation leads participants with implicit knowledge to generalize the initially learned structure to novel input.

CHAPTER 3

Consolidation and Generalization in Visual Statistical Learning

The study presented in this chapter builds on the unsupervised transfer learning paradigm introduced in the previous chapter and adopts it to study the influence of consolidation - phases of offline processing - on implicit structural learning and transfer. I present a series of three experiments that demonstrate that after asleep, but not awake, consolidation participants with implicit knowledge are able to generalize structural knowledge. It furthermore confirms that this effect is specific to sleep and cannot be explained by a time-of-day effect.

3.1 Consolidation, Abstraction, and Statistical Learning

An influential view on memory in Psychology and Neuroscience distinguishes between three distinct phases: *encoding*, *storage*, and *retrieval* (Melton, 1963). While encoding and retrieval encompass how a memory is initially formed and later used again, storage describes everything that happens in between. This is not akin to passive storage of information on a computer hard drive but is an active phase in which memories are transformed by a process called *memory consolidation* during offline phases while we are awake or asleep (Squire et al., 2015). Memories can get stronger or weaker, and crucially, they can undergo qualitative changes such as abstraction. As mentioned in previous chapters, the key feature of abstraction is the extraction of commonalities between several instances (Blackburn, 2008). This role of memory consolidation as an enabler of abstraction and generalization is the focus of the current study. As a point on terminology, this reorganization function is often called *systems consolidation*, in contrast to the shorter-lived processes of *synaptic consolidation*. For this dissertation, the focus is on the system-level reorganization of memory and the resulting qualitative changes in memory-guided behavior. Therefore, the term consolidation will mainly be used to refer to systems consolidation. The next subchapters will give an overview of previous research on consolidation and abstraction (3.1.1), consolidation and statistical learning (3.1.2), and consolidation and explicitness of knowledge (3.1.3).

3.1.1 Consolidation and Abstraction

An instrumental role of consolidation processes in abstraction and generalization has previously been demonstrated for both artificial (Ellis et al., 2021; Wittkuhn et al., 2021) and biological systems (Chambers, 2017; Diekelmann & Born, 2010; Klinzing et al., 2019; Lerner & Gluck, 2019; Lewis & Durrant, 2011; Rasch & Born, 2013). These studies used computational (Ellis et al., 2021; G. E. Hinton et al., 1995; McClelland et al., 1995; Singh et al., 2022) and empirical approaches to demonstrate the major effect of consolidation under sleep-based (Djonlagic et al., 2009; Lutz et al., 2017, 2018; Schapiro, McDevitt, et al., 2017; Sweegers et al., 2014) and awake conditions (Hennies et al., 2014).

An influential theoretical account and computational model of systems consolidation is the *complementary learning systems* (CLS) framework (McClelland et al., 1995; O'Reilly et al., 2014). Based on neural network models and findings about the neural correlates of memories in the brain, this framework suggests that human memory function is based on an interaction of two complementary memory systems. The first one, the *hippocampus*, is fast learning, realized by high sparsity and plasticity of neural connections within the system. The second, situated throughout the *neocortex*, is slower learning, realized by overlapping representations and lower plasticity within the system. According to this framework, human memory performance is based on an interaction of the two systems. The fast-learning hippocampus builds episodic memories online, and later "teaches" the slow-learning neocortex during offline phases by reactivating its own representations, leading to corresponding activity in the neocortex. This interaction does not consist of simply moving memories from one storage site to another but serves two essential interrelated functions. First, the necessity of two learning systems with different properties became apparent in early connectionist studies of knowledge representation in neural network models, which showed *catastrophic interference*. This means that new information fed to the system interfered with previously stored information, leading to forgetting what had already been learned. The two complementary learning systems help with this, as the first system is fast enough to quickly acquire regularities, while the second system is slow enough to incorporate new information without wiping out old knowledge. The second function of the interaction of the systems in CLS is abstraction. In this view, the neocortical learning system is slow and only reliably stores information after multiple repetitions. As different but similar memories are fed from the hippocampus to the neocortex, the overlap between the different memories is activated in the neocortex for every one of them. In contrast, the differences between them, the unique properties of episodes, are only activated for the specific instances. Therefore, the commonality is represented stronger and longer lasting, leading effectively to an abstraction of the shared properties of several memories. The properties of the two learning systems have been linked to episodic and semantic memory systems, respectively.

This basic version of the CLS framework, as proposed several decades ago, has undergone revisions and criticism. A selection of these is discussed in this paragraph. Evidence has accumulated that a simple one-to-one mapping of memory systems and brain regions is an oversimplification not supported by empirical findings on functional neuroanatomy (Sherman et al., 2024). It has also been suggested that empirically demonstrated hippocampal functions going beyond *episodic memory* and extending to statistical learning (Covington et al., 2018; Schapiro et al., 2014, 2012, 2016) can be explained by the presence of additional complementary learning systems within the hippocampus (Schapiro, Turk-Browne, et al., 2017). Furthermore, it was suggested that the reactivation of memories for consolidation is a guided process influenced by factors such as strength and depth of initial encoding (Denis et al., 2020; Schapiro et al., 2018), learned rules and schemas (Y. Liu et al., 2019; Preston & Eichenbaum, 2013), utility (Y. Liu et al., 2021; Mattar & Daw, 2018), and the potential for generalization (Sun et al., 2023).

Finally, research on sleep's role in systems consolidation has been connected to accumulating research on sleep stages, suggesting different roles for *slow wave sleep* (SWS) and *rapid eye movement* (REM) sleep in the reorganization of memory through consolidation (Diekelmann & Born, 2010; Gilboa & Marlatte, 2017; Kumaran & McClelland, 2012; O'Reilly et al., 2014; Rasch & Born, 2013; Schapiro, McDevitt, et al., 2017; Singh et al., 2022; Squire et al., 2015; Witkowski et al., 2020).

The core idea of the CLS framework of enabling abstraction and generalization based on the representational overlap of reactivated memories has been echoed in several different conceptualization of the qualitative effects of consolidation on memory (Chambers, 2017; Diekelmann & Born, 2010; Klinzing et al., 2019; Rasch & Born, 2013; Singh et al., 2022; Sun et al., 2023; Sweegers et al., 2014; Tse et al., 2007; Winocur et al., 2010). For the purpose of this dissertation, it is precisely this feature that is most important as it is highly relevant for the interpretation of the experiments discussed in the current chapter.

3.1.2 Consolidation and Statistical Learning

The findings on the influence of consolidation on statistical learning are mixed and divergent, showing clear effects of sleep on performance for some setups (Durrant et al., 2013, 2016, 2011) and no or very limited effects for others (Arciuli & Simpson, 2012; Hallgató et al., 2013; Kim et al., 2009; McDevitt et al., 2022; Nemeth et al., 2010; Quentin et al., 2021; Simor et al., 2019).

For auditory statistical learning using probabilistic sequences, it was shown that consolidation over long (8 hours) and short (90 minutes) phases of sleep led to higher performance, as compared to awake consolidation over the same period (Lewis & Durrant, 2011). Furthermore, longer delays (24 hours) lead to stronger performance improvement than shorter delays (30 minutes) (Durrant et al., 2013). The cross-modal transfer of auditory statistical knowledge to visual sequences was found to be contingent on a long delay (24 hours) (Durrant et al., 2016). In all of these studies, the amount of slow-wave sleep predicted improved performance or transfer.

For the *alternating serial reaction time* (ASRT) task, measuring simultaneously statistical learning of simple (adjacent) and complex (non-adjacent) regularities, results are more complicated. It was shown that while statistical learning of adjacent regularities happens fast and then plateaus, not being altered by a period of awake or asleep consolidation (Hallgató et al., 2013; Nemeth et al., 2010), learning of non-adjacent regularities, is slower, and more gradual, and further increases during consolidation (Quentin et al., 2021; Simor et al., 2019).

For temporal VSL, it was shown that test performance is the same directly after the familiarization with the stimuli or after delays of 30 minutes, 1 hour, 2 hours, 4 hours (Arciuli & Simpson, 2012), or 24 hours (Kim et al., 2009). This suggests that temporal VSL is stable and not dependent on awake or asleep consolidation. McDevitt et al. (2022) were the first to study the effect of sleep on spatial VSL. The results showed that in a standard condition, sleep did not benefit spatial VSL, as compared to awake consolidation. When interference was present, in the form of two successive stimuli sets, learning was found in a REM sleep and an active-wake group but not in a non-REM sleep and a quiet-wake group. This suggests that consolidation phases have influences on spatial VSL, although their interpretation is not straightforward. It is not either sleep or awake consolidation that improves learning; instead, the details of the consolidation phase matter for both cases.

All of the studies discussed so far focused on the effect of consolidation on learning specific (simple or complex) item-item associations. The study presented in the current chapter goes beyond this by investigating how consolidation influences the abstraction from and generalization of such specific item-item associations. In my case, the generalization is the structural transfer of what has been abstracted to novel input. Furthermore, most previous studies on consolidation and statistical learning did not consider the actual state of acquired knowledge and simply assumed implicitness of knowledge based on the properties of the used statistical learning paradigms. However, one study did consider the quality of knowledge and found that while implicit representations are strengthened during a 24-hour consolidation period, explicit representations are decaying (H. Liu et al., 2023). This finding stresses the importance of measuring the state of participants' knowledge in statistical learning research. This was already done in the previous chapter on statistical and structure learning (Chapter 2) and will again be a major topic for the analysis in the current chapter. Overall, the current investigation goes beyond previous work by systematically investigating the influence of sleep and awake consolidation on the abstraction and generalization of the structure of learned items for both explicit and implicit knowledge.

3.1.3 Consolidation and Explicitness

Consolidation has been previously linked to the explicitness of knowledge. It was demonstrated that knowledge that was represented implicitly after initial encoding can become explicit after sleep (Fischer et al., 2006; Wagner et al., 2004; Zander et al., 2017). Furthermore, it was demonstrated for different domains that implicit and explicit representations can be differentially influenced by sleep (Lerner & Gluck, 2019; H. Liu et al., 2023; Robertson et al., 2004). These findings further support the notion that tracking the explicitness of representations in the current study is highly relevant for reliable and meaningful analysis.

3.2 Implicit Generalization Through Consolidation

3.2.1 Experiment 3a: Sleep Enables Generalization of Implicit Knowledge

While the previous chapter investigated how unsupervised learning of structural knowledge can be immediately applied to future learning or direct testing, the current chapter focuses on how this is influenced by intermediate phases of offline processing; i.e. consolidation. For this purpose, the experiments in this chapter build on the new unsupervised transfer learning paradigm introduced in the previous chapter (see Figure 3.1). The key idea there was that in a first training phase, participants are exposed to visual scenes created from combinations of fixed shape pairs with a single shared underlying structure: horizontal or vertical orientation. Following this, in a second training phase, they are exposed to new scenes made from a set of new shapes, which are grouped in pairs of both underlying structures: horizontal and vertical orientation. While in the previous chapter, this setup was used with a minimal break between the two learning phases of only two minutes, the current chapter focuses on how the duration and state of consciousness between the two learning phases influence learning and transfer of knowledge. Experiment 3a directly tests the effect of sleep on unsupervised structural transfer by introducing a 12-hour period during the night between the two learning phases. As with Experiment 1a in the previous chapter, the central question of interest here is not specific quantitative levels of learning but how first learning patterns of one type of structure influences subsequent learning of patterns of the same and novel structures and how the explicitness of knowledge moderates this. Based on the literature discussed in 3.1.1, the hypothesis here is that consolidation during sleep will enable participants with implicit knowledge to abstract the shared structure of multiple learned pairs, which can then, in turn, be generalized to novel input. This would demonstrate for unsupervised implicit learning, something previously shown only for more explicit learning under guidance (supervised and reinforcement learning).

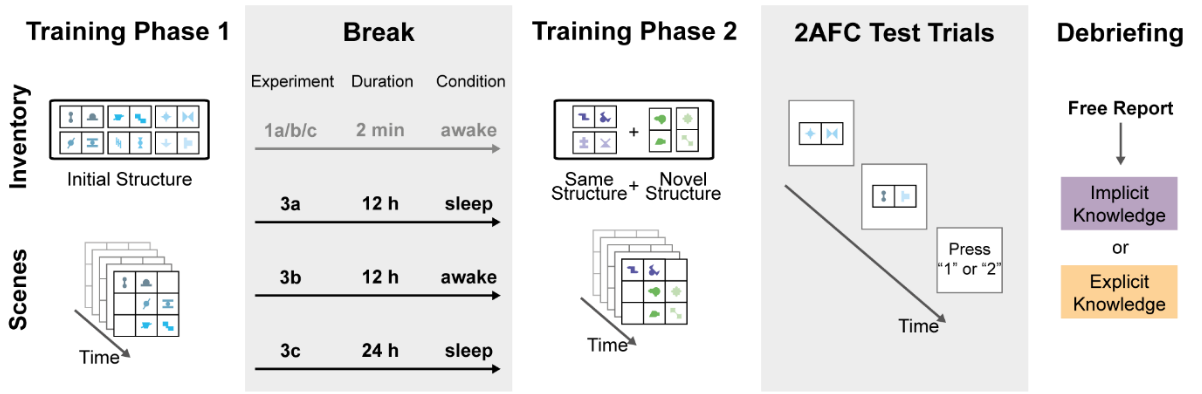

Figure 3.1 Statistical Learning Transfer Paradigm - 2 **Training Phase 1** The first training phase follows the classical spatial VSL paradigm, presenting scenes made from shape pairs without segmentation cues. All colors are only for illustration; for participants, everything is black and white. For this learning phase, all pairs have the same orientation (horizontal or vertical), counterbalanced between participants. **Break** The break after the first training phase varies between Experiments, as described in the figure. **Training Phase 2** The second training phase consists of novel scenes made from novel shapes. In this phase, all participants see horizontal and vertical pairs. **2AFC Test Trials** In all 2AFC test trials, participants are presented with a real pair from the training phases and a foil pair made by combining shapes of two real pairs. They need to decide which of the two shape pairs seems more familiar. **Debriefing** After the experiment, participants answer a set of open questions which are used to assess whether they have explicit knowledge of the presence of pairs in the input.

Participants

259 participants (127 female, mean age = 25.6, SD = 8.5) were recruited via prolific.co. The hourly compensation was £ 6.3. All participants had normal or corrected-to-normal vision. The sample size was chosen to match that of Experiment 1a. The study was approved by the Psychological Research Ethics Board of the Central European University, and all participants provided informed consent. In order to ensure that participants had overnight sleep during the experiment as intended, several constraints and checks were implemented (see Appendix B).

Materials

The materials of the main part of the experiment were identical to Experiment 1a. Additionally, participants filled out the *Pittsburgh Sleep Quality Index* (PSQI) (Buysse et al., 1989) and the *Groningen Sleep Quality Scale* (GSQS) (Meijman et al., 1988).

Procedure

The procedure (see Figure 3.1) within the main tasks was identical to Experiment 1a. However, in this experiment, participants completed the first training phase in the evening at around 09:00 p.m., followed by the GSGS and PSQI questionnaires. Twelve hours later in the morning at 09:00 a.m., they completed the second training phase, followed by all the test trials, and finally, they completed the GSGS again.

Results

The same exclusion criteria as for Experiment 1a were used. This led to 18 exclusions for failed attention checks and two exclusions for response bias. For this experiment, I also employed exclusion criteria related to sleep quality. 47 participants were excluded because they reported bad sleep quality for the night before the experiment or the night of the experiment as measured by the Groningen Sleep Quality Scale (GSQS, score below 9). 31 participants were excluded because they reported a bad habitual sleep quality measured with the Pittsburgh Sleep Quality Index (PSQI, score below 10). 161 participants remained after exclusions. Participants were categorized as explicit or implicit as in Experiment 1a. This led to 21 participants being categorized as explicit and 140 as implicit. Bayes Factors for Bayesian t-tests (Rouder et al., 2009) for implicit participants in Experiments 3a, b, and c used an r-scale parameter of .5 instead of the default √2/2, reflecting a change in prior after Experiment 1a found small effects.

The results (Figure 3.2) showed that participants with implicit knowledge (n = 140) performed above chance for pairs of the first training phase (M = 53.1, SE = 1.0, d = 0.27, t(139) = 3.24, p = .009, BF = 16.8) and for pairs of the *same structure* (M = 58.8, SE = 2.5, d = 0.30, t(139) = 3.51, p = .004, BF = 37.6) but not pairs of a *novel structure* (M = 46.8, SE = 2.6, d = 0.11, t(139) = -1.24, p = .435, BF = 0.27) in the second training phase. The performance for *same* and *novel structure* pairs is significantly different (d = 0.24, t(139) = 2.82, p = .027, *BF* = 5.4). Participants with explicit knowledge (n = 21) show the same pattern of results as they did in Experiment 1a, performing above chance for pairs of the first training phase (M = 74.8, SE = 4.2, d = 1.3, $t(20)$ = 5.9, $p < .001$, BF = 2,361) and for pairs of the *same structure* (M = 66.7, SE = 5.8, d = 0.63, $t(20)$ = 2.87, p = .038, BF = 5.3) but not pairs of a *novel structure* (M = 51.2, SE = 5.8, d = 0.04, $t(20)$ = 0.20, p = .841, BF = 0.23) in the second training phase. For explicit participants, the difference between pairs of the *same* and *novel structure* is not significant (d = 0.39, $t(20)$ = 1.77, p = .272, BF = 0.86), but we see a strong positive correlation between learning in the first learning phase and learning pairs of the same structure in the second learning phase (r = .52, p = .015). Again, no significant correlation is observed between the first learning phase and *novel structure* pairs in the second phase (r = .27, p = .232).

I conducted the same type of matched sample analysis as for Experiment 1a here (see Appendix C for details). As in Experiment 1a, the matched sample showed the same pattern as the full sample (see Figure 3.3). As a critical analysis, we can see that for the matched implicit sample, there is a significant difference between learning pairs of the *same* and of the *novel structure* (d = 0.98, $t(20)$ = 4.51, $p < .001$, BF = 127), suggesting generalization of the structure.

In Chapter 2, retroactive interference was shown for explicit participants, as they performed worse on pair learning if there was a second training phase with novel stimuli between initial training and test (Experiments 1a vs. 2a). As an exploratory test of release of interference after consolidation, I compared the performance of explicit participants in the current Experiment 3a with these experiments. While we see that the explicit participants performance for Experiment 3a (M = 74.8) is descriptively between the performance in Experiment 1a (M = 67.9) and Experiment 2a (M = 84.5), it is not significantly different from either (Experiments 3a vs 1a: d = 0.29, $t(51.5)$ = 1.11, $p < .270$, BF = 0.43; Experiments 3a vs 2a: d = 0.51, $t(40.7)$ = -1.67, $p < .103$, BF = 0.91). However, given the low sample sizes for explicit participants, this comparison was likely underpowered.

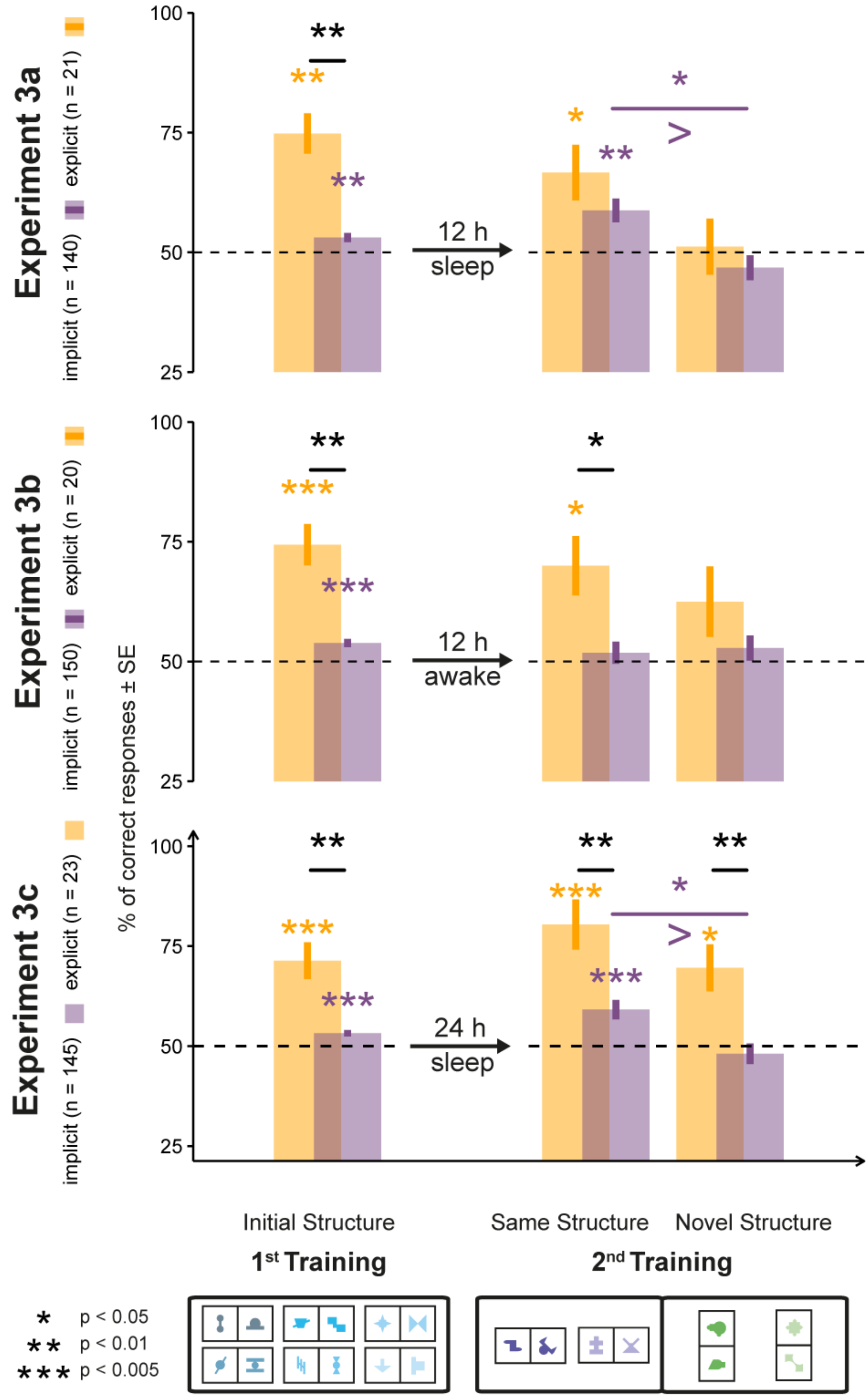

Figure 3.2 Experiments 3a-c Results. Results of 2AFC familiarity tests. Test trials are grouped along the x-axis according to which training phase they appeared in, and for the second training phase, according to whether they follow the same or a different structure (horizontal or vertical) than the pairs of phase 1. The y-axis represents the proportion of correct responses. Bars represent the standard error; color coding indicates implicit and explicit subgroups. The dotted line at 50% shows the chance level.

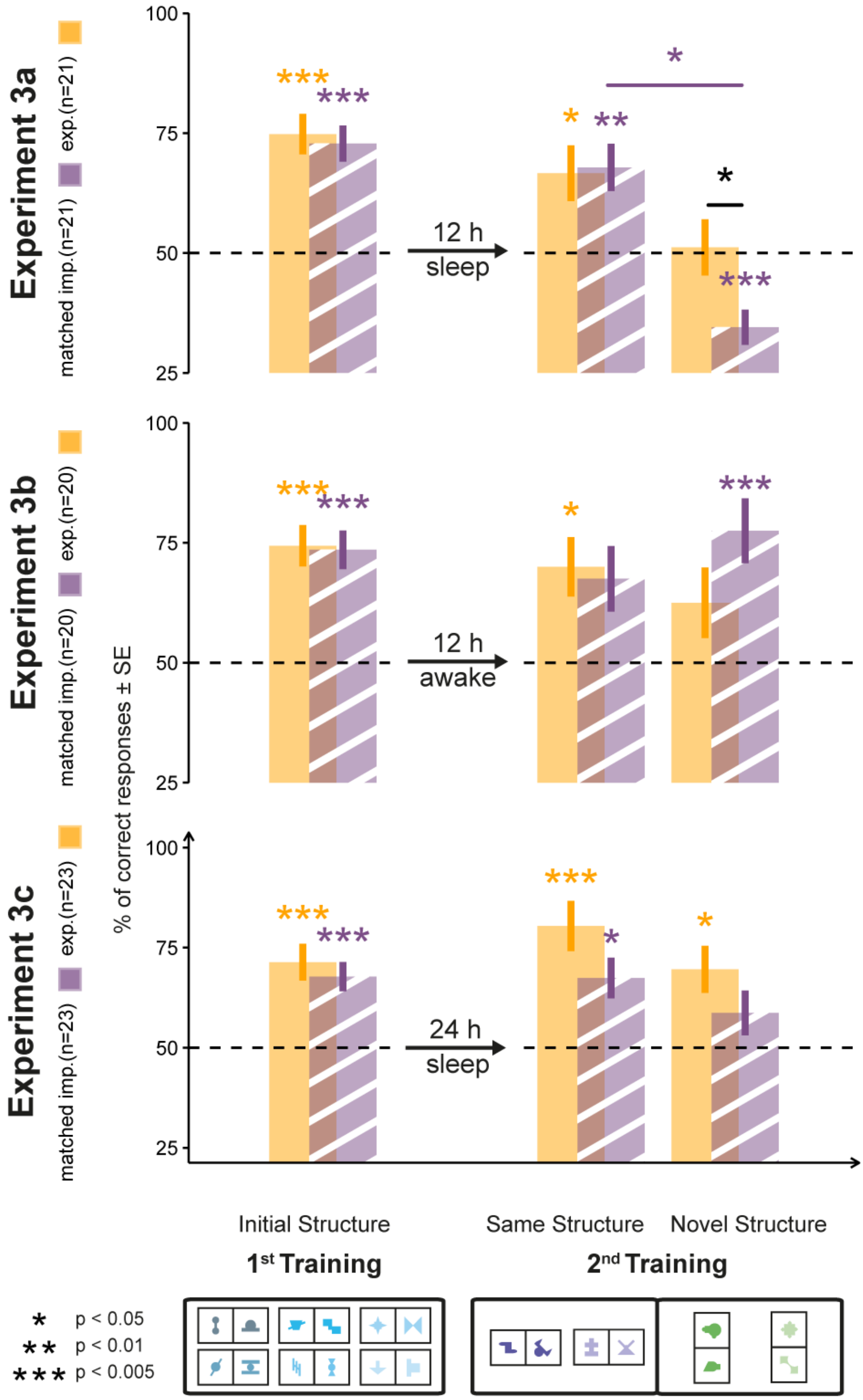

Figure 3.3 Experiments 3a-c Matched Sample Results. Results of 2AFC familiarity tests for the subsample of implicit participants found by matching to explicit participants' phase 1 performance. Otherwise, the plots are as described in the legend of Figure 3.2.

Discussion

We saw that in Experiment 3a, participants with implicit knowledge completely switched their behavior compared to the implicit participants in Experiment 1a and now behave as the explicit participants in Experiment 1a did. Therefore, the results of the current experiment demonstrate that consolidation during sleep can enable the abstraction and transfer of structural knowledge in unsupervised implicit learning. Matched sample analysis again demonstrated that this pattern of behavior is not moderated by the specific level of learning in the first training phase. However, the results of Experiment 3a alone are insufficient to demonstrate that the effect found is specific to sleep. Therefore, I conducted two further experiments that realize control conditions that are best practice in research on the effect of consolidation on learning (Németh et al., 2023).

3.3 Testing the Role of Sleep and Time-of-Day

3.3.1 Experiment 3b: The Effect of Consolidation on Generalization is Sleep Specific

Experiment 3b was designed to test whether the effect found in Experiment 3a is specific to sleep or is a general effect of consolidation. For this purpose, the 12-hour consolidation phase was moved to the daytime without any sleep occurring. Comparing the results of Experiments 3a and 3b is, therefore, a direct experimental test of the effect of sleep on unsupervised implicit structural transfer.

Participants

275 participants (134 female, mean age = 28.9, SD = 9.8) were recruited via prolific.co. The hourly compensation was £ 6.3. All participants had normal or corrected-to-normal vision. The sample size was chosen to match that of Experiment 1a. The study was approved by the Psychological Research Ethics Board of the Central European University, and all participants

provided informed consent. In order to ensure that participants conducted the experiment during the day and did not sleep during the consolidation phase, several constraints and checks were implemented (see Appendix B for details).

Materials

The materials were identical to those of Experiment 3a.

Procedure

The procedure (see Figure 3.1) was the same as Experiment 3a, with the difference that the first session took place in the morning at around 09:00 a.m. and the second in the evening at 09:00 p.m. As there was no night of sleep between the first and second sessions, participants filled out the GSQS only once in this experiment.

Results

The same exclusion criteria as Experiment 3a were used. This led to 29 exclusions for failed attention checks and three exclusions for response bias. 28 participants were excluded because they reported bad sleep quality for the night before the experiment as measured by the Groningen Sleep Quality Scale (GSQS, score below 9). 28 participants were excluded because they reported a bad habitual sleep quality measured with the Pittsburgh Sleep Quality Index (PSQI, score below 10). Additionally, 17 participants in this experiment were excluded as they reported sleeping during the day. This left us with 170 participants after exclusions. Participants were categorized as explicit or implicit in the same way as in Experiment 1a. This led to 20 participants being categorized as explicit and 150 as implicit. Bayes Factors from Bayesian t-tests for implicit participants reported for Experiments 3a, 3b, and 3c used an r-scale parameter of .5 instead of the default $\sqrt{2}/2$, reflecting that Experiment 1a found small effect sizes for this group.

The results (Figure 3.2) showed that participants with implicit knowledge (n = 150) that did not sleep during the consolidation phase performed above chance for pairs of the first training phase (M = 53.9, SE = 0.9, d = 0.36, t(149) = 4.46, p < .001, BF = 1,020), but not for pairs of the *same structure* (M = 51.8, SE = 2.3, d = 0.07, t(149) = 0.80, p = .999, BF = 0.17) or pairs of a *novel structure* (M = 52.8, SE = 2.6, d = 0.09, t(149) = 1.08, p = .999, BF = 0.22) in the second training phase. Participants with explicit knowledge (n = 20) again perform as they did in Experiment 1a, performing above chance for pairs of the first training phase (M = 74.4, SE = 4.3, d = 1.27, t(19) = 5.67, p < .001, BF = 1,281) and for pairs of the *same structure* (M = 70.0, SE = 6.2, d = 0.72, t(19) = 3.24, p = .026, BF = 10.4) but not pairs of a *novel structure* (M = 62.5, SE = 7.4, d = 0.38, t(19) = 1.70, p = .530, BF = 0.78) in the second training phase. As in Experiment 1a, for explicit participants, the difference between pairs of the *same* and *novel structure* is not significant (d = 0.19, t(19) = 0.86, p = .999, BF = 0.32), but again we do see a strong positive correlation between learning in the first learning phase and learning pairs of the *same structure* in the second learning phase (r = .56, p = .010). Again, no significant correlation is observed between learning in the first learning phase and learning pairs of the *novel structure* in the second learning phase (r = .42, p = .065).

I conducted the same type of matched sample analysis as for experiments 1a and 3a here (see Appendix C for details). As previously, the matched sample showed a pattern similar to that of the full sample (Figure 3.3). Critically, we can see that for the matched implicit sample, there is no significant difference between learning pairs of the *same* and of the no*vel structure* (d = 0.29, t(19) = -1.29, p = .214, BF = 0.59), suggesting no generalization of the structure.

Discussion

The results of Experiment 3b show that in the absence of sleep, participants with implicit knowledge do not generalize structural knowledge. This supports the idea that the type of offline processing necessary for the implicit abstraction of the shared structure of multiple learned patterns does not occur during all kinds of offline phases but is specific to sleep.

3.3.2 Experiment 3c: The Results of Experiments 3a+b Are Not Based On Time-of-Day

The results of Experiment 3b suggest that the effect found in Experiment 3a was specific to sleep. However, there is still an alternative explanation. It was previously suggested that generalization is more easily achieved in the morning than in the evening (Tandoc et al., 2021). This suggests that a simple comparison between an AM-PM condition and a PM-AM condition is insufficient to show the effect of sleep on generalization as it confounds the effect of sleep with a possible effect of time of day. To test for a possible time-of-day effect, Experiment 3c uses the same setup as the previous experiments but introduces a PM-PM condition, i.e., a 24-hour consolidation phase from one evening to the next. This means that participants in this condition sleep after the first learning phase, as participants in Experiment 3a, but they need to generalize their knowledge not in the morning but in the evening, as participants in Experiment 3c. The key idea here is that replicating the findings of the AM-PM condition in the PM-PM condition strongly supports an actual effect of sleep and against a time-of-day effect.

Participants

275 participants (129 female, mean age = 27.9, SD = 8.9) were recruited via prolific.co. The hourly compensation was £ 6.3. All participants had normal or corrected-to-normal vision. The sample size was chosen to match that of Experiment 1a. The study was approved by the Psychological Research Ethics Board of the Central European University, and all participants provided informed consent. In order to ensure that participants had overnight sleep during the experiment as intended, several constraints and checks were implemented (see Appendix B for details).

Materials

The materials were identical to Experiment 3a.

Procedure

The procedure (see Figure 3.1) was identical to Experiment 3a. However, in this experiment, participants completed the first session in the evening at around 09:00 p.m. and the second session 24 hours later again in the evening at 09:00 p.m.

Results

The same exclusion criteria as Experiment 3a were used. This led to 17 exclusions for failed attention checks and five exclusions for response bias. 51 participants were excluded because they reported bad sleep quality for the night before the experiment or the night of the experiment as measured by the Groningen Sleep Quality Scale (GSQS, score below 9). Additionally, 33 participants were excluded because they reported a bad habitual sleep quality measured with the Pittsburgh Sleep Quality Index (PSQI, score below 10). This left us with 169 participants after exclusions. Participants were categorized as explicit or implicit in the same way as in Experiment 1a. This led to 145 participants being categorized as implicit, 23 categorized as explicit, and one being excluded for being categorized as fully explicit (having explicit knowledge about the predominant pair orientation). Bayes Factors from Bayesian t-tests for implicit participants reported for Experiments 3a, 3b, and 3c used an r-scale parameter of .5 instead of the default $\sqrt{2}/2$, reflecting that Experiment 1a found small effect sizes for this group. The results (Figure 3.2) showed that participants with implicit knowledge (n = 145) performed above chance for pairs of the first training phase (M = 53.2, SE = 0.8, d = 0.34, $t(144)$ = 4.09, p = .001, BF = 260) and for pairs of the *same structure* (M = 59.1, SE = 2.4, d = 0.31, $t(144)$ = 3.78, p = .001, BF = 89.4) but not pairs of a *novel structure* (M = 48.1, SE = 2.6, d = 0.06,

t(144) = -.74, p = .459, BF = 0.17) in the second training phase. The performance for pairs of the *same* and *novel structure* was significantly different (M_{diff} = 11.04, d = 0.23, t(144) = 2.75, p = .021, BF = 4.4). Participants with explicit knowledge (n = 23) show the same pattern of results as they did in Experiment 1a, performing above chance for pairs of the first training phase (M = 71.4, SE = 4.6, d = 0.9, t(23) = 4.6, p = .001, BF = 217), for pairs of the *same structure* (M = 80.4, SE = 6.3, d = 1.01, t(23) = 4.85, p = .001, BF = 347), and pairs of a *novel structure* (M = 69.6, SE = 5.9, d = 0.69, t(23) = 3.33, p = .012, BF = 13.6) in the second training phase. As in Experiment 1a, for explicit participants, the difference between pairs of the *same* and *novel structure* is not significant (M_{diff} = 10.78, d = 0.26, t(22) = 1.27, p = .437, BF = 0.44), but here we do not see a significant positive correlation between learning in the first learning phase and learning pairs of the *same structure* in the second learning phase (r = .109, p = .620).

I conducted the same type of matched sample analysis as for Experiment 1a here (see Appendix C for details). As in Experiment 1a, the matched sample descriptively showed the same type of pattern as the full sample (Figure 3.3). However, the critical analysis of the difference between learning pairs of the *same* and of the *novel structure* for the matched implicit sample failed to reach significance (M_{diff} = 8.66, d = 0.22, t(22) = 1.05, p = .304, BF = 0.46).

To compare directly the effect of type of consolidation on implicit structure learning, I entered the data of participants with implicit knowledge from Experiments 1a, 3a, 3b, and 3c into a 4x2 ANOVA, with consolidation type (no consolidation, 12-h-sleep, 12-h-awake, and 24-h-sleep consolidation) and structure type (*same* or *novel structure*) as factors. The obtained results showed the typical pattern of a cross-over interaction with no significant main effects (consolidation type: F(3, 623) = 0.18, p = .910, BF = 0.003, η_p^2 = .0009; structure type: F(1, 623) = 1.52, p = .218, BF = 0.17, η_p^2 = .002) but a significant interaction (F(3, 623) = 7.43, p < .001, BF = 1979, η_p^2 = .03). Post-hoc tests revealed significant differences between the no-consolidation group (Exp. 1a) and the two asleep-consolidation groups (Exp. 3a and 3c), where the no-consolidation group showed stronger learning of *novel structure* (Exp. 1a vs. Exp. 3a: $p = .004$, $BF = 44.3$; Exp. 1a vs. Exp. 3c: $p = .012$; $BF = 12.7$), while the asleep-consolidation groups showed stronger learning of *same structure* pairs (Exp. 1a vs. Exp. 3a: $p = .015$, $BF = 8.8$; Exp. 1a vs. Exp. 3c: $p = .011$; $BF = 14.3$). No other significant differences were found.

As it has previously been reported that sleep can lead to explicitness of previously implicit representations (Fischer et al., 2006; Wagner et al., 2004; Zander et al., 2017) I tested for differences in the proportion of participants with explicit knowledge across experiments. X^2-tests showed no significant differences in the proportion of explicitness between the conditions with no consolidation (Experiment 1a) and the condition with 12-hours consolidation including sleep (Experiment 3a: $X^2(1, 392) = 0.054$, $p = .816$) or the condition with 24-hours consolidation including sleep (Experiment 3c: $\mathrm{X}^2(1, 387) = 0.166$, $p = .683$).

In Experiments 1a, 1b, 3a, 3b, and 3c participants with explicit knowledge showed the same descriptive pattern of higher performance for same over novel structure pairs. However, this difference failed to reach significance within these experiments. We, therefore, analyzed them again collapsed over all these experiments. The results showed an overall higher performance for the *same structure* pairs than *novel structure* pairs for explicit participants ($d = 0.34$, $t(233.99) = 2.56$, $p = .011$, $BF = 3.12$). Furthermore, over all experiments, explicit participants' performance for Phase 1 only significantly correlated with the learning of *same structure* pairs ($r = .41$, $p < .001$, $BF = 4{,}460$), not *novel structure* pairs ($r = .14$, $p = .137$, $BF = 0.62$).

Discussion

The results of Experiment 3c replicate the results found in Experiment 3a and, therefore, demonstrate that the implicit transfer of structural knowledge was in fact specific to sleep. This is further confirmed by jointly analyzing the data of experiments with all different consolidation conditions (Experiments 1a, 3a, 3b, 3c).

3.4 General Discussion

The research presented in this chapter studied the influence of consolidation on implicit structural learning and transfer by adopting the unsupervised transfer learning paradigm introduced in the previous chapter. Over three experiments, I found that after asleep, but not awake, consolidation participants with implicit knowledge are able to generalize structural knowledge. Following best practice in consolidation research, I confirmed that this effect is specific to sleep and cannot be explained by a time-of-day effect.

These results are in line with results previously reported in the literature showing abstraction based on consolidation (3.1.1) but extend them to the domain of unsupervised learning and implicit representations. This suggests that the same underlying mechanism, or at least the same logic of processing, is at play, spanning implicit and explicit representation and supervised and unsupervised learning. The results also support the central idea of this thesis, that what is canonically called statistical learning is part of a larger unsupervised learning system, incorporating more abstract conceptual information as well as low-level co-occurrence statistics. Combining the findings of this chapter with the structural novelty effect reported in the previous chapter clearly shows that there are complex interactions between representations on different levels of abstraction in unsupervised learning. Furthermore, we see support for the idea that consolidation and explicitness of knowledge are important moderators for these interactions.

The results of the current chapter are in line with the interpretation given in the previous chapter. The structural novelty effect found there was interpreted as a result of a structure-level interference, where the representational overlap between previously learned and newly encountered pairs leads to proactive interference, i.e., hindering the learning of the new pairs. Based on the findings of the current chapter, we can extend this interpretation by a consolidation-driven abstraction of the shared feature encoded by the representational overlap of the old pairs.

This emergence of representation on several hierarchical levels amounts to a factorization of the representation, representing the abstract feature of horizontality in its own right, allowing for its interference-free application to novel input. This parsimonious interpretation explains both the structural novelty effect in the absence of consolidation and the generalization behavior after consolidation with the representational overlap of learned pairs under the assumption of a restructuring of representations during sleep. This assumption seems reasonable considering previous empirical findings and theoretical accounts of consolidation (3.1.1) and the findings presented in the current chapter.

A hypothesis derived from linking the current findings to previous research on the effect of consolidation on memory and complementary learning systems is that for the participants with implicit knowledge, the statistical learning part, i.e., learning specific pairs, mainly depends on networks within the hippocampus (Schapiro, Turk-Browne, et al., 2017), while the sleep-dependent abstraction observed in the current experiments mainly depends on hippocampal-neocortical interactions (McClelland et al., 1995). Apart from testing this hypothesis directly using neural measurements appropriate for human sleep research (Uji & Tamaki, 2023), future studies could also go deeper into sleep physiology to study the effect of specific sleep stages (slow-wave-sleep vs. rapid-eye-movement-sleep) and investigate if the same patterns are found as for previous research focusing on more explicit representations (Chambers, 2017; Klinzing et al., 2019; Pöhlchen & Schönauer, 2020). A further related open question is how the quick abstraction and generalization are realized for participants with explicit knowledge and how that relates to the slower consolidation-dependent process observed for participants with implicit knowledge.

In the previous chapter, I reported a retroactive interference for participants with explicit knowledge, showing that a second training phase with novel items reduces their performance on items of the first training phase (comparison of Experiments 1a and 2a). Comparing the results of the conditions without consolidation (Experiment 1a) to the one with asleep consolidation (Experiment 3a), we do not have significant evidence for release of interference based on consolidation (although this comparison may have been underpowered). This is at odds with previous results showing a release from retroactive interference by asleep consolidation (Abel et al., 2023; Ellenbogen et al., 2006) for the domain of declarative memory. In contrast to the structure-level proactive interference reported for implicit participants in Experiment 1a (structural novelty effect) was released after asleep consolidation in Experiment 3a. This suggests overall important differences in how prior and novel representations interact for implicit and explicit representations. For participants with implicit knowledge, I found proactive interference specific to orientation and released by consolidation, while for participants with explicit knowledge, I found retroactive interference not specific to orientation and (probably) not released by consolidation. These findings might be related to previous research showing divergent consolidation trajectories in auditory statistical learning for participants with explicit and implicit knowledge (H. Liu et al., 2023). There, it was shown that explicit representations tend to decay over a 24-hour period, while implicit representations are strengthened. However, this study only compared a no-consolidation group to a 24-hour consolidation group. Therefore, It is not possible to deduce the role of sleep in their findings. Overall, the findings reported in this chapter support the argument of the previous chapter that tracking explicitness or implicitness of knowledge is critical in statistical learning research, as participants with explicit and implicit knowledge again show important quantitative and qualitative differences in their performance.

Interestingly, for the current domain, I did not replicate previous findings suggesting that a phase of asleep consolidation would lead from implicit to explicit representation (Fischer et al., 2006; Wagner et al., 2004; Zander et al., 2017) and, therefore increase the proportion of participants with explicit knowledge in conditions containing asleep consolidation. For our data the proportion of participants with explicit knowledge was independent of sleep. This leads to the question of which features or context variables of a task moderate whether or not such an effect is found.

The current results are in line with previous results suggesting that for the alternating serial reaction time (ASRT) task, learning of specific simple chunks happens online, during training sessions, while the acquisition of more complex higher-order rules crucially depends on offline phases (Quentin et al., 2021). In a similar setup, it was also demonstrated that previous exposure to such higher-order rules led participants to interpret subsequent new input in line with these rules (Kóbor et al., 2020). However, in the context of the ASRT task, higher-order learning describes second- or third-order temporal transitions between specific elements (i.e., non-adjacent associations), as compared to simple adjacent transitions. These are, therefore, directly observable in the input and not latent factors as the underlying orientation in the study presented in the current chapter. Furthermore, a benefit of (some types of) asleep consolidation has also been demonstrated in statistical learning using probabilistic input as compared to deterministic input (within chunk conditional probabilities of 1) (Durrant et al., 2013, 2011), in cross-modal transfer of statistical learning (Durrant et al., 2016), and for statistical learning in the presence of interference (McDevitt et al., 2022). Combining the findings of these previous studies with the findings presented in the current chapter might suggest that the critical boundary for consolidation dependence in unsupervised learning is not just between observable (statistical learning) and more latent (structure learning) features but between simple (adjacent, unimodal, specific, deterministic, and interference-free) and complex (non-adjacent, cross-modal, abstract, probabilistic, or interference-based) regularities.

Summarizing the results of the current and the previous chapter, we see that while participants with explicit knowledge can immediately generalize structural knowledge from one unsupervised learning context to another, participants with implicit knowledge show a structural novelty effect in immediate transfer, only generalizing after a phase of asleep, but not awake, consolidation. The results for the participants with implicit knowledge highlight the complex interactions of representations over different levels of abstraction and support the necessity of novel paradigms that can advance our understanding of how what is canonically called statistical learning is related to the unsupervised learning of more abstract, conceptual knowledge. Furthermore, critical moderating roles of consolidation and the explicitness of knowledge have been clearly demonstrated.

CHAPTER 4

Spatio-Temporal Visual Statistical Learning

Visual statistical learning is classically investigated in two independent lines of research: temporal and spatial VSL. However, visual input is not either spatially or temporally structured in real-world environments, but always both. In the current chapter, I present a series of five experiments using a novel spatio-temporal visual statistical learning paradigm based on moving spatially defined patterns in and out of the participants' view over time. I first validate that learning is possible in this new setup (Experiments 4a+b) and then show that participants use both the temporal regularities (Experiment 5c) and the perceived motion (5b) to learn the spatial patterns with different levels of noise (Experiment 5a).

4.1 Spatial and Temporal Regularities in Visual Statistical Learning

As real-world regularities in the visual input are not either spatial or temporal but always both, it has been suggested that understanding visual input must build on a combination of both of them (Gepshtein & Kubovy, 2000; Hochberg, 1968; Johansson, 1973; Rolls, 2012; Stone, 1998; Wallis & Rolls, 1997). Obviously, spatial regularities in one moment are not independent of spatial regularities in the next. Indeed, spatio-temporal stability is sometimes cited as a defining feature of objects and object cognition (Baillargeon, 2008; Piaget, 1954). What distinguishes work done within the statistical learning paradigms from previous work on the role of spatio-temporal regularities in visual input is the focus on unsupervised, implicit learning of stimuli in unsegmented input (see definition of statistical learning in Chapter 1.1.1).

However, only a small set of studies in visual statistical learning has started to explore the connection between spatial and temporal features of the visual input. Turk-Browne and

Scholl (2009) trained participants on typical spatial or temporal visual statistical learning paradigms but then tested participants using both spatial and temporal tests, regardless of the type of initial training. The results showed a significant transfer between spatial and temporal setups. This demonstrates that representations built during VSL are flexible enough to be applied in tests that vary from the original presentation mode without directly demonstrating spatio-temporal processing. Kirkham et al. (2007) showed that infants could learn spatio-temporal sequences defined by the order of a set of global positions. This demonstrates a form of spatio-temporal processing; however, it does not realize the same type of relative spatial association usually used in spatial VSL research. In short, the feature of global position can be seen as more akin to shape identity in usual temporal VSL setups as it realizes a unique stimulus that is predictive of the following unique stimulus. Tummeltshammer et al. (2017) showed that learning forward and backward transitional probabilities is constrained to a purely temporal setup and is not possible in a spatio-temporal setup. However, the specific experimental design used might have biased participants in the spatio-temporal condition to learn patterns spatially in one particular arrangement, as will be discussed in more detail in the discussion section of this chapter. Xu et al. (2023) showed that attention is not only guided by spatial and temporal regularities separately but also by spatio-temporal regularities. Although this is suggestive of general implicit spatial-temporal learning, it does not focus on learning reappearing chunks in unsegmented visual input, as is the focus of the current work. Finally, Yan et al. (2023) investigated the temporal associations formed between arrays containing spatial associations and found that participants formed predictions about temporal sequences based on spatial configurations rather than single objects. Although this is somewhat complementary to the approach taken in the current study, the specific paradigm used by Yan et al. is a strong deviation from setups usually referred to as statistical learning as it combined reinforcement and unsupervised learning, with partially segmented input and several hours of exposure (as compared to several minutes usually used in VSL experiments).

All of these studies contributed specific ways of combining space and time in visual statistical learning paradigms. What is missing, however, is a paradigm providing a systematic way of studying how spatio-temporal regularities are used in implicit learning, using typical properties of statistical learning paradigms to establish continuity with the rich existing SL literature.

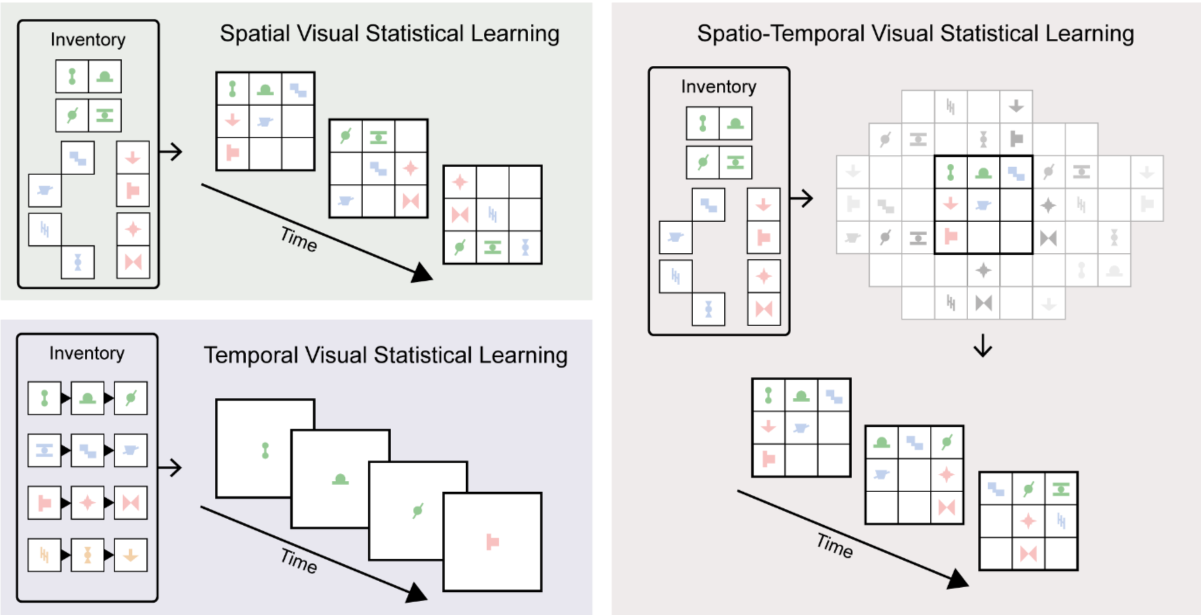

Figure 4.1 Spatio-temporal VSL Setup. The **top left panel** shows the standard spatial visual statistical learning setup. The **bottom left panel** shows the standard temporal visual statistical learning setup. The **right panel** shows the new spatio-temporal visual statistical learning setup (stVSL). There, the visual scenes are conceptualized as part of a larger visual environment, populated with the pairs of the inventory. Participants only see a 3x3 snapshot at a given time, akin to the 3x3 scene used in spatial VSL. However, the following snapshot is given by moving the shapes under the aperture by one grid cell, making the succession of snapshots temporarily dependent on each other, as compared to them being identical and independently distributed (i.i.d.) as in the spatial VSL setup.

4.2 The Spatio-Temporal Visual Statistical Learning Paradigm

As mentioned above, in the real world, regularities in the visual input are not either spatial or temporal but always both. This means that spatial regularities in one given moment are not independent of spatial regularities in the next moment. For example, the way the percept of a moving object changes in the eye of an observer provides important information about the spatio-temporal structure of the object, as specific views of the object predictably follow each other. Similarly, observing half an object emerging from behind an occluder allows for a clear prediction of what will be seen next, given knowledge of the object and the observed movement trajectory. However, the classic spatial VSL paradigm isolates spatial regularities, creating a situation where the presented scenes are supposed to be independent and identically distributed (i.i.d.). Although it is useful to be able to separate spatial and temporal regularities and study them independently in spatial VSL and temporal VSL, the interaction of the two during natural vision can only be understood by combining them. Therefore, I developed a new spatio-temporal VSL (stVSL) paradigm in which spatially defined patterns move in and out of the observer's view over time.

The single grid-based scenes presented in an independent and identically distributed (i.i.d.) fashion in spatial VSL studies can be considered as snapshots of a grid-like environment populated with the spatial patterns of the underlying inventory (Figure 4.1). For the new spatio-temporal VSL paradigm, instead of sampling random scenes from this environment, a part of the environment was visible through an aperture. Shapes moved in and out of the participants' view by periodically shifting the environment under the aperture one grid cell at a time. This also led to partial presentations of shape pairs when they were moving in and out. This added spatial uncertainty was balanced with the overall temporal coherence, which counters any violation of the spatial structure over time. I.e., although I see half a pair right now, the other half will reliably become visible after the next movement.

Using this new paradigm, I show that participants can learn the underlying spatial structure in this setup (Experiment 4a and 4b), that the extent of learning is not a function of added spatial noise alone (Experiment 5a), that participants crucially rely on the temporal regularity to recover the spatial patterns (Experiment 5c), and that participants use perceived motion as a cue to temporal regularities (Experiment 5b).

4.3 Proof-of-Concept Experiments

To test if learning is possible in the new spatio-temporal VSL paradigm, I conducted Experiment 4a using the same training structure and tests as a classic spatial VSL study (Fiser & Aslin, 2001) while letting the scenes unfold over time in the new spatio-temporal VSL fashion. For comparison, I also conducted an online replication of the classic experiment (Experiment 4b).

4.3.1 Experiment 4a: spatio-temporal VSL Proof-of-Concept

Participants

20 participants (6 female, mean age = 25, SD = 6.5) were recruited via prolific.co. The hourly compensation was £ 6.7. All participants had normal or corrected-to-normal vision. The sample size was based on the original study by Fiser and Aslin (2001). 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 were taken from Fiser and Aslin (2001) and consisted of 12 abstract black shapes on a white background. The shapes were grouped randomly to form six pairs (two horizontal, two vertical, and two diagonal) for each participant. 144 scenes were created by placing one horizontal, one vertical, and one diagonal pair in a 3x3 grid without any segmentation cues. Each shape's maximum horizontal and vertical extension was 50% of the size of one grid cell. As this was an online study, participants conducted it on their own computers using Google Chrome, Safari, or Opera browser. Only desktop and laptop computers were admissible, but no smartphones or tablets were. Stimuli were presented using custom JavaScript code built with the *jsPsych* library (version 6.1.0) (Leeuw, 2015). As participants used different devices (screen size and resolution), the visual angle of the shapes was not the same for all participants. Instead, the 3x3 grid extended over 600x600 pixels and was centered in the middle of the screen. The remaining screen outside the grid was empty (white).

Procedure

Participants first passively observed the familiarization phase before conducting the test phase. For the familiarization phase, participants received only minimal instructions, stating that they should pay attention to what was happening on the screen and that they would be asked simple questions about it later. The pair structure of the scenes was not mentioned.

In this experiment, the scenes moved in and out of the screen. Starting from one completely visible scene, the scene moved out by one grid cell at a time, while a new scene simultaneously moved in by one grid cell at a time. There were no segmentation cues between scenes, meaning that all the stimuli were part of one continuous stream for the participants. Each motion took .5 seconds and was animated as a constant speed translation along the horizontal or vertical axis. The image then stood still for two seconds between motions. The ordering of scenes was random for each participant, constrained so that no two identical pairs would be visible simultaneously. Participants saw left, right, up, and down motion. Periodically, the motion direction changed, going from one orientation, i.e., horizontal motion, to the other, i.e., vertical motion. A change of direction occurred after 6, 9, or 12 steps. Overall, all participants saw leftward, rightward, upward, and downward motions for the same number of steps.

As individual scenes were seen longer in Experiment 4a than 4b (they move in and out over several steps), Experiment 4a used only half of the original 144 scenes (balanced for pair frequency and co-occurrence). The familiarization phase of Experiment 4a took nine minutes.

The test phase consisted of 36 2-alternative forced choice (2AFC) trials. In each trial, participants saw a real pair and a foil pair after each other (randomized order, two-second presentation, and one-second inter-stimuli-interval) and indicated which of the two was more familiar based on the familiarization phase by pressing "1" or "2" on the keyboard. Overall, six foil pairs, two horizontal, two vertical, and two diagonal, were created by re-combining shapes from different pairs of the familiarization phase. Each real pair was tested once with each foil pair. After the test phase, participants answered open questions to assess their explicit knowledge of the pair structure (see Appendix A).

Results

Two participants were excluded from the analysis for having verbalizable explicit knowledge of the pairs. The remaining participants performed significantly above the chance level of 50%: $M = 57.7$, $SE = 2.4$, $t(17) = 3.18$, $p = .005$, $d = 0.75$, $BF = 8.8$ (see Figure 4.2). As in the previous chapters, all Bayes Factors (BF) reported were calculated using the *BayesFactor* R package (Rouder et al., 2012). In my interpretations, I conservatively counted results as significant if my criteria for both p-values ($< .05$) and *BF* (> 3) were met.

Discussion

The results showed that participants could learn pairs in the novel spatio-temporal VSL setup, therefore providing a first proof-of-concept for the feasibility of this paradigm.

4.3.2 Experiment 4b: Online Replication of Spatial VSL

In order to compare the stVSL proof-of-concept experiment with the previously used spatial VSL setup, I conducted an online replication of Experiment 1 of Fiser and Aslin (2001).

Participants

20 participants (7 female, mean age = 24.7, SD = 5.5) were recruited via prolific.co. The hourly compensation was £ 6.7. All participants had normal or corrected-to-normal vision. The sample size was based on the original study by Fiser and Aslin (2001). 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 identical to Experiment 4a.

Procedure

The procedure was identical to Experiment 4a, with the difference that the visual scenes were presented following the traditional spatial VSL setup (Fiser & Aslin, 2001) rather than the novel spatio-temporal VSL setup. Therefore, participants saw one scene after another for two seconds with a one-second inter-trial interval between them. The order of scenes was randomly chosen for each participant. An attention check was included as the static input may have been less engaging than the dynamic input in Experiment 1a. For the attention check, text appeared in the central cell of the grid, prompting participants to press the space bar. Simultaneously, five black squares appeared in randomly chosen cells of the grid to mimic the overall appearance of the used visual scenes. The attention check disappeared every 2 seconds and reappeared after .5 seconds. The number of times the attention check was shown before the space bar was pressed was recorded.

As individual scenes were seen for a longer duration in Experiment 4a than in 4b (because they are moving in and out over several steps), Experiment 4a used only half of the original 144 scenes (balanced for pair frequency and co-occurrence). Overall, the familiarization phase of Experiment 4a took nine minutes, while for Experiment 4b it took seven minutes and 12 seconds. Therefore, the amount of exposure was not completely identical for the two experiments. However, the goal was not to compare the two setups directly but to test whether participants can implicitly learn the pairs in the new stVSL setup.

Results

Two participants were excluded from the analysis for having verbalizable explicit knowledge of the pairs. The remaining participants performed significantly above the chance level of 50%: $M = 56.5$, $SE = 2.5$, $t(17) = 2.60$, $p = .019$, $d = 0.61$, $BF = 3.2$ (see Figure 4.2).

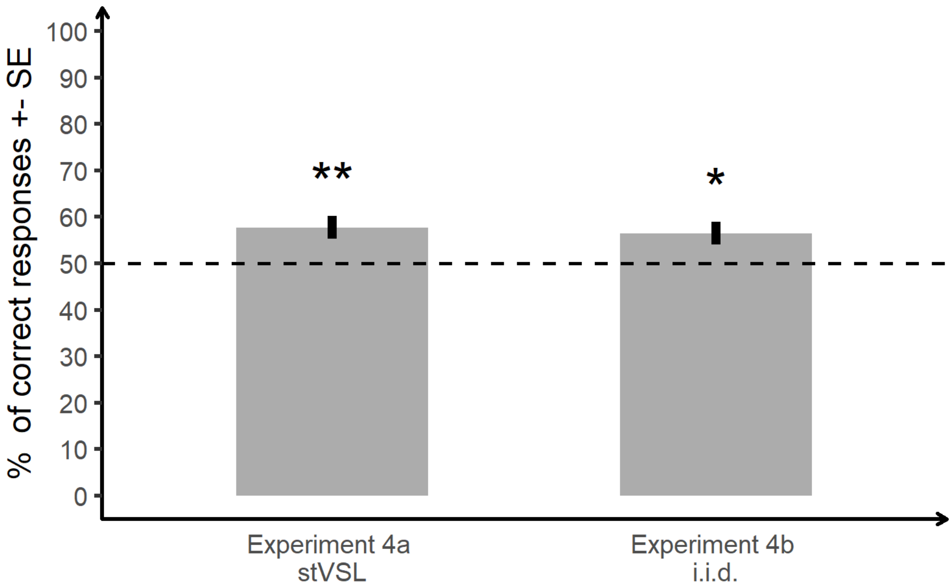

Figure 4.2 Experiments 4a and b Results. The y-axis represents the participants' mean performance on the 2-alternative forced choice (2AFC) trials, used as the measure of learning of pairs embedded in the familiarization stream. Error bars represent the standard error. The dotted line indicates the chance level of 50%. Stars represent the significance of the difference from chance. * $p < 0.05$; ** $p < 0.01$.

Discussion

Participants learned the spatial structure equally well in both the classic i.i.d. (Experiment 4b) and the novel spatio-temporal VSL setup (Experiment 4a). This supports the feasibility and usefulness of the novel stVSL paradigm.

4.4 The Role of Temporal Coherence and Perceived Motion

Experiment 4a demonstrated that learning is possible in the spatio-temporal VSL setup. But which features of this paradigm contributed to learning to what extent? One possible explanation is that participants did, in fact, use temporal coherence of the input to learn the spatial patterns despite the spatial noise (partial presentations) added by the spatio-temporal setup. An alternative explanation is that spatial VSL is simply robust enough to spatial noise to show similar test performance with or without this added noise. The experiments presented here tested this directly in two ways: introducing pairs with different noise levels within participants and realizing a contrast between the presence and absence of temporal coherence between participants. To better understand the novel paradigm, I further introduced a condition to investigate the role of perceived motion as a cue to temporal regularities.

Experiment 5a introduced pairs with different levels of spatial noise (e.g., number of partial presentations) to the stVSL paradigm to test whether the strength of learning is a simple function of the amount of noise. Experiment 5b removes the motion animations to probe the effect of temporal coherence in the absence of obvious perceptual cues. Experiment 5c removes the temporal coherence altogether to provide experimental evidence for its effect on learning spatial patterns.

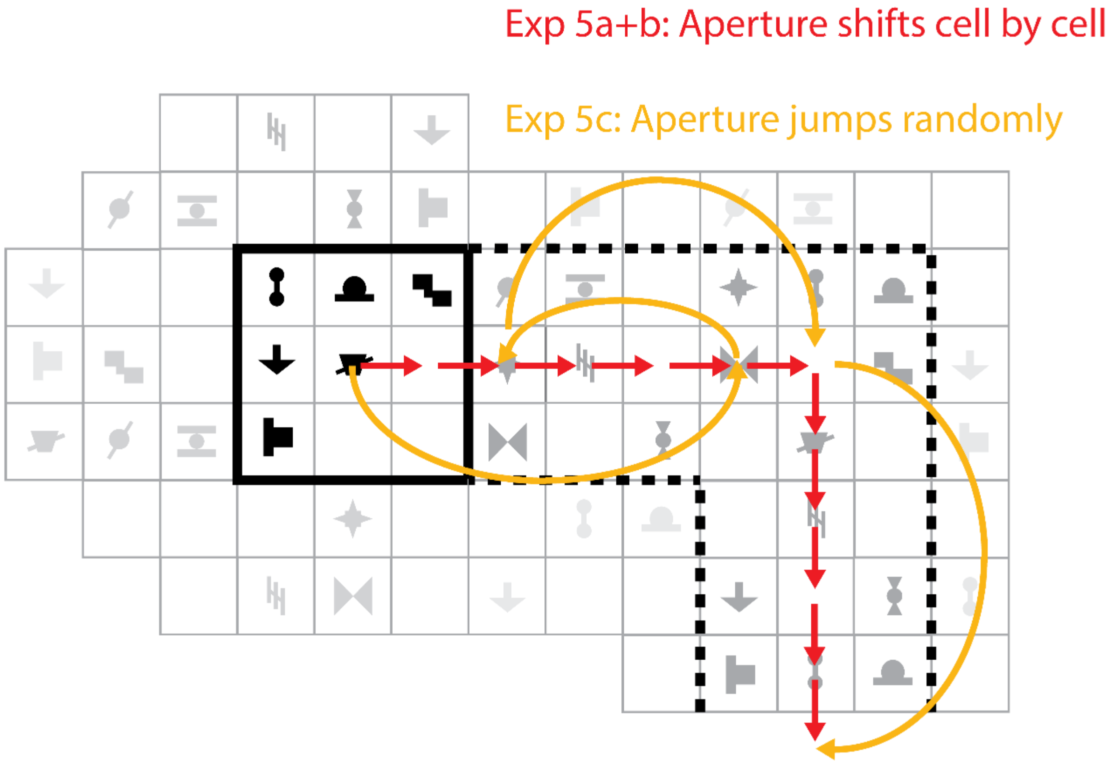

Figure 4.3 Presentation Modes in Experiments 5a-c. This graphic visualizes the temporal relationship of subsequent spatial patterns in Experiments 5a-c. In Experiments 5a and 5b, visualized in red, the visual aperture shifts by one cell at a time, leading to a sequence of temporally coherent scenes. In Experiment 5c, visualized in yellow, the visual aperture visits the same overall parts of the environment, leading to the same number of partial presentations of pairs. However, as visualized with the yellow arrows, the order of visual scenes is random and, therefore, not temporally coherent.

4.4.1 Experiment 5a: The Role of Spatial Noise in Spatio-Temporal VSL

Participants

88 participants (39 female, mean age = 26.9, SD = 8.0) were recruited via prolific.co. The hourly compensation was £ 6.7. All participants had normal or corrected-to-normal vision. The sample size was chosen to achieve a power of 80% for three parallel comparisons (i.e., alpha = .1666…) with medium effect sizes (cohen's $d = 0.5$) and rounded up to account for expected exclusions. 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 identical to Experiment 4a.

Procedure

The general procedure of Experiment 5a was identical to Experiment 4a, except for the specific movement directions. Participants no longer saw all directions of movements the same number of times; instead, participants were randomly assigned to one of two conditions, having more horizontal or more vertical movement.

In the horizontal condition, 75% of the movements were along the horizontal axis, with equal movements to the left and right. 25% of movements were along the vertical axis, with equal movements up and down. For horizontal movement, the change of movement direction occurred after nine, 12, 15, or 18 steps. For vertical movement, the change occurred after three or six steps. Through this biased movement pattern, horizontal pairs had more partial presentations than vertical pairs, as they were shown partially only during horizontal movement. This pattern of movement and partial presentations were inverted for the vertical condition. Referring to both conditions simultaneously, I speak of *parallel pairs*, being aligned with the predominant movement direction (horizontal pair in the horizontal condition), and *orthogonal pairs*, being orthogonal to the predominant movement direction (vertical pairs in the horizontal condition). In both conditions, *diagonal pairs* had the overall highest number of partial presentations, as they were shown partially during both horizontal and vertical movement. The conditional probabilities of the shapes of one pair, i.e., the probability that if one shape of a pair is visible, the other shape is visible as well, were .916 for the orthogonal pairs, .75 for the parallel pairs, and .6 for the diagonal pairs.

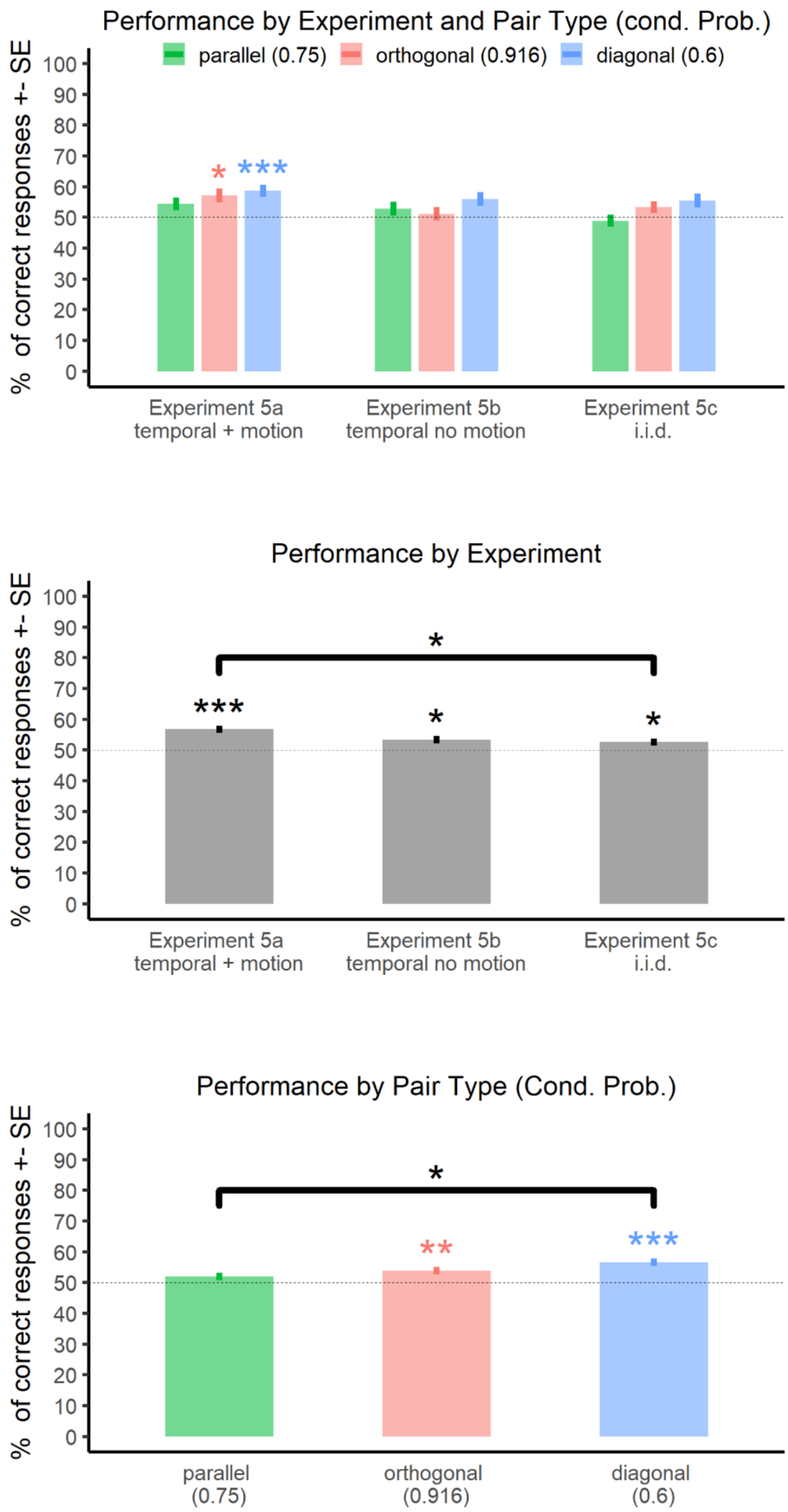

Figure 4.4 Experiments 5a-c Results. The y-axis represents the performance on 2AFC trials. Error bars represent the standard error. The dotted line indicates the chance level. Stars represent the significance of statistical tests: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.005$. **Top panel:** results of pair type by experiment. **Middle panel:** main effects of experiment averaged over pair types. **Lower panel**: main effects of pair type averaged over experiments.

Results

Before analysis, three participants were removed for response bias (bias for one of the two response buttons > 2.5 SD), and 15 participants were removed for acquiring explicit knowledge of the task structure. For data of explicit participants, see Appendix D.

One-sample t-tests showed that performance for the *orthogonal* (M = 57.1, SE = 2.3, $t(69)$ = 3.17, p = .018, d = 0.38, BF = 12.2) and the *diagonal* (M = 58.7, SE = 2.0, $t(69)$ = 4.4, p < .001, d = 0.53, BF = 535) pairs was significantly different from chance. The performance for the *parallel pairs* was not different from chance: M = 54.4, SE = 2.0, $t(69)$ = 2.2, p = .157, d = 0.26, BF = 1.2. The reported significant tests are correct for multiple comparisons over all three experiments (5a-c) using the Holm-Bonferroni method (Holm, 1979).

Combining the data for all three pair types, we see an overall significant difference from chance: M = 56.7, SE = 1.1, $t(69)$ = 5.96, p < .001, d = 0.71, $BF = 1.4*10^5$.

4.4.2 Experiment 5b: The Role of Perceived Motion in Spatio-Temporal VSL

Participants

89 participants (31 female, mean age = 33.4, SD = 10.9) were recruited via prolific.co. The hourly compensation was £ 6.7. All participants had normal or corrected-to-normal vision. The sample size was chosen to achieve a power of about 80% for three parallel comparisons (i.e., alpha = .1666…) assuming medium effect sizes (cohen's d = 0.5) and rounded up to account for expected exclusions. 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 identical to Experiment 4a.

Procedure

The only difference between Experiments 5a and 5b was that in 5b, the animated motion, a 0.5-second smooth transition of the shape between scenes, was removed and replaced with a 0.5-second blank screen. This means that the temporal coherence was the same as in 5a, i.e., if only one shape of a pair moved in on the last step, the other one will reliably move in on the next step. However, the strong perceptual cue to temporal coherence given by motion was removed. Comparing the results of Experiment 5a and 5b will, therefore allow us to tease apart the effects of the spatio-temporal co-occurrence statistics itself with the effect of perceived motion as a cue to temporal coherence. Additionally, as the static input may have been less engaging, an attention check was included during the familiarization phase, as in Experiment 4b.

Results

Prior to analysis, six participants were removed for response bias, one participant was removed for failing attention checks (response time to attention check > 3 SD), and 10 participants were removed for acquiring explicit knowledge of the structure of the task. For data of explicit participants, see Appendix D.

One-sample t-tests showed that performance for all of the three pair types was not significantly different from chance: *parallel* (M = 52.9, SE = 2.2, $t(71)$ = 1.3, p = .006, d = 0.15, BF = 0.29), *orthogonal* (M = 51.2, SE = 2.1, $t(71)$ = 0.5, p = .999, d = 0.06, BF = 0.15), *diagonal* (M = 56.0, SE = 2.2, $t(71)$ = 2.7, p = .064, d = 0.32, BF = 3.5). The reported significant tests are correct for multiple comparisons over all three experiments (5a-c) using the Holm-Bonferroni method (Holm, 1979).

Combining the data for all three pair types, we see an overall significant difference from chance: M = 53.4, SE = 1.2, $t(71)$ = 2.84, p = .006, d = 0.33, BF = 5.2.

4.4.3 Experiment 5c: The Role of Temporal Coherence in the Spatio-Temporal VSL

Participants

90 participants (30 female, mean age = 26.2, SD = 8.9) were recruited via prolific.co. The hourly compensation was £ 6.7. All participants had normal or corrected-to-normal vision. The sample size was chosen to achieve a power of about 80% for three parallel comparisons (i.e., alpha = .1666…) assuming medium effect sizes (cohen's $d = 0.5$) and rounded up to account for expected exclusions. 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 identical to Experiment 4a.

Procedure

The only difference between Experiment 5a and 5c was that in 5c, the temporal coherence of the stimuli presentation was removed (Figure 4.3). Participants still saw the same images which participants in Experiment 5a saw between movements. However, instead of moving them in and out of the grid by one cell at a time, they were temporally shuffled and presented without animations. This realizes visual input with the exact same level of spatial noise as in Experiment 5a but with an i.i.d. presentation instead of a temporally structured one. Additionally, as the static input may have been less engaging, an attention check was included during the familiarization phase, as in Experiment 4b.

Results

Prior to analysis, two participants were removed for response bias, two participants were removed for failing attention checks (response time to attention check > 3 SD), and 10 participants were removed for acquiring explicit knowledge of the structure of the task. For data of explicit participants, see Appendix D.

One-sample t-tests showed that performance for neither of the three pair types was significantly different from chance: *parallel* (M = 48.9, SE = 2.0, t(75) = -.55, p = .999, d = 0.06, BF = 0.146), *orthogonal* (M = 53.4, SE = 2.0, t(75) = 1.75, p = .338, d = 0.2, BF = 0.54), *diagonal* (M = 55.5, SE = 2.3, t(75) = 2.4, p = .112, d = 0.28, BF = 1.86). The reported significant tests are correct for multiple comparisons over all three experiments (5a-c) using the Holm-Bonferroni method (Holm, 1979).

Combining the data for all three pair types, we see an overall significant difference from chance according to t-test p-value, but not Bayes Factor: M = 52.6, SE = 1.2, t(75) = 2.18, p = .032, d = 0.25, BF = 1.18.

For further analysis, I entered Experiment 5a, 5b, and 5c into one 3x3 mixed-ANOVA with pair type (*parallel*, *orthogonal*, *diagonal*) as a within-subject factor and experiment (5a, 5b, 5c) as a between-subject factor. The results showed a significant main effect of pair type (F(2, 430) = 3.74, p = .025, η^2 = .012) and of experiment (F(2, 215) = 3.22, p = .042, η^2 = .012) but no significant interaction (F(4, 430) = 0.62, p = .649, η^2 = .004). Tukey's SHD post-hoc tests showed that the *diagonal pairs* were learned significantly better than the *parallel pairs* (p = .020) and that learning in Experiment 5c (i.i.d. condition) was significantly worse than in Experiment 5a (p = .043). The results for all three experiments are visualized in Figure 4.4.

4.4.4 Discussion of Experiments 5a-c

Over Experiments 5a-c, I found that participants learned pairs with low and high spatial noise in the temporally coherent condition (5a), whereas they learned them significantly worse or not at all without the perceived motion (5b) and without the temporal coherence (5c). The performance for the condition with temporal coherence but no perceived motion (5b) is descriptively between the other two conditions but not significantly different from either. This suggests a gradual effect where the effect of temporal coherence is strengthened by the perceptual cue to its presence. Overall, these results are direct experimental evidence (5a vs. 5c) that participants use temporal regularities to acquire spatial patterns, and they furthermore support the idea that perceived motion is an important cue to temporal regularity.

Surprisingly, participants learned the diagonal pairs better than the coherent pairs despite their lower conditional probability. This exploratory finding suggests that the learning of spatial regularities is not a simple function of spatial and temporal coherence but that other factors, such as prior biases, are at play.

Finally, this study was the first application of the novel stVSL paradigm, and all present experiments used the same arbitrary length of exposure in the familiarization phase. To generalize my findings, Supplementary Experiment 1 replicated Experiment 5a, using twice the amount of training. I found overall similar results in both experiments with no significant differences. See Appendix E for details.

4.5 General Discussion

In the current chapter, I presented a series of five experiments that established a novel spatio-temporal visual statistical learning paradigm in order to investigate how spatial and temporal regularities in the visual input are combined in implicit learning. The results showed that implicit learning is possible in such a setup, leading to a similar level of performance as purely spatial learning, and provided experimental evidence that participants used the temporal coherence of the input in the implicit learning of spatial patterns. The results furthermore suggested, without reaching statistical significance, that perceived motion is a crucial feature allowing participants to use the temporal regularities in the input for learning spatial patterns.

The current work goes beyond previous studies on spatial and temporal regularities in VSL in important ways. While Turk-Browne and Scholl (2009) showed that VSL is flexible enough to allow for temporal tests after spatial learning and spatial tests after temporal learning, their findings did not go beyond showing general associations between shapes of a chunk. Importantly, in their setup, there was no meaningful spatial structure to test for after temporal learning and no meaningful temporal structure to test for after spatial learning. In contrast to this, the current study is specifically designed to demonstrate the role of temporal coherence for the learning of spatially defined structure, therefore meaningfully connecting both domains in an ecologically relevant way. In their spatial context condition, Tummeltshammer et al. (2017) presented infants with spatial structure in a setup similar to the current stVSL paradigm. However, there are a number of key differences. Their null result for the specifics of the temporal ordering of stimuli aligns with my interpretation that this type of setup leads to the representation of spatial patterns. Still, it is noteworthy that their specific way of moving stimuli through the screen, using only one direction of motion for each participant, might have been a case that lends itself more readily to learning temporal structure. This is plausible as in this setup one shape of a pair always moves in first and the other moves in second. This is not the case in my setup, using all cardinal directions. However, in their setup, but not in mine, there is an additional strong cue for representing the structure spatially. The same shapes and shape pairs appeared on the screen multiple times simultaneously. This could lead to a kind of pop-out effect of the spatial regularity, potentially overshadowing all effects of the temporal regularity observed at the borders of the screen. These two effects, the connection of movement directions and temporal ordering, and the pop-out effect of multiple simultaneous presentations, could easily be studied in isolation and interaction using my new stVSL setup. Yan et al. (2023) investigated the temporal associations formed between arrays containing spatial associations and found that participants formed predictions about temporal sequences based on spatial configurations rather than single objects. The authors framed this as a statistical learning effect, which can be seen as complementary to the current work. While I study how temporal regularities affect the learning of spatial patterns, Yan et al. studied the role of spatial patterns in learning temporal patterns. However, as Yan et al. used a combination of reinforcement and unsupervised learning, and they used much more extensive training than usual in the SL literature (hours instead of minutes), their results can not be easily related to the statistical learning literature as a whole. However, studying the role of learned spatial patterns in learning temporal patterns in an unsupervised SL setup would, in fact, be complementary to the current work and advance our understanding of the role of spatio-temporal regularities in visual statistical learning.

A critical way in which the current study goes beyond previous attempts of studying the connection of spatial and temporal regularities in VSL is that it provides direct evidence of an effect of temporal regularities in implicitly learning spatial patterns (significant difference between Experiments 5a and 5c). This suggests that the process underlying spatial VSL does not consider only momentary spatial co-occurrence but also coherence over time. A related type of learning has been demonstrated in neural network models of unsupervised learning of invariant visual object representation by implementing a *trace learning rule* (Wallis & Rolls, 1997). Units in such models keep a short-term memory trace of their previous activation, which leads to weights updating based on a succession of stimuli. As different views of the same object follow each other more often than views of different objects, this leads to learning invariant object representations by associating the different versions (views) of objects encountered. Such simple neural implementations are one candidate for the mechanistic implementation of the learning process demonstrated behaviorally in the current study.

Another line of research from outside the VSL domain related to the phenomenon studied here, concerns the role of (partial) occlusion in visual perception. Under the heading of amodal completion, the majority of studies in this line investigate the role of stimuli-driven processes independent of top-down influences (Kanizsa, 1985; van Lier & Gerbino, 2015). Recently, a smaller number of studies started to investigate the role of prior knowledge in this process, using stimuli that can only be completed appropriately by knowing the structure of the whole stimuli (Hazenberg et al., 2014; Hazenberg & van Lier, 2016; Yun et al., 2018). This is related to the current work, as the partial presentations of shape pairs moving in and out of the visible aperture in my setup can be considered instances of partial occlusion. In the current form of my paradigm, I am considering the role of such partial occlusion during learning, while the amodal completion studies cited above consider the role of occlusion during object perception. This can be considered two complementary cases, and future work could aim to combine both in one setup. I.e., studying the effect partial presentations during learning have on the perception of partially occluded stimuli during testing. The topic of occlusion in spatio-temporal VSL will be revisited in the next chapter.

The research presented in the current chapter illustrates both the feasibility and the necessity of the novel spatio-temporal VSL paradigm. It does so by showing that people can learn in this setup (feasibility) and that temporal coherence is directly used for learning spatial patterns (necessity). The current results showed that it is not merely spatial or temporal co-occurrence statistics that drive VSL, but that spatio-temporal regularity as a whole is of crucial importance. How does this relate to the research in the previous two chapters, which focused on how a combination of low-level co-occurrence statistics and top-down influences of prior knowledge drives VSL? Taken together, we can see that the human unsupervised learning system works by combining low-level spatio-temporal statistics and higher-level features. Going beyond this, the next chapter will demonstrate more intricate connections between the spatio-temporal VSL paradigm and the study of higher-level features in VSL. We will see that a perceived overall motion direction and overall item arrangements can exert top-down influences on what is implicitly learned or inferred from visual input. Furthermore, a variation of the stVSL paradigm will show, in line with the results presented in Chapter 2, that explicitness of knowledge is a crucial moderator of how participants can apply their acquired knowledge.

CHAPTER 5

Higher-Level Influences in Spatio-Temporal Visual Statistical Learning

The main argument of this dissertation is that what is usually called visual statistical learning is part of a larger unsupervised learning system that operates by combining lower-level spatio-temporal co-occurrence statistics with higher-level top-down biases, flexible combining the multitude of available features across levels to achieve a congruent and comprehensive, yet parsimonious interpretation of the world. While the previous chapters treated spatio-temporal VSL and hierarchical VSL separately, the current chapter highlights their intertwined nature by demonstrating that (1) top-down biases can not only be abstracted from specific chunks (as demonstrated in Chapters 2 and 3) but can also be induced by higher-level spatio-temporal statistics, and that (2) differential behavioral outcomes for participants with explicit and implicit knowledge arise not only in VSL based transfer-learning (as demonstrated in Chapters 2 and 3) but also in a prediction-based variant of the spatio-temporal VSL paradigm.

5.1 Biases Induced by Higher-Level Features

In Chapters 2 and 3, I demonstrated that participants could develop biases about the type of structures present in a learning context based on abstracting and generalizing common features from multiple patterns learned during visual statistical learning. In addition to these learned biases, the previous chapters also presented evidence of some pre-existing biases, such as biases for horizontality over verticality (Chapter 2) and a potential bias for diagonal structures during stVSL (Chapter 4). In the current chapter, this is augmented with a third type of bias: biases induced by higher-level spatio-temporal features of the input statistics.

The experiments presented in the current sub-chapter identify two relevant higher-level features that can induce biases: the perceived *overall motion direction* and the perceived *overall shape arrangement*. I refer to them as *higher-level* here as they transcend the item-item co-occurrence statistics usually studied in VSL experiments and instead describe the overall input, i.e., properties that each item has (e.g., motion direction) or properties that describe the collection of all present items jointly (e.g., shape arrangement). The influence of such features has so far been largely neglected in VSL research, with a few exceptions discussed in this chapter's General Discussion section (5.3.1).

To enable the systematic investigation of the effect of such higher-level features on VSL, two major changes to the spatio-temporal VSL paradigm introduced in the previous chapter were necessary (see Figure 5.1). First, each participant now saw motion along only one orientation: horizontal or vertical motion. Second, an occluder was overlaid on the central grid cells, orthogonal to the motion direction. The combined effect of these two changes was that participants now only perceived motion along a single orientation and only saw shapes next to each other along a single, orthogonal orientation. This dissociation allowed for directly comparing these two higher-level spatial and temporal features by measuring participants' preferences for horizontal or vertical pairs. An additional effect of these changes was that some pairs were now only presented temporally, i.e., their shapes always followed each other but were never visible next to each other, while others were only presented spatially, i.e., their shapes always were visible next to each other but never followed each other. The role of perceived motion in this setup is specifically targeted by contrasting this setup with and without motion animations between Experiments 6a and 6b.

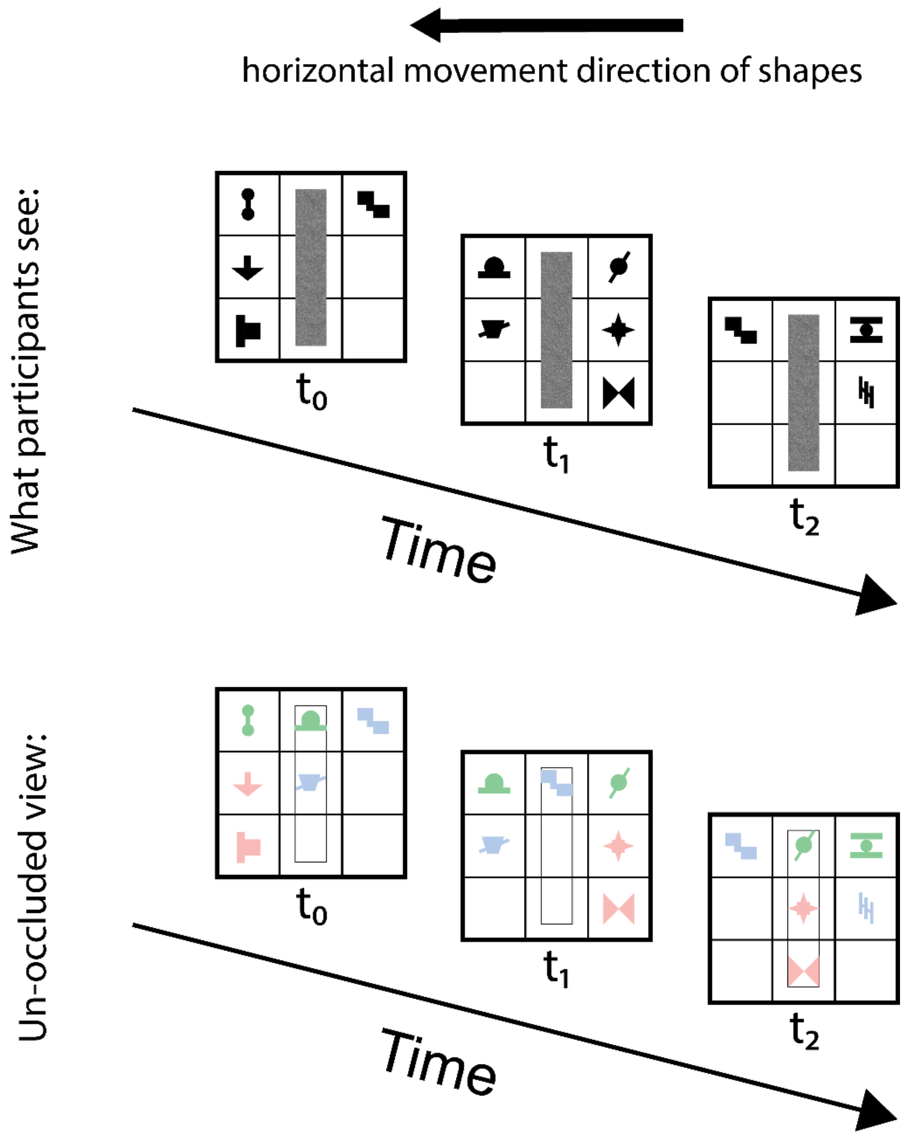

Figure 5.1 Experiments 6a and b Setup. Experiment 6a+b introduces two key changes to the basic stVSL setup. (1) Participants see only either horizontal or vertical movement (counter-balanced between groups). (2) An occluder is overlaid over the three central grid cells perpendicular to the movement direction. The combined effect of these two changes is that pairs aligned with the movement direction are now presented only temporally, while pairs perpendicular to the movement direction are only presented spatially. The graphics visualize the condition using only horizontal motion. The **top panel** directly shows what participants see in the experiment. The **bottom panel** is only for illustration purposes as it highlights the underlying structure by color coding pairs and making the occluder transparent.

5.1.1 Experiment 6a: Biasing VSL by Perceived Motion

Experiment 6a tested the effect of overall perceived motion direction and directly compared the effect of spatial and temporal presentation on learning spatially defined patterns. This was enabled by changes in stimuli presentations and the addition of novel tests. These tests allowed probing whether participants learned only general associations between pairs or also their specific spatial arrangements and whether they developed biases about types of arrangements.

Participants

132 (60 female, mean age = 27.1, SD = 11.1) participants were recruited via prolific.co. The hourly compensation was £ 6.7. The sample size calculation was built on the previous experiments but assumed a lower alpha level of .005 to account for the higher number of multiple comparisons (more tests in this experiment) and assumed a higher rate of exclusions. The study was approved by the Psychological Research Ethics Board of the Central European University, and all participants provided informed consent.

Materials

This experiment used the same materials as the stVSL experiments in Chapter 4. In addition, a rectangular occluder image was created, which consists of 1/f-noise and has 50% of the width and height of the grid cells. This way, if placed in the center of a grid cell, it covers the central part so that the shape present in the cell is not visible. This study was again conducted online, using the same setup as described for Experiment 4a in Chapter 4.

Procedure

The general procedure was built on Experiment 4a of Chapter 4 but introduced two changes. First, every participant now saw motion along only one orientation, i.e., only horizontal motion (left, right) or vertical motion (up, down). Participants first saw motion in one direction for 60 steps, then motion in the opposite direction for 60 steps, then 48 steps in the first motion direction, and finally 48 steps in the second motion direction. At each motion direction change, an attention check appeared, identical to the one used in Experiment 4b of Chapter 4.

Second, occluders were overlaid over the three central grid cells perpendicular to the movement direction, so the shapes within those cells were not visible (Figure 5.1). The combined result of these two changes was that *orthogonal pairs* (pairs orthogonal to the motion direction) were now only presented spatially, without any temporal coherence. I.e., both shapes of those pairs were always fully visible together, and they never followed each other. *Parallel pairs* (pairs aligned with the motion direction) were now never visible next to each other in the still images between movements (2 sec) and were only partially visible next to each other during the short movement period (0.5 sec). However, the shapes had perfect temporal coherence, i.e., they always followed each other in the same position of the grid. *Diagonal pairs* had the same type of temporal coherence and partial co-occurrence as parallel pairs, but the shapes did not follow each other in the same grid cell but in a different spatial arrangement. Furthermore, the higher-level statistics describing the scenes transcending item co-occurrences were altered by these changes so that participants only saw motion along one orientation (horizontal or vertical), and they only saw shapes next to each other along one orientation (horizontal or vertical) orthogonal to the motion direction. This orthogonality allows a direct comparison of the effects of these two manipulations.

After the nine-minute familiarization phase, participants completed a 2-alternative forced-choice (2AFC) test phase. Before completing the same real pair vs. foil pair tests (called *standard learning trials* here) as in the other experiment, participants completed test trials focusing on spatial arrangement. In *spatial learning trials* test trials, both options consisted of the shapes of the same real pair but were presented once in their correct spatial arrangement (e.g., horizontal) and once in an opposite arrangement (e.g., vertical). Diagonal pairs were tested against themselves in a different diagonal arrangement. The same logic was applied to the foil instead of the real pairs in *bias trials*. Note that in those trials, there was no correct orientation (i.e., no orientation seen in the familiarization phase). Therefore, there was no correct answer, and the question was if participants had overall biases to choose either horizontal or vertical options.

Results

Prior to analysis, two participants were removed for response bias (bias for one of the two responses buttons > 2.5 SD), two participants were excluded for failing the attention check (having the attention check message appear at least ten times over all three instances of the attention check), and 18 participants were removed for acquiring explicit knowledge of the structure of the task (for data of explicit learners see Appendix D).

Standard Learning Trials. One-sample t-tests showed that in the *standard learning trials* performance for the *parallel pairs* was significantly above chance: $M = 58.2$, $SE = 2.4$, $t(109) = 3.4$, $p = .004$, $d = 0.33$, $BF = 23.5$. The performance for the *orthogonal* ($M = 48.6$, $SE = 2.6$, $t(109) = -0.52$, $p = .692$, $d = 0.05$, $BF = 0.12$) and *diagonal* ($M = 54.5$, $SE = 2.3$, $t(109) = 1.99$, $p = .148$, $d = 0.19$, $BF = 0.70$) pairs was not different from chance. These results suggest that participants learned only the temporally presented parallel pairs.

Spatial Learning Trials. One-sample t-tests showed that in the *spatial learning trials*, performance for the *parallel pairs* was significantly above chance: $M = 65.5$, $SE = 2.7$, $t(109) = 5.68$, $p < .001$, $d = 0.54$, $BF = 1.1*10^5$. The performance for the *diagonal pairs* was not different from chance: $M = 52.5$, $SE = 2.6$, $t(109) = 0.95$, $p = .692$, $d = 0.09$, $BF = 0.16$. The performance for the *orthogonal pairs* was significantly below chance: $M = 36.8$, $SE = 2.6$, $t(109) = -5.2$, $p < .001$, $d = 0.49$, $BF = 1.2*10^4$.

Bias Trials. In the bias trials, there were no correct response options. The data for all *parallel* and *orthogonal* foil pairs was considered together and scored for choosing the parallel or orthogonal option. For every participant, the proportion of parallel choices was expressed in percent and then subtracted by 50 to get a measure of bias away from a chance level of zero. Positive bias suggests more parallel choices, while negative bias suggests more orthogonal choices. One-sample t-tests showed that participants chose the parallel options significantly more often: $M = 8.0$, $SE = 2.0$, $t(109) = 3.99$, $p < .001$, $d = 0.38$, $BF = 149$. The trials for diagonal pairs were not considered in the analysis. The reported significance tests above are correct for multiple comparisons over all three test types using the Holm-Bonferroni method (Holm, 1979).

For further analysis, the *standard learning trials* and *spatial learning trials* data were entered into one 2x3 mixed-ANOVA with test type (*standard learning*, *spatial learning*) and pair type (*parallel*, *orthogonal*, *diagonal*) as within-subject factors. The results showed a significant main effect of pair type ($F(2, 545) = 25.6$, $p < .001$, $\eta^2 = .19$) and a significant interaction ($F(4, 545) = 7.4$, $p < .001$, $\eta^2 = .06$). The main effect of test type was not significant, $F(1, 545) = 1.28$, $p = .26$, $\eta^2 = .01$. Tukey's SHD post hoc tests showed significantly higher performance for the *parallel pairs* than the *orthogonal pairs* in the *standard learning trials* ($p = .025$), and in the *spatial learning trials* ($p < .001$). The diagonal pairs showed significantly higher performance than the orthogonal pairs only in the *standard learning trials* ($p < .001$). The parallel pair showed significantly higher performance than the diagonal pair only in the *spatial learning trials* ($p = .001$).

A direct comparison was conducted to test whether the high deviation from chance in the *spatial learning trials* was based solely on the bias also measured in the *bias trials* or if it additionally included knowledge about the orientation of the specific pairs. For this purpose, the results for the *parallel* and *orthogonal pairs* in the *spatial learning trials* were separately transformed into a measure of deviation from chance, as described for the *bias trials*. Paired t-tests showed that the deviation from chance for the *parallel pairs* ($t(109) = -2.60$, $p = .021$, $d = 0.30$, $BF = 2.6$) but not the *orthogonal pairs* ($t(109) = -1.92$, $p = .058$, $d = 0.22$, $BF = 0.62$) was significantly higher than the bias measured in the *bias trials*. This suggests that participants had knowledge about the actual orientation of pairs they have learned in addition to a general bias. The results for all test types are visualized in Figure 5.2.

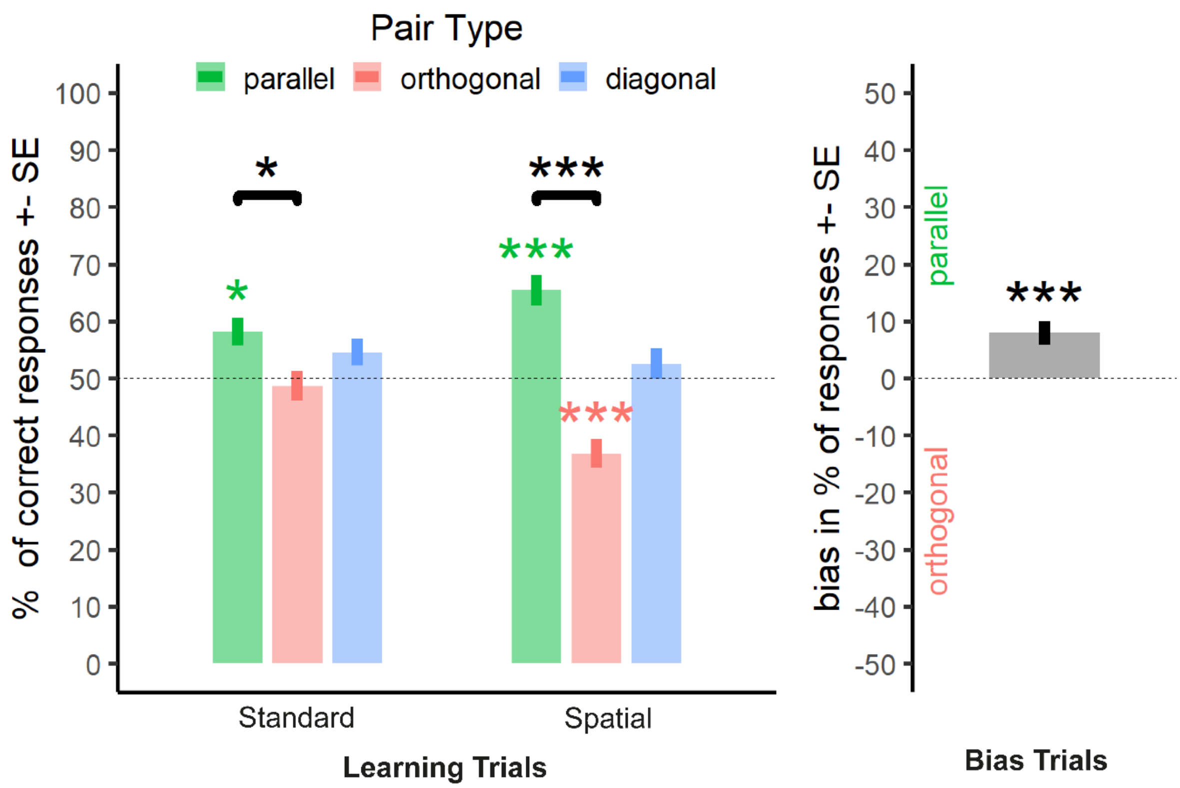

Figure 5.2 Experiment 6a Results. The y-axis represents the participants' mean performance on 2AFC trials. Error bars represent the standard error. The dotted line indicates the chance level of 50% or 0%. Stars represent the significance of the difference from chance. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.005$. The *standard learning trials* was a standard learning test using one real pair from the training phase and one foil pair created by combining shapes of two real pairs. It measures learning of item co-occurrence. The *spatial learning trials* test showed the same real pair twice. Once in its correct orientation and once rotated by 90°. It measures learning of the spatial arrangement of learned pairs. The *bias trials* test showed the same foil pair twice. Once horizontally and once vertically. There is no correct response, and it measures bias for one of the orientations.

Discussion

In this setup, participants only reliably learned the temporally presented parallel pairs. The results of the *bias trials* suggest that participants developed an overall bias to choose options aligned with the motion direction perceived during the training, independent of the orientation in which a pair appeared originally. Additionally, we see that participants also have knowledge about the actual orientation of pairs they have learned.

These results clearly show that to understand statistical learning, we need to consider top-down effects of higher-level biases. However, the exact nature of the interaction between learning of specific pairs, the perceived motion direction, and the observed overall bias remain unclear. Did participants preferably learn temporally presented pairs and then generalize the learned structure to the tests about orientation bias? Or did the perceived motion directly induce a bias that can explain both the preferred pair learning and the bias in orientation tests? Experiment 6b was designed to answer this question.

Regardless of the underlying mechanism, we see the strength such biases can have in visual statistical learning, as participants showed no sign of learning the spatially defined orthogonal pairs, even though these pairs were regularly seen in the input with only one other adjacent shape. Treating statistical learning as being exclusively based on lower-level co-occurrence statistics would have wrongly predicted strong learning for these pairs.

5.1.2 Experiment 6b: Biasing VSL by Spatial Arrangements

Was the effect in Experiment 6a driven by the underlying spatio-temporal co-occurrence statistics, i.e., a preference for temporally presented pairs, or by an overall feature of the stimuli presentation, i.e., the perceived horizontal or vertical motion? Experiment 6b was designed to answer this using the same spatio-temporal structure as Experiment 6a but removing the motion animation.

Participants

117 (52 female, mean age = 30.4, SD = 11.9) participants were recruited via prolific. Co. The hourly compensation was £ 6.7. The sample size was chosen to match that of Experiment 6a. The study was approved by the Psychological Research Ethics Board of the Central European University, and all participants provided informed consent.

Materials

This experiment used the same materials as Experiment 6a.

Procedure

The procedure was identical to Experiment 6a, apart from the movement animation. Whereas in Experiment 6a, between the 2-second stimuli displays shapes moved in and out of the visible aperture with a .5-second animated movement, in Experiment 6b, the animation is replaced by a .5 blank screen. Note that this manipulation did not alter the spatial and temporal structure. Parallel pairs were still presented only temporally, and orthogonal pairs were still presented only spatially.

Results

Prior to analysis, two participants were removed for response bias, six participants were excluded for failing the attention check (having the attention check message appear at least ten times over all three instances of the attention check), and seven participants were removed for acquiring explicit knowledge of the structure of the task (for data of explicit learners see Appendix D).

Standard Learning Trials. One-sample t-tests showed that in the *standard learning trials* the performance for all of the pairs was not different from chance: *parallel* (*M* = 51.3, *SE*

= 2.5, $t(99) = 0.51$, $p = .99$, $d = 0.05$, $BF = 0.13$), *orthogonal* ($M = 48.5$, $SE = 2.9$, $t(99) = -0.51$, $p = 1.00$, $d = 0.05$, $BF = 0.126$), *diagonal* ($M = 50.0$, $SE = 2.9$, $t(99) = 0.00$, $p = 1.00$, $d = 0.00$, $BF = 0.11$). Overall, these results suggest that participants did not reliably learn any pairs in this experiment.

Standard Learning Trials. One-sample t-tests showed that in the *spatial learning trials*, the performance for all of the pairs was not different from chance: *parallel* ($M = 47.0$, $SE = 2.7$, $t(99) = -1.09$, $p = 1.00$, $d = 0.11$, $BF = 0.198$), *orthogonal* ($M = 56.5$, $SE = 2.7$, $t(99) = 2.41$, $p = .107$, $d = 0.24$, $BF = 1.7$), *diagonal* ($M = 51.5$, $SE = 2.5$, $t(99) = 0.59$, $p = 1.00$, $d = 0.06$, $BF = 0.13$).

Bias Trials. Data for this test was converted to a measure of bias away from chance, as in Experiment 6a. One-sample t-tests showed that participants chose the orthogonal options significantly more often: $M = -9.0$, $SE = 2.0$, $t(99) = -4.55$, $p < .001$, $d = 0.46$, $BF = 1{,}051$.

To test whether the significant difference between *parallel* and *orthogonal pairs* in the spatial learning trials, found in Experiment 6a along the direction of the overall bias (results of *bias trials*), is also present here, a paired t-test was performed. The results showed significantly higher performance for the *orthogonal pair* trials: $t(99) = -2.16$, $p = .033$, $d = 0.35$, $BF = 1.03$. A direct comparison of both measures was conducted to test whether the results of the *spatial learning trials* were in line with the bias measured in the *bias trials*. For this purpose, the results for the parallel and orthogonal pairs in the *spatial learning trials* were separately transformed into a measure of deviation from chance, as described for the *bias trials*. Paired t-tests showed that neither the deviation from chance for the *parallel pairs* ($t(99) = -2.20$, $p = .061$, $d = 0.25$, $BF = 1.1$) nor for the *orthogonal pairs* ($t(99) = -0.87$, $p = .389$, $d = 0.11$, $BF = 0.16$) were significantly different from the bias measured in the *bias trials*. The reported significant tests above are correct for multiple comparisons over all three test types using the Holm-Bonferroni method (Holm, 1979). The results for all test types are visualized in Figure 5.3.

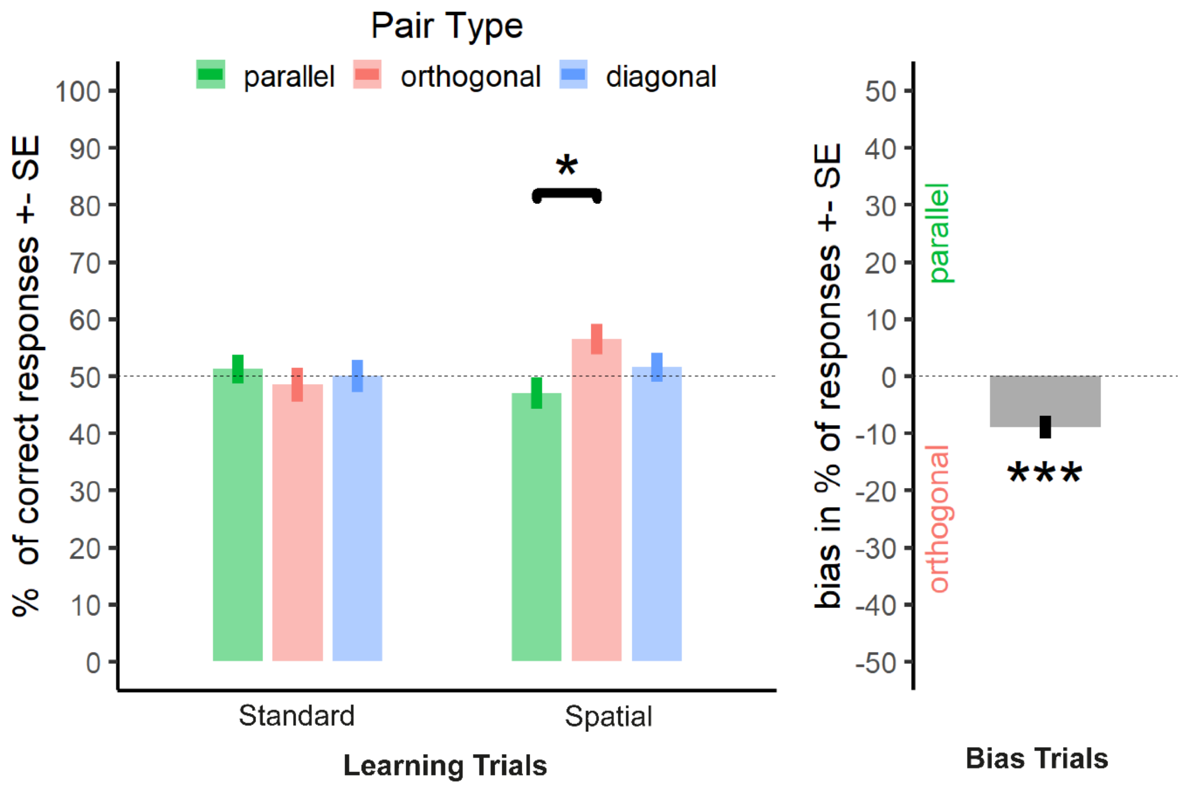

Figure 5.3 Experiment 6b Results. The y-axis represents the participants' mean performance on 2AFC trials. Error bars represent the standard error. The dotted line indicates the chance level of 50% or 0%. Stars represent the significance of the difference from chance. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.005$. The **Standard learning trials** was a standard learning test using one real pair from the training phase and one foil pair created by combining shapes of two real pairs. It measures learning of item co-occurrence. The **Spatial learning trials** showed the same real pair twice. Once in its correct orientation and once rotated by 90°. It measures learning of the spatial arrangement of learned pairs. The **Bias trials** showed the same foil pair twice. Once horizontally and once vertically. There is no correct response, and it measures bias for one of the orientations.

Discussion

I found no learning of specific pairs in this setup. This suggests that the learning of temporally presented parallel pairs in Experiment 6a was not solely based on the temporal regularity but critically influenced by the observed motion. The instructions for Experiments 6a and 6b were identical, stating that shapes were moving in and out. Therefore, it can be assumed that participants in 6b knew about the movement without directly observing it.

While in Experiment 6a, participants showed an overall bias to choose options aligned with the motion direction, in Experiment 6b, participants preferred options perpendicular to it. This suggests that in Experiment 6a, the overall bias was induced by the observed motion. The opposite bias observed in Experiment 6b might be due to the shapes being seen next to each other perpendicular to but not along the movement direction, i.e., due to the overall perceived shape arrangement. This suggests a flexible and parsimonious acquisition of biases based on the salience of the inducing features. In Experiment 6a, the feature apparently inducing the bias in 6b, overall shape arrangement, is present; however, it seems to be overshadowed by the feature of perceived motion.

5.2 Implicit and Explicit Online Prediction in Spatio-Temporal VSL

Traditional studies in visual statistical learning separate learning and testing into two different phases of the experiment. However, researchers have developed different online measures of VSL, trying to capture learning while it happens. In this sub-chapter, I first give an overview of the online measures previously employed in VSL research and then introduce a spatio-temporal VSL-based online measure that shows strikingly different results for participants with explicit and implicit knowledge.

Siegelman et al. (2018) introduced an online reaction time measure for temporal VSL based on self-paced progression through the familiarization phase. As usual in temporal VSL, participants see one shape after another on the screen, where some shapes form triplets. However, there is no fixed presentation rate, and participants move to the next shape via button press. The idea is that reaction times for predictable stimuli, the second and third shapes of a triplet, will be faster than those for unpredictable stimuli, the first shape of a triplet. Indeed, the authors found such an effect, which increased throughout the learning phase, constituting a measure of the trajectory of VSL. Crucially, this study utilized explicit task instructions, telling participants about the presence of predictable patterns before the experiment. As discussed and empirically demonstrated in Chapter 2, explicit instructions can alter learning outcomes in SL. Specifically, it was previously demonstrated that in temporal VSL tasks, instructions are only irrelevant for very fast presentation rates (Bertels et al., 2015), which were not achieved in this study due to the self-paced nature of the experiment. Based on this, the measure introduced by Siegelman et al. (2018) is most likely a measure of explicit learning during VSL.

Batterink and Paller (2017) introduced an EEG-based measurement of auditory statistical learning based on *neural entrainment*. The main idea here is that rhythmically presented stimuli will cause a rhythmical brain response at the same frequency. This is obvious and can be explained as purely perceptual for the syllable frequency in an auditory SL task. However, neural entrainment on the triplet frequency can only be explained by triplet-based processing and, therefore, successful segmentation of the input stream. In several studies, this entrainment on the triplet frequency was demonstrated (Batterink et al., 2023; Batterink & Paller, 2017; Choi et al., 2020; Moreau et al., 2022), therefore establishing an online, EEG-based, task-less measurement of temporal statistical learning. The same general idea of using neural entrainment as a measure for temporal SL has also been realized with intracranial recordings in patient populations (Henin et al., 2021; Sherman et al., 2023).

Dale et al. (2012) used computer mouse movements as a measure for the learning of temporal sequences of global positions. They found that while participants could learn explicitly or implicitly, explicitness was correlated with predictive behavior. Similarly, Sznable et al. (2023) used a temporal VSL task and found again that predictive behavior, but also predictive brain responses, were specific to participants with explicit knowledge.

All of the studies discussed above focused on measuring temporally defined structures in statistical learning. In contrast, the study presented in this sub-chapter employs a novel task to measure statistical learning of spatially defined structures online.

5.2.1 Experiment 7: An online prediction-based measure for stVSL

Participants

47 participants (19 female, mean age = 27.2, SD = 7.1) gave informed consent prior to the experiments. Participants were university students in Budapest, Hungary, who received 1,500 Hungarian Forint for their participation. As in the previous experiments, participants conducted the experiment at home on their own devices. The study was approved by the Hungarian United Ethical Review Committee for Research in Psychology (EPKEB), and all participants provided informed consent.

Materials

This experiment used the same materials as Experiment 4a in Chapter 4.

Procedure

The basis of the procedure was the spatio-temporal VSL setup introduced for Experiment 4a in Chapter 4. Additionally, there was an ongoing prediction task, querying participants about their beliefs about which shapes would move in on the next step throughout the training phase (see Figure 5.4). Every two to four steps (i.e., animated motions), a blue arrow appeared outside the grid, next to one grid cell, pointing at it for 500 ms. This was immediately followed by a single shape from the set of 12 shapes in the same position as the blue arrow. Participants had to respond by pressing a button, indicating whether they believed that this shape would move into the visible grid from this position during the next step. The cue and query shape position was always chosen to be next to a currently partially presented shape pair along the movement direction so that the next shape moving in was always predictable. The shape presented outside the grid would move in on the next step in 50% of the trials. Overall, there were 72 queries, using each shape an equal number of times as correct and incorrect options. The query shape presented outside the grid was always chosen to be a shape not currently visible in the grid. Participants had 1000ms to respond to the query. If they failed to respond within that time, a warning appeared telling them to respond more quickly on the following trials. For half of the participants, pressing "j" indicated believing that the shape would move in, and pressing "f" indicated believing that it would not move in. For the other 50% of the participants, this mapping was inverted. There was no fixed order of individual shapes or of correct/incorrect trials. However, the choice of the next query was biased for a higher probability of shapes that had appeared less frequently so far and for trial type (correct/incorrect) that had appeared less frequently. As for the previous experiments, the "path" of participants, i.e. the specific stream of shape combinations, was created by ordering pre-existing scenes and moving them in and out of the visible 3x3 grid without segmentation cues. However, as the specific constraints used for the task (equal number of appearances of shapes, no use of the shapes present in the currently visible scene) strongly reduced the number of admissible paths, a loop created paths until one in line with the constraints was found for each participant.

The training phase was followed by the same 2AFC test trials and post-experiment questionnaire used in Experiment 4a in Chapter 4.

Results

As with the previous experiments, participants were divided into explicit and implicit groups based on verbalized knowledge about the presence of pairs in the visual input. 31 participants were categorized as implicit, while 16 were categorized as explicit. The rate of explicitness of about 34% was much higher than in the previous experiments (mean = 12.3%, SD = 3.8, range = [5.4; 19.5]). This meant participants with implicit and explicit knowledge could be analyzed separately and compared to each other. The reason for the higher rate of explicitness is not clear.

2AFC Results. 2AFC trials were collapsed over the horizontal and vertical pairs, referred to as parallel pairs here, but analyzed separately for diagonal pairs. This was based on the assumption that the novel training phase (presenting a shape outside the grid every 2-4 steps) would affect horizontal and vertical pairs in the same way while it has the potential to influence diagonal pairs differentially. Horizontal and vertical pairs are always presented outside the grid when the movement direction is parallel to their pair orientation. For diagonal pairs, by definition, the movement direction is always diagonal to them. This might be relevant as in Experiment 6a, we saw that the perceived direction of motion could bias learning towards pairs parallel to that direction. Therefore, to avoid overlooking a potential interaction of this effect with the effects of the presentation of shapes outside the grid in the novel setup, I opted to analyze diagonal pairs separately.

The results (see Figure 5.5) showed that participants with explicit knowledge performed above chance for both parallel (M = 88.5, SE = 0.8, $t(15)$ = 11.66, p < .001, d = 2.92, BF = $1.7 \cdot 10^6$) and diagonal (M = 75.0, SE = 1.4, $t(15)$ = 4.32, p = .002, d = 1.08, BF = 58.4) pairs. In contrast, participants with implicit knowledge performed above chance only for parallel (M = 57.9, SE = 0.5, $t(30)$ = 2.84, p = .016, d = 0.51, BF = 5.4) but not for diagonal (M = 49.7, SE = 0.6, $t(30)$ = -0.09, p = .932, d = 0.02, BF = 0.19) pairs. The performance of participants with explicit knowledge was significantly higher than the performance of participants with implicit knowledge for both parallel ($t(35.08)$ = -7.08, p < .001, d = 2.07, BF = $3.4 \cdot 10^5$) and diagonal pairs ($t(23.86)$ = -3.84, p < .001, d = 1.30, BF = 188.8) participants, as shown by two-sample t-tests. The reported significant tests are correct for multiple comparisons using the Holm-Bonferroni method (Holm, 1979).

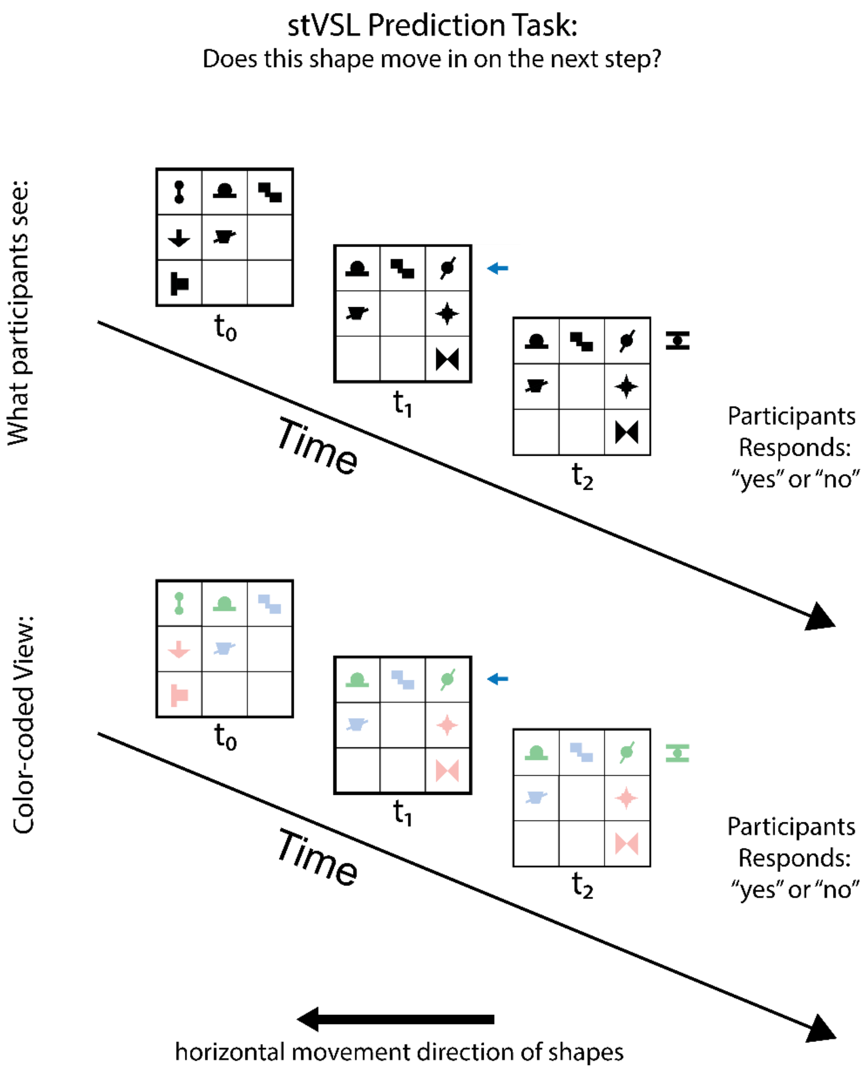

Figure 5.4 Experiment 7 Setup. Experiment 7 builds on the stVSL setup where the currently visible display is not independent of the previous one. Additionally, every 2-4 steps, participants need to make a prediction about whether or not a specific shape will move into the visible aperture on the next step. First, a cue in the form of a blue arrow appears outside the grid (seen in t_1). Next, a shape appears at the same position (seen in t_2), and participants need to indicate whether they believe that this shape will move in from there at the next step. The **top panel** directly shows what participants see in the experiment. The **bottom panel** is only for illustration purposes, as it highlights the underlying structure by color coding pairs.

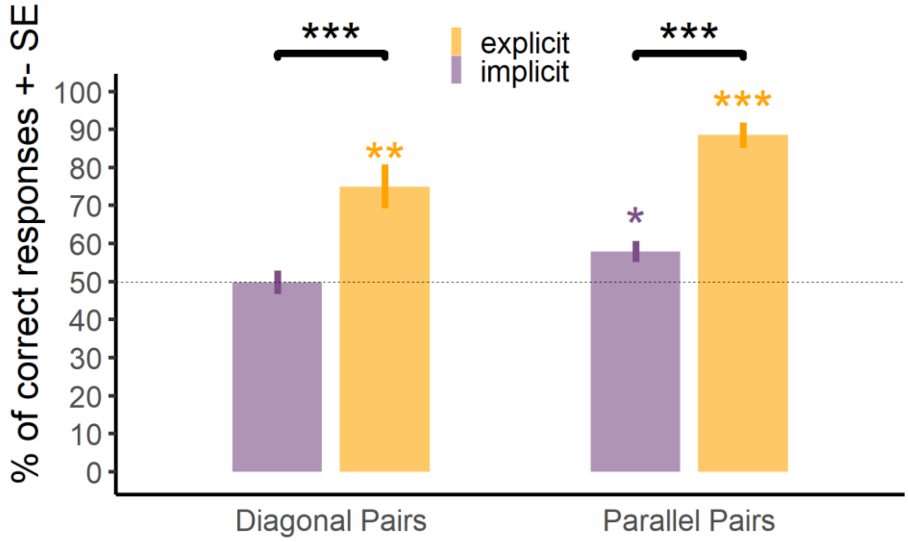

Figure 5.5 Experiment 7 Results – 2AFC. The y-axis represents the participants' mean performance on different 2AFC trials. Error bars represent the standard error. The dotted line indicates the chance level of 50%. Stars represent the significance of the difference from chance. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.005$. Colors indicate groups of participants with explicit or implicit knowledge.

Online Prediction Task Results. As the main interest of this experiment was the trajectory of learning throughout the whole training phase contrasted for participants with explicit and implicit knowledge, I decided to initially analyze the data by group-level *signal detection theory* (SDT). For this purpose, for every test trial, a group level d' was calculated separately for participants classified as explicit and implicit. The calculation of d' was done using the *psycho* R package (Makowski, 2018) *dprime* function. These d' values were then analyzed in a linear regression with the timepoint of the trial as the predictor. The results (see Figure 5.6) showed that for explicit participants, time was a significant predictor of task performance as measured by group level d' ($r^2 = .22$, $F(1, 70) = 20.28$, $p < .001$). In contrast, for implicit participants, time was not a significant predictor of task performance as measured by group level d' ($r^2 < .001$, $F(1, 70) = 0.002$, $p = .966$).

This significant linear trajectory for the group of explicit participants clearly demonstrates that explicit participants learn and that the group steadily increases its prediction performance. However, this cannot be taken as evidence that individual participants also follow a linear trajectory in their prediction performance. As has been shown for other domains (Gallistel et al., 2004), a steady increase over time on the group level can result from step-like functions on the individual level with the step points distributed over time.

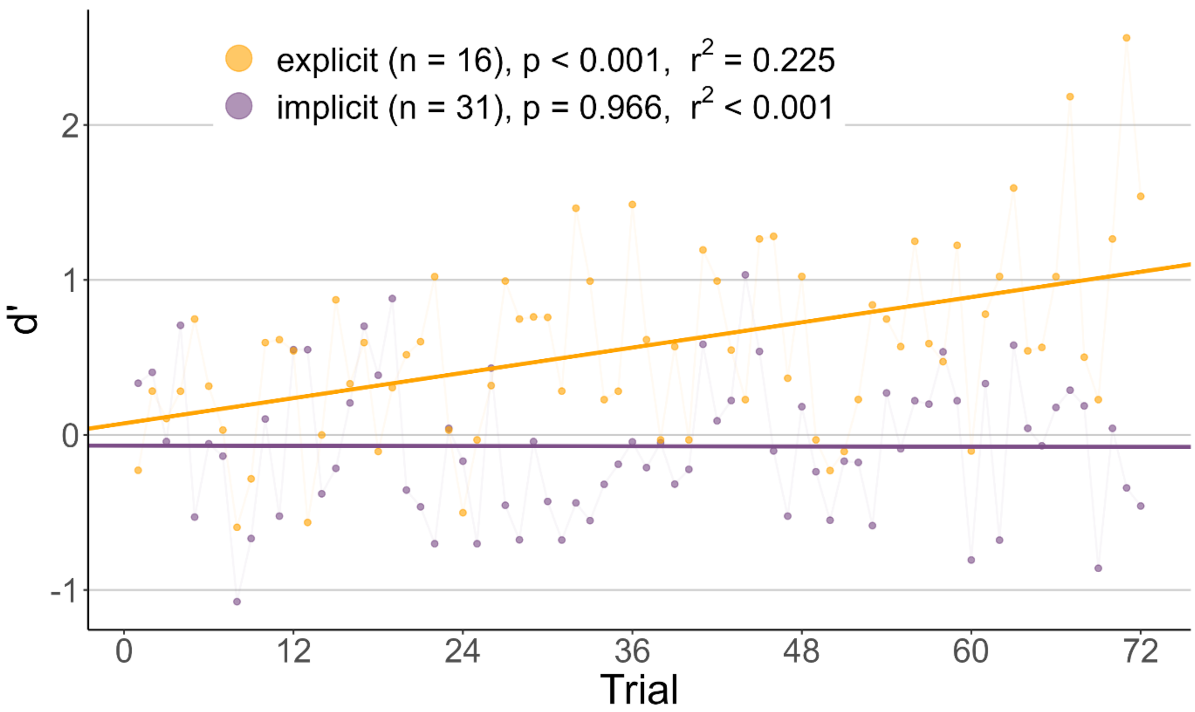

Figure 5.6 Experiment 7 Results – Group Level SDT. The colored points visualize group level d' values for each point in time (= trial number) separately for participants with explicit and implicit knowledge. The colored lines visualize linear regression of these d' values by trial number, showing a significant increase over time only for participants with explicit knowledge.

As an additional analysis, I ran a logistic regression for each individual participant, regressing the correctness of choice by time point. As shown in Figure 5.7, for explicit participants, this shows mostly positive slopes and a higher chance of correct responses at the end of the training.

This is not the case for participants with implicit knowledge. t-tests confirm that the slopes were, on average, different from zero for explicit (M = 0.013, SE = 0.003, t(14) = 4.05, p = .001, d = 1.05, BF = 33.3) but not implicit participants (M = 0.000, SE = 0.002, t(28) = -0.043, p = .966, d = 0.008, BF = 0.20). Furthermore, I correlated the participants' probability of correct choice on the last trial, as predicted by the logistic regressions, with the participants' performance on the 2AFC task. For participants with explicit knowledge, this showed a large significant correlation between choice probability and 2AFC parallel trials (r = .65, p = .008) and a medium non-significant correlation between choice probability and 2AFC diagonal trials (r = .28, p = .315). For participants with implicit knowledge, this showed small, non-significant correlations for both 2AFC parallel trials (r = .12, p = .521) and 2AFC diagonal trials (r = .06, p = .774).

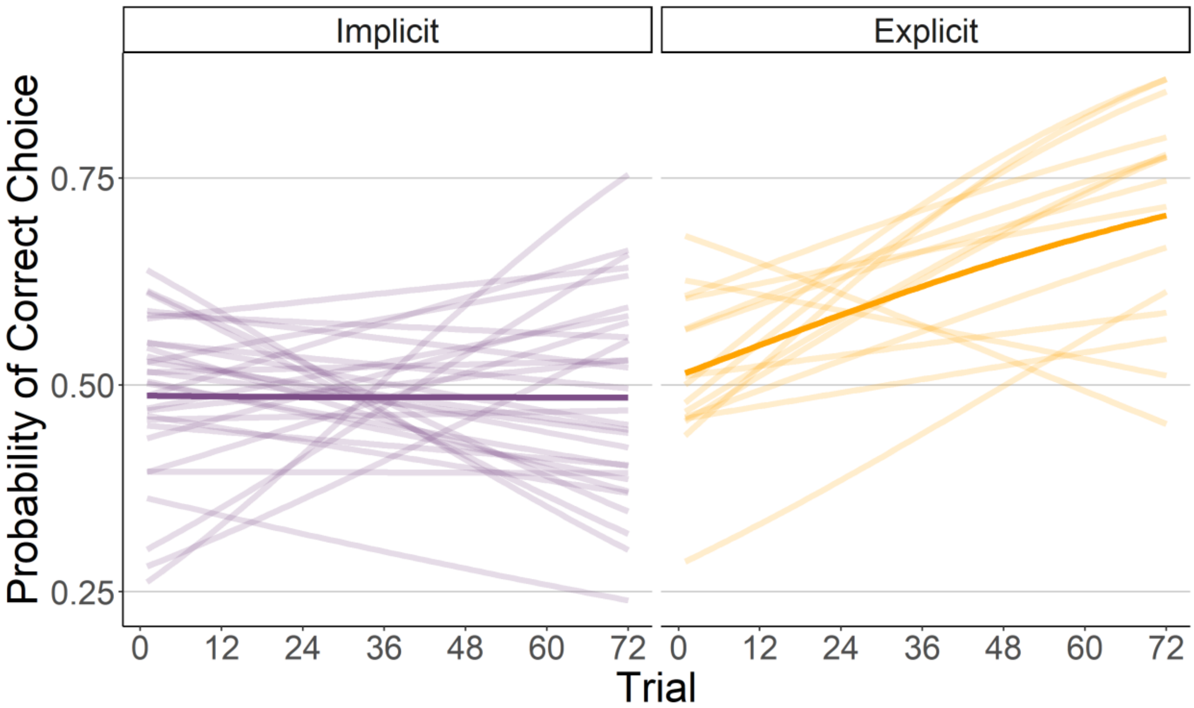

Figure 5.7 Experiment 7 Results – Participant Level Analysis. The faintly colored lines visualize each participant's correct responses logistically regressed by time. The stronger colored lines represent group averages of those regressions. In line with Figure 5.6, this again suggests an increase in prediction performance over time only for participants with explicit knowledge.

Discussion

The results of this experiment demonstrated that while both participants with explicit and implicit knowledge could learn in this setup according to a standard 2AFC familiarity task, only participants with explicit knowledge could apply their knowledge in an online prediction task. This is in line with similar results for statistical learning of temporally defined patterns discussed above. Additionally, for participants with explicit knowledge, their prediction performance was correlated with their 2AFC performance.

5.3 General Discussion

The experiments presented in the current chapter demonstrated connections between the research on abstraction and generalization in VSL presented in Chapters 2 and 3 and the research on spatio-temporal VSL presented in Chapter 4. We see that once we leave the most narrowly controlled setups typical for VSL research, top-down, bottom-up interactions naturally emerge. Furthermore, we again see that the explicitness of knowledge acquired during visual statistical learning is a critical moderator of how this knowledge can be applied to visual input. Overall, this chapter provides further support for the central idea of this thesis, that what is usually called visual statistical learning needs to be understood as part of the overall human unsupervised learning system, operating by combining lower-level spatio-temporal co-occurrence statistics with higher-level top-down biases, combining the available features flexibly across levels.

5.3.1 Higher-level Spatio-temporal Features in VSL

The results showed that while in Experiment 6a, participants had an overall bias to choose options aligned with the motion direction, in Experiment 6b, participants preferred options perpendicular to it. This suggests that in Experiment 6a, the observed motion induced the overall bias. The opposite bias observed in Experiment 6b might be due to the shapes being seen next to each other perpendicular to but not along the movement direction, i.e., due to the overall perceived shape arrangement. This demonstrates a flexible and parsimonious acquisition of biases based on the salience of the inducing features. In Experiment 6a, the feature that is apparently inducing the bias in 6b, overall shape arrangement, is present; however, it seems to be overshadowed by the feature of perceived motion.

As discussed in the introduction of this chapter (5.1), the influence of *higher-level* features of the visual input, transcending the item-item co-occurrence statistics usually studied in VSL experiments, has so far been largely neglected in VSL research. However, there are a few exceptions. While an overall shape arrangement has been utilized in a VSL study before (Jun & Chong, 2016), there it was part of a predictable sequence, therefore being subject to statistical learning in the classical sense itself. In the current study, this overall feature of the input is used in a way not suitable for statistical learning to investigate how it will then interact with statistical learning proper. Another relevant previous study focused on the interaction of statistical learning and *statistical summary perception* (Zhao et al., 2011), a type of visual processing extracting summary statistics over an input set. It was found that having participants focus on summary statistics will impede statistical learning and vice versa. The overall motion direction utilized in the current study could be considered a summary statistic. However, critical differences in how this is utilized in the studies remain. For Zhao et al. (2011), the summary statistics are computed over the same input feature that encodes the item identity (the orientation of a line), this is not the case for the motion direction in the current study. Furthermore, the current study can, if at all, be only considered a fringe case of statistical summary perception, as all shapes follow the same motion direction at a given point in time; i.e., there is no variance at all. These differences highlight how this previous study and the current one are complementary in their design and scope, focusing on different aspects of interactions of low-level co-occurrence and higher-level group statistics. Interestingly, their differences generate a conceptual space of possible studies which could be conducted by varying the parameters of (1) variance in summary statistics, (2) overlap of features used for summary statistics and item identity, and (3) guiding participants attention to different aspects (as Zhao et al. did).

A phenomenon in the object recognition literature related to the current study is *aperture viewing* (Morgan et al., 1981). In aperture viewing, an object or pattern is moving behind a narrow slit so that only small parts of it are visible at any given point in time. It has been shown that people are generally able to recognize known objects and patterns in a range of slit sizes and movement speeds. Whereas early work suggested that this effect is solely based on low-level integration of enduring activation (*retinal painting*), it has later been shown that, at least in some circumstances, higher-level perceptual mechanisms need to be recruited (Morgan et al., 1981). A newer version of research in this area investigates object perception using *minimal videos* (Ben-Yosef et al., 2020). There, it was shown that objects could be recognized in a video recording that is strongly degraded in both the spatial (down-sampling and/or cropping) and the temporal (removing video frames) domains. At the point of degradation termed minimal video, any further reduction of spatial or temporal information leads to a substantial drop in object identification accuracy and image interpretability. This highlights how spatial and temporal information interact crucially in object perception. While this type of research focuses on perception, the current study investigates learning. Future studies could combine both aspects in investigating how the interaction of spatial and temporal information in learning relates to the interaction of spatial and temporal information in later recognition.

It should be noted that the higher-level features used in the current study, overall motion direction and overall shape arrangement, might be only two examples of relevant features that could influence visual statistical learning. Other features, such as presentation rates, speed of motion, complexity and amount of stimuli, and others, could influence VSL on their own as well as in interaction with each other and the lower-level statistics.

5.3.2 Explicitness and Prediction in spatio-temporal VSL

In Experiment 7, I showed that while both participants with implicit and explicit knowledge can learn in this novel setup according to a classic 2AFC measure, only participants with explicit knowledge can use their knowledge in an online prediction task. These results for spatially defined patterns, presented in the spatio-temporal VSL setup, are in line with previous results from purely temporal VSL showing an association between explicitness of knowledge, prediction performance, and predictive brain activity (Dale et al., 2012; Sznabel et al., 2023). Furthermore, the finding that participants with implicit knowledge perform above chance for the 2AFC but not the prediction task mirrors previous results in temporal VSL, showing differential behavioral outcomes for different measures of VSL within participants (Bays et al., 2016). The results of Experiment 7 and the related results in VSL research discussed here can be seen as instances of a more general phenomenon of dissociation of different measures of learning in the memory literature (Ingram et al., 2012; Jurjut et al., 2017; Kuchibhotla et al., 2019; Yonelinas, 2002) often neglected when summarizing results over several VSL studies.

It should be noted that in my experiment, the participants with explicit knowledge showed a significantly higher level of learning for the 2AFC trials than the participants with implicit knowledge. This difference could, in principle, explain the difference in prediction behavior. Unfortunately, the sample sizes of the current study make a matched sample analysis, as in Chapters 2 and 3, unpractical for the current study.

The current results are also in line with the results of Experiments 1a-c and 2a-b in Chapter 2, demonstrating a difference in how participants with explicit and implicit knowledge can apply their acquired knowledge. While in Chapter 2, we saw that participants with explicit, but not implicit, knowledge can use it for generalization, here we see that participants with explicit, but not implicit, knowledge can use their knowledge for prediction. Building on the finding of Chapter 3 that participants with implicit knowledge are able to generalize after a phase of asleep consolidation hints at the possibility that the same might be the case for prediction. That is, the transformation of implicit knowledge during sleep might change it into a representational format suitable for prediction. However, this is speculative as, in contrast to the case of generalization, there is no established mechanism of representational change during sleep that could account for prediction.

CHAPTER 6

General Discussion

This dissertation set out to achieve two goals. The first goal was a reconceptualization showing that what is canonically called visual statistical learning is part of a more extensive, hierarchically structured system of human unsupervised learning in which both bottom-up and top-down influences play an important role. The second goal was the development of a joint spatio-temporal visual statistical learning paradigm that enables systematic investigation of how the temporal and spatial statistics of the input interact in unsupervised learning. Overall, the findings presented in this dissertation support a view of visual statistical learning as an interaction of lower-level spatio-temporal co-occurrence statistics and higher-level top-down biases. Therein, available features across levels are flexibly combined to achieve a congruent and comprehensive yet parsimonious interpretation of the world.

To summarize the main results of this dissertation, I identified three types of higher-level biases. First, pre-existing biases that cannot be explained by properties of the experiment, such as a bias of diagonally over cardinally oriented structures (Chapter 4, Experiments 5a-c) and a bias of horizontally over vertically oriented structures (Chapters 2, Experiments 2b). Second, biases based on the abstraction of learned low-level co-occurrence statistics, such as a structural transfer or structural novelty effect (Chapters 2 and 3, Experiments 1a-c, 2a-b, 3a-b). Third, biases based on observed higher-level spatial or temporal input features, such as a perceived overall motion direction or shape arrangement (Chapter 5, Experiments 6a-b). Additionally, two critical moderators of this hierarchical learning system were identified: explicitness and consolidation of knowledge. The explicitness of knowledge influenced how it could be applied to novel input for generalization (Chapters 2, Experiments 1a-c, 2a-b) or prediction

(Chapters 5, Experiment 7). Consolidation of knowledge enabled generalization of learned structures even without explicitness (Chapter 3, Experiment 3a, 3c). Finally, the research presented in Chapters 4 and 5 demonstrated that participants used the temporal statistics of the input in learning spatially defined patterns and identified essential moderators of this process, such as perceived motion.

6.1 Connections to Previous Research

The studies presented in this dissertation are related to a broad range of research from within and outside the statistical learning literature. Here, I will summarise important connections already discussed in the previous chapters and introduce novel connections that concern aspects of the research spanning multiple chapters.

6.1.1 Consolidation and Statistical Learning

Previous research found mixed results for the effect of consolidation on statistical learning (see Chapter 3.1.2), with some studies showing that asleep or awake consolidation helps increase or retain test performance. The effect of consolidation, especially sleep-based, on abstraction has been demonstrated outside the field of statistical learning (see Chapter 3.2.1). The experiments conducted for this dissertation showed that sleep-based consolidation also plays a critical role in abstraction from representations built during visual statistical learning and, therefore, in the domain of unsupervised learning. Following the complementary learning systems (CLS) framework (McClelland et al., 1995) and recent additions to it (Schapiro, Turk-Browne, et al., 2017), this might be realized by an interaction between the hippocampus and neocortex or between different learning systems within the hippocampus. However, although suggesting a similar mechanism for my findings and previous findings on CLS seems natural, this does not automatically mean that the same neural substrate implements both. Only studies targeted at the neural substrate of the represented chunks and more abstract, structural representations could answer such questions.

The explanation given for my findings in Chapter 3 suggests that both the structural novelty effect and the sleep-based generalization found for participants with implicit knowledge can be explained by the representational overlap of the chunk representations. The idea is that while in the absence of consolidation, the representational overlap leads to proactive interference hindering the learning of similar pairs, after consolidation, the shared structure has been abstracted and now guides future learning. However, confirming this interpretation would require further studies. Empirical studies employing neural measures that allow tracking the representations and representational similarities of shapes and pairs during learning and consolidation could test this interpretation. Furthermore, existing models of CLS mechanisms (O'Reilly et al., 2014; Schapiro, Turk-Browne, et al., 2017) could potentially be extended to capture structure-based novelty effects.

Interestingly, while the proactive interference (Kliegl & Bäuml, 2020) reported for participants with implicit knowledge was released by asleep consolidation in my experiments, this was not significant for the retroactive interference (Dewar et al., 2007) reported for participants with explicit knowledge. This is in contrast to previous results showing a release from retroactive interference by asleep consolidation (Abel et al., 2023; Ellenbogen et al., 2006) for declarative memory. However, due to the smaller number of participants with explicit knowledge in my studies, this comparison might have been underpowered, and more targeted studies are needed to clarify these findings.

How can the disparate findings on the influence of consolidation on statistical learning be reconciled? Using the alternating serial reaction time (ASRT) task, it was shown that learning specific simple chunks happens online while learning more complex higher-order rules depends crucially on offline phases (Quentin et al., 2021). Higher-order, in this context, means second- or third-order temporal transitions between specific elements and, therefore, non-adjacent associations as compared to simple adjacent transitions. These higher-order rules are, thus, directly observable in the input and not latent factors such as the underlying orientation used for the experiments presented in Chapters 2 and 3 of this dissertation. Furthermore, asleep consolidation has been demonstrated to have an effect on statistical learning using probabilistic input as compared to deterministic input (within chunk conditional probabilities of 1) (Durrant et al., 2013, 2011) on the cross-modal transfer of statistical learning (Durrant et al., 2016), and on statistical learning in the presence of interference (McDevitt et al., 2022). These findings, combined with my results presented in Chapters 2 and 3, might suggest that the critical boundary for consolidation dependence in unsupervised learning is not just between observable (statistical learning) and more latent (structure learning) features. Instead, consolidation might be critical for complex (non-adjacent, cross-modal, abstract, probabilistic, or interference-subjected) but not simple (adjacent, unimodal, specific, deterministic, and interference-free) regularities.

6.1.2 Explicitness and Statistical Learning

Previous research in VSL debated whether explicit task instructions - mentioning the presence of pairs, triplets, or any fixed item combinations - would lead to different learning outcomes. While some authors suggested that explicit instructions do not matter (Arciuli et al., 2014), others have demonstrated that this strongly depends on context factors (Bertels et al., 2015). In short, it was shown that explicit instructions do not matter when using extremely short presentation times but do matter for timings more typically used in VSL research. Another study that failed to find any VSL effect for their setup without explicit instructions found a strong effect in a sample that received instructions (Himberger et al., 2019). All of these contributions focused on how explicit instructions change performance quantitatively, i.e., more or less learning. In contrast, the results of Experiment 1c in Chapter 2 of this dissertation showed a critical effect of explicit instructions on both quantitative and qualitative behavior. Explicit instructions led not only to higher levels of learning but also to generalization. Furthermore, comparing these results to those of the participants reaching explicitness on their own in Experiment 1a demonstrates that similar behavior can occur independently of whether explicit knowledge was achieved by instructions or spontaneously. This is similar to recent reports in the field of reinforcement learning, where explicit instructions change behavior qualitatively in a way that is observed spontaneously in only a few participants (Castro-Rodrigues et al., 2022).

Furthermore, the results reported in this dissertation highlight how critical it is to consider the explicitness of knowledge as a covariate in SL research. While many studies in SL simply assume implicitness of knowledge, the current results suggest that the sample will often include a subsample of participants who develop explicitness of knowledge. This is highly problematic as I have found strong quantitative and qualitative differences between participants with explicit and implicit knowledge. Not identifying the explicit subsample can lead to the computation of statistical artifacts, e.g., group means that do not meaningfully represent the sample, which might mask existing effects. While this is already problematic for quantitative differences, it can be even worse for qualitative differences. In the case of Experiment 1a in Chapter 2 of this dissertation, not considering the explicitness of knowledge as a covariate would have masked both the generalization behavior of participants with explicit knowledge and the structural novelty effect of participants with implicit knowledge. In short, my conclusion is that explicitness matters and should be tracked in SL research.

6.1.3 Hierarchies in Statistical Learning

Previous studies argued for multiple ways in which statistical learning could be a part of a hierarchical learning system. It was recently suggested that statistical learning competes with rule learning in explaining perceptual input (Maheu et al., 2022). The hierarchical aspect of this interpretation is that both statistical learning and rule rearning are situated at a lower level, working in parallel, while the higher level of the hierarchy realizes an arbitration mechanism, which decides whether statistical or rule learning is fitting for the current input. This is in contrast to my interpretation, according to which forms of learning based on abstraction, such as rule learning, are situated at a higher level of the hierarchy than statistical learning. It should be noted that my view of statistical and rule learning as different levels of a hierarchy does not suggest a simple feed-forward hierarchy in which the output of a statistical learning process feeds into a rule learning process. Instead, as discussed in the Introduction chapter and empirically demonstrated in Chapters 2 and 3, there is an ongoing interaction between levels of abstraction in which each level constraints connected levels while simultaneously being constrained by them.

It was also recently suggested that statistical learning leads to a compositional hierarchy of representation (Lee et al., 2021). According to this view, learned chunks, which consist of multiple elements such as shapes, can themselves become elements for creating larger chunks. For example, a quadruplet of shapes can be represented as consisting of two pairs of shapes. In this view, which parts of a chunk should be represented as smaller chunks flexibly depends on the experience with the input. For example, if I have learned the shape pair AB before, I might represent the quadruplet ABCD as consisting of the pairs AB and CD. This idea of a hierarchy of composition is neither the same as my proposed hierarchy of abstraction nor is it at odds with it. Both are relevant for the representation of real-world visual input. An overarching understanding of unsupervised visual learning will need explain how this system can realize both a hierarchy of composition and a hierarchy of abstraction simultaneously.

Finally, for temporal visual and auditory statistical learning, it was demonstrated via intracranial recordings that the structure of the input stream was represented at different levels along the neural processing hierarchy (Henin et al., 2021). While transitional probabilities were most strongly represented in modality-specific early sensory regions, such as areas in the occipital cortex and superior temporal gyrus (STG), information about item position and identity was most strongly represented in not modality-specific associative regions, such as inferior frontal gyrus (IFG), anterior temporal lobe (ATL) and the hippocampus. In principle, the approach chosen by Henin et al. (2021) could be extended to localizing the neural substrate of more abstract features learned over several chunks, such as the ones used in my studies. However, Henin et al.'s neural frequency tagging methodology is naturally suited for measuring the learning of temporal but not spatial regularities. This is the case as it measures the power of the frequency spectrum at the chunk frequency, which is not defined for classical spatial VSL. Therefore, alternative measures would be necessary to meaningfully combine these two lines of research.

6.1.4 Eye Movements and Statistical Learning

Recent research found connections between statistical learning and eye movements (Arato et al., 2023; Zolnai et al., 2022). For the temporally defined alternating serial reaction time (ASRT) task, it was shown that eye movements were a signature of anticipating upcoming predictable stimuli (Zolnai et al., 2022). For spatially defined visual statistical learning, it was shown that learning the pairs of shapes embedded in visual scenes leads to more within-pair transitions of fixations (Arato et al., 2023). This eye-movement-based signature of visual statistical learning was also correlated with the performance in a classical 2-alternative forced choice (2AFC) familiarity task. These findings naturally lead to the question of how eye movements might be related to the VSL phenomena studied in this dissertation. This subchapter will discuss potential connections, highlighting possible future avenues of research.

First, eye movements are a potential explanation for the preference for horizontal pairs found in Chapter 2, Experiment 2b. There, I found that in the absence of any induced bias to an orientation, participants prefer horizontal structures to vertical ones. As discussed in that chapter, preferences for horizontality have been demonstrated for several domains, including visual processing (Lim & Sinnett, 2012), face perception (Balas et al., 2015; Dakin & Watt, 2009), and, importantly, direction of spontaneous saccades as captured with eye-tracking (Foulsham et al., 2008; Gilchrist & Harvey, 2006; Tatler & Vincent, 2008; Van Renswoude et al., 2016). In principle, this bias for saccade direction alone could be enough to account for preferential learning of horizontal pairs as more spontaneous horizontal saccades would naturally bias attention to horizontal associations, i.e., a participant will more often see the shapes of a horizontal pair in succession as compared to the shapes of a vertical pair. Such an effect could, for example, be investigated with novel studies correlating spontaneous horizontal eye movements with learning of horizontal and vertical pairs or with studies experimentally manipulating eye movement patterns.

Second, eye movements could also play a role in transferring learned structure described in Chapters 2 and 3. If within-pair fixation transitions are becoming increasingly more likely throughout learning, as demonstrated by Arato et al. (2023), and only pairs of one orientation (horizontal or vertical) are used, such as in the first training phase of the experiments of Chapters 2 and 3, we would expect a high number of fixation transitions along this one orientation at the end of the first training phase. If this biased transition pattern is preserved for the context of an experiment and therefore also present in the second learning phase, this would constitute a mechanistic explanation of the found generalization pattern. However, how would this relate to the implicit participants' performance pattern and the observed effect of consolidation? The absence of generalization for participants with implicit knowledge in the immediate transfer would not be surprising based on the findings of Arato et al., as they found that for implicit participants, the eye-movement effect takes longer to develop than for explicit participants and only shows after more than 100 trials, which is longer than the first training phase in the Experiments in Chapters 2 and 3. Furthermore, a potential effect of eye movements does not need to be at odds with the memory-transforming effects of consolidation suggested in the previous chapters, as they could work alongside or in interaction with eye-movement-based effects.

Third, in Chapters 4 and 5, we saw strong effects of perceived motion and motion direction in visual statistical learning, which could also be connected to eye movements. The motion might lead to more transitions along the perceived motion direction, leading both to a general effect of motion (Chapter 4, Experiment 5a-c) and biases for specific directional associations based on biased motion direction (Chapter 5, Experiment 6a). The effect of a bias towards the overall shape arrangement in the absence of motion (Chapter 5, Experiment 6b) could also be related to a transfer of transition patterns from the initial training.

We can see that there are a multitude of potential connections between eye movements and the findings of this dissertation. However, firm statements on such connections will require more research that employs eye-tracking to VSL transfer learning and spatio-temporal VSL paradigms, potentially alongside experimental manipulations that bias the direction of fixation transitions.

6.1.5 Object-Like Representations in Statistical Learning

It was previously shown that representations of chunks arising during visual statistical learning have properties of object representations, leading to the idea that VSL leads to essentially object-like representations (Fiser & Lengyel, 2022). First, it was demonstrated that chunks learned during VSL show *object-based attention* (Lengyel et al., 2021), suggesting that the shapes making up the chunks are represented as one unified object after learning. Second, purely statistically defined chunks learned in the visual domain showed zero-shot generalization to the haptic domain (Lengyel et al., 2019). In short, after learning the co-occurrence statistics of shapes based on VSL, participants interacted with them manually as if they were solid objects within but not across chunk boundaries. Third, it was demonstrated that spatial VSL reduces the perceived numerosity of multi-chunk scenes, again suggesting that chunks are perceived as single unified objects (Zhao & Yu, 2016). Taken together, we see that chunks formed during VSL show several important properties associated with object perception and cognition.

The results presented in Chapters 2 and 3 add to this evidence by demonstrating for VSL chunks a further hallmark of object cognition: nested organization of representations into categories or concepts. Objects are not simply represented as an unordered set of elements, but their representation is structured into higher-level categories based on their common underlying structure or shared features (Ashby & Maddox, 2005; Richler & Palmeri, 2014). Similarly, as demonstrated by my experiments, the chunks learned during VSL can be nested into representations of their shared structures, such as the category of horizontal pairs, which in turn is used to interpret novel input. This adds to the existing evidence suggesting that statistical learning is a vehicle for acquiring object-like representations.

6.2 Limitations and Future Directions

The findings presented in this dissertation demonstrate that visual statistical learning should be understood as part of a larger unsupervised learning system. I suggested that this builds on an interaction of lower-level spatio-temporal co-occurrence statistics and various higher-level features. However, for every answer provided by this research, new questions emerged.

What exactly is represented at the "higher level"? The effects found in Chapters 2 and 3 could be explained by a general bias for one orientation, i.e., *horizontality,* or by a more specific representation of a structure, i.e., *horizontal pair*. Such differences could be investigated in experiments studying the transfer from oriented pairs to triplets or vice versa. Furthermore, the chunk orientation is only one feature. How do the current findings generalize to other features? A challenge here is to define chunks with features that can be abstracted, that are not obviously perceived from the input without learning the chunks, and that have two versions that are in some sense orthogonal to replicate the switch from novelty effect to generalization.

Another open question is: what is the time course of learning and generalization within the different learning phases? The current studies on transfer learning had a strict separation between training and test phases, so the collected data cannot answer these questions. A challenge for future studies is to devise a paradigm where implicit learning in spatial VSL can be captured online. The current thesis successfully developed such a measure for explicit learning, as the prediction results for participants with explicit knowledge in Experiment 7 were highly correlated with their offline 2AFC performance. However, this was not the case for participants with implicit knowledge. Previous studies developing online measures for VSL focused on the temporal domain (Batterink & Paller, 2017; Henin et al., 2021; Quentin et al., 2021; Sherman et al., 2023; Zolnai et al., 2022) and partially used explicit instructions (Siegelman et al., 2018). However, these measures developed for capturing the learning of temporal statistics cannot be directly applied to learning spatial statistics. Developing such a measure for spatial VSL is, therefore, a critical challenge for advancing VSL research.

As mentioned above (6.1.1), although the findings on the role of consolidation in abstraction presented in this dissertation mimic results of the complementary learning systems framework, this does not show that the same neural mechanisms and substrate are at play. Future studies tracking neural representations are necessary for testing this interpretation.

Similarly, the suggested links (6.1.4) between my findings and eye movements so far are speculative, although in line with previous results. These interpretations can only be tested by the direct application of eye-tracking to the paradigms introduced in this dissertation.

The spatio-temporal VSL paradigm has the potential for extension into several directions. Giving participants control over the direction of movement would extend SL research toward an active learning framework, as previously demonstrated with a gaze-contingent VSL paradigm (Arato et al., 2023). Such a setup lends itself to investigating how what participants already learned interacts with their exploration behavior. Furthermore, biasing the visual environment to contain more regularities or specific types of regularities in some regions opens an avenue to investigate interactions of local spatial regularities (i.e., what co-occurs with what) and global spatial regularities (i.e., what occurs where). This has the potential to provide a natural link to both the abstraction/generalization experiments presented in Chapters 2 and 3 of this dissertation and to research on the learning of global spatial (and conceptual) spaces such as cognitive maps (Behrens et al., 2018; Tolman, 1948). Naturally, when abstraction and interactions between abstract and specific information come into play, the spatio-temporal VSL paradigm could and should be extended to multi-session consolidation versions, given my previous findings on the role of consolidation in abstraction.

6.3 Conclusions

The research presented in this dissertation established meaningful connections between what is usually called statistical learning and the learning of more abstract features, therefore establishing a connection between previously disparate lines of research. The results showed important interactions between lower-level co-occurrence statistics and higher-level biases, providing evidence that the ecological role of statistical learning cannot be understood by only studying it in isolation. Furthermore, by developing a spatio-temporal VSL paradigm, the research presented in this dissertation joined two previously largely disparate lines of research within statistical learning. The results showed not only that temporal statistics are used in the implicit learning of spatial patterns but also that interactions arise that would not be predicted by studying the learning of spatial and temporal SL in isolation. Additionally, the dissertation research emphasized qualitative differences based on the explicitness of knowledge in unsupervised learning. It, thus, supports the notion that tracking explicitness in SL research is essential. Taking all these findings together, this dissertation demonstrates that the narrow limitation and control that enabled the initial success of SL research need to be carefully and incrementally overcome to understand the role of SL in the overall human cognitive system. It does so by introducing two new VSL paradigms that enable novel, systematic ways of investigating the human unsupervised learning system.

Appendix

A. Post-Experiment Questionnaire

After completing the test phase, participants in all experiments gave written responses to several open questions. Assessment of explicit knowledge was based on the responses to:

1. Please explain with your own words what you think the experiment was about.
2. In the first part of the experiment, where you only passively watched the screen: Did you notice any regularities in how the shapes were arranged? If yes, please describe them.

Any references to fixed pairs or reappearing combinations of shapes were counted as evidence for explicit knowledge.

B. Consolidation Studies: Details on Sleep and Time-of-Day

In order to ensure that participants had overnight sleep during Experiments 3a and 3c and that they did not sleep during the day in Experiment 3b, several constraints and checks were implemented. First, participants were not taken from the full prolific pool but restricted to several European countries within the same time zones. Country of residence is one of the attributes of participants that prolific.co verifies. In order to roughly approximate the geographic distribution of participants in Experiment 1a (see Figure B.1), I chose countries from two time zones. GMT±00:00 (Countries: UK and Portugal) and GMT+01:00 (Countries: Germany, France, Spain, Czech Republic, Denmark, Hungary, Italy, Netherlands, Poland, Slovenia, Switzerland). Second, as I did not expect Prolific's residence information to be perfectly predictive of where participants were while they conducted the experiment, participants were asked what the current time at their location was. Third, at the start of the second part of the experiment, participants were asked how much they slept between the first and second parts. This was used to exclude participants from Experiment 3b who slept during the day.

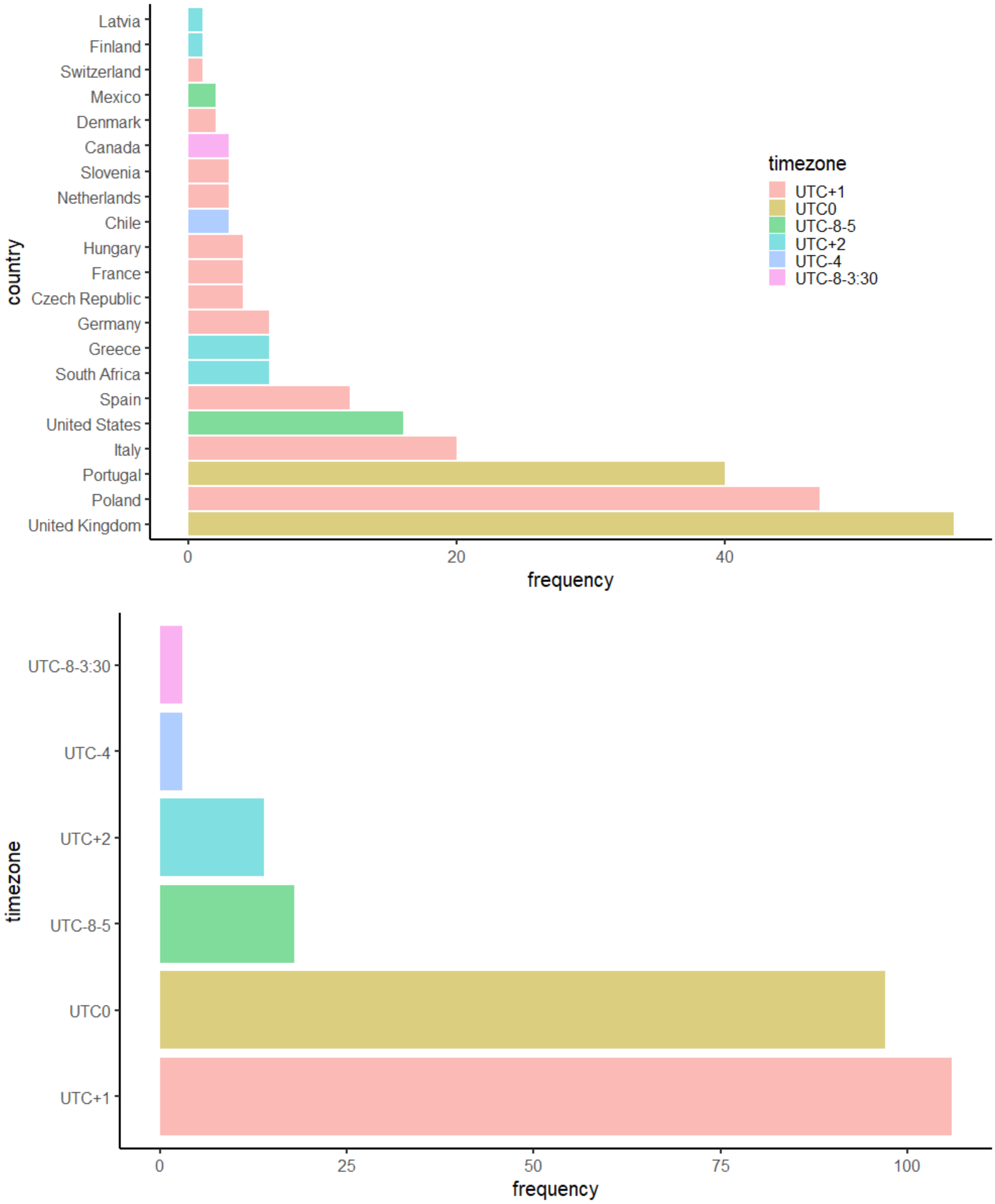

Figure B.1 Experiment 1a - Country and Time Zone. The **top panel** shows the country of residency for the participants of Experiment 1a, used as a proxy for the time zone they were in, visualized in the **bottom panel**. Participants for the subsequent consolidation studies were chosen from the countries of the most frequent time zones: GMT±00:00 (Countries: UK and Portugal) and GMT+01:00 (Countries: Germany, France, Spain, Czech Republic, Denmark, Hungary, Italy, Netherlands, Poland, Slovenia, Switzerland).

C. Matched Sample Analysis

As reported in the main text of Chapters 2 and 3, explicit participants show higher average learning in the first learning phase, which could be what enables the generalization of the learned structure. To test this idea, I conducted a matched sample analysis (Ho et al., 2007). The general idea of this analysis is to create a sub-sample of the implicit participants that perform like the explicit participants for the trials of the first training phase.

For all experiments, in a first step, I ran six applicable matching algorithms implemented in the *MatchIt* R package (Ho et al., 2011). The six matched implicit samples were then compared to the original explicit sample according to four metrics: standardized mean difference, variance ratio, mean of the empirical cumulative density function, and maximum of the empirical cumulative density function. All values for all experiments can be seen in Supplementary Tables 1-4. “Unbalanced” denotes the values for the full, non-matched implicit sample. All values for the used matching methods are evaluated as an improvement from those values. *Std. Mean Diff.* describes how far the mean of the matched sample is from the comparison sample (explicit participants); values closer to zero are better. The *variance ratio* is the ratio of the variances of the matched and the comparison sample; the best possible value is 1. The *eCDF* (empirical cumulative density function) contains more information than the mean and variance ratio as they capture the whole distribution of values. Two commonly used simple metrics based on the *eCDF* are the *mean* and *maximum difference* of the eCDFs of the matched and comparison group. Generally, values closer to zero are better. The best-fitting matching algorithm was not exactly aligned for all experiments. For consistency reasons, I chose the overall best-fitting method for all experiments: *nearest neighbor matching with replacement*.

Experiment 1a Results

The matched sample showed a pattern similar to that of the original full sample in the second training phase. A 2x2 ANOVA using the *novel* and *same structure* pairs for the original explicit and the matched implicit data showed a significant interaction ($F(1, 87) = 8.53$, $p = 0.004$, $BF = 10.7$, $\eta_p^2 = 0.09$) and post-hoc comparisons showed a significant difference between *novel* and *same structure* trials for the matched implicit data ($p = 0.012$; $BF = 3.6$). This analysis suggests that the difference between the two groups is not based on different learning strengths.

Experiment 3a Results

As in Experiment 1a, the matched sample showed the same pattern as the full sample. As a critical analysis, we can see that for the matched implicit sample, there is a significant difference between learning pairs of the *same* and of the *novel structure* ($d = 0.98$, $t(20) = 4.51$, $p < 0.001$, $BF = 127$), suggesting generalization of the structure.

Experiment 3b Results

The matched sample showed a similar pattern as the full sample. Critically, we can see that for the matched implicit sample, there is no significant difference between learning pairs of the *same* and of the *novel structure* ($d = 0.29$, $t(19) = -1.29$, $p = 0.214$, $BF = 0.59$), suggesting no generalization of the structure.

Experiment 3c Results

As in Experiment 1a, the matched sample descriptively showed the same type of pattern as the full sample. However, the critical analysis of the difference between learning pairs of the *same* and of the *novel structure* for the matched implicit sample failed to reach significance ($M_{\text{diff}} = 8.66$, $d = 0.22$, $t(22) = 1.05$, $p = 0.304$, $BF = 0.46$).

Supplementary Table 1

Experiment 1a - Overview of Balance Metrics for the Used Matching Algorithms

Matching Method	Std. M. Diff.	Var. Ratio	eCDF mean	eCDF max
Unbalanced	0.486	2.92	0.169	0.358
NN with replacement	**-0.002**	**0.95**	**0.003**	**0.059**
NN without replacement	0.078	1.31	0.026	0.206
Optimal pair matching	0.078	1.31	0.026	0.206
Optimal full matching	-0.018	1.05	0.012	0.059
Coarsened exact matching	-0.021	1.12	0.021	0.118
Subclassification	-0.049	1.19	0.020	0.059

Supplementary Table 2

Experiment 3a - Overview of Balance Metrics for the Used Matching Algorithms

Matching Method	Std. M. Diff.	Var. Ratio	eCDF mean	eCDF max
Unbalanced	1.125	2.909	0.273	0.548
NN with replacement	**0.432**	**2.424**	**0.105**	**0.429**
NN without replacement	0.103	1.089	0.030	0.238
Optimal pair matching	0.432	2.424	0.105	0.429
Optimal full matching	0.112	1.079	0.033	0.238
Coarsened exact matching	0.142	1.130	0.037	0.294
Subclassification	0.248	1.175	0.060	0.238

Supplementary Table 3

Experiment 3b - Overview of Balance Metrics for the Used Matching Algorithms

Matching Method	Std. M. Diff.	Var. Ratio	eCDF mean	eCDF max
Unbalanced	1.067	3.284	0.290	0.49
NN with replacement	**0.293**	**2.260**	**0.079**	**0.30**
NN without replacement	0.043	0.950	0.018	0.25
Optimal pair matching	0.293	2.260	0.079	0.30
Optimal full matching	0.068	1.073	0.023	0.25
Coarsened exact matching	0.043	1.161	0.023	0.25
Subclassification	0.089	1.106	0.039	0.25

Supplementary Table 4

Experiment 3c - Overview of Balance Metrics for the Used Matching Algorithms

Matching Method	Std. M. Diff.	Var. Ratio	eCDF mean	eCDF max
Unbalanced	0.819	5.352	0.222	0.494
NN with replacement	**0.450**	**3.006**	**0.106**	**0.391**
NN without replacement	0.164	1.327	0.043	0.304
Optimal pair matching	0.450	3.006	0.106	0.391
Optimal full matching	0.151	1.479	0.047	0.304
Coarsened exact matching	0.026	1.099	0.016	0.064
Subclassification	0.240	2.404	0.071	0.304

D. stVSL Results for Explicit Participants

In the main text, all results for the stVSL experiments included only participants with implicit knowledge. Here, I give a short overview of the results for the explicit group of participants, discussing them descriptively in relation to the results of the implicit group of participants.

Experiment 5a. The participants with explicit knowledge show better overall performance and do not show superiority in learning diagonal pairs. (N explicit: 15; Proportion of explicit participants: 17.05%)

Experiment 5b. The participants with explicit knowledge show better overall performance and do not show superiority in learning diagonal pairs. (N explicit: 15; Proportion of explicit participants: 16.67%)

Experiment 5c. The participants with explicit knowledge show better overall performance and do not show superiority in learning diagonal pairs. (N explicit: 10; Proportion of explicit participants: 11.11%)

Comparison of 5a, 5b, 5c: Descriptive Results: 5a is not superior to 5c for the explicit group, suggesting no or smaller effect of the temporal coherence.

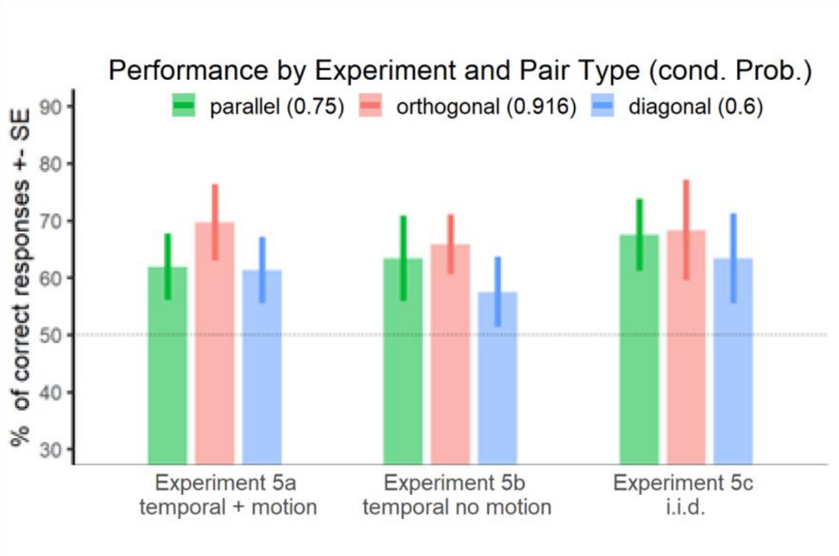

Figure D.1 Experiment 5a-c Explicit Results. This figure visualizes 2AFC test performance for participants with explicit knowledge in Experiments 5a-c. The y-axis represents the performance on 2AFC trials. Error bars represent the standard error. The dotted line indicates the chance level.

Experiment 6a. Explicit participants learn the parallel but not the orthogonal pairs, as the implicit participants. Furthermore, they also learn the diagonal pairs. It seems like the overall bias observed in participants with implicit knowledge is not present, and we see that participants know specifically for the parallel pairs they learned which orientation they had. (N explicit: 18; Proportion of explicit participants: 14.06%)

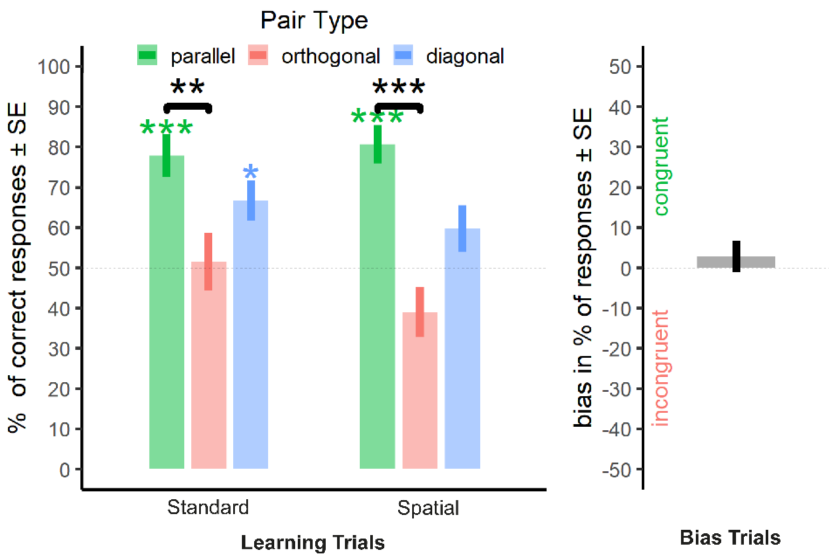

Figure D.2 Experiment 6a Explicit Results. The y-axis represents the explicit participants' mean performance on 2AFC trials. Error bars represent the standard error. The dotted line indicates the chance level of 50% or 0%. Stars represent the significance of the difference from chance. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.005$. The *Standard learning trials* was a standard learning test using one real pair from the training phase and one foil pair created by combining shapes of two real pairs. It measures learning of item co-occurrence. The *Spatial learning trials* showed the same real pair twice. Once in its correct orientation and once rotated by 90°. It measures learning of the spatial arrangement of learned pairs. The *Bias trials* showed the same foil pair twice. Once horizontally and once vertically. There is no correct response, and it measures bias for one of the orientations.

Experiment 6b. Descriptive Statistics: It seems likely that participants have learned all three types of pairs. They clearly seem to know the orientation of the parallel pairs, potentially also of the orthogonal pairs, but not of the diagonal pairs. In contrast to the implicit group, there seems to be no overall bias. (N explicit: 7; Proportion of explicit participants: 6.54%)

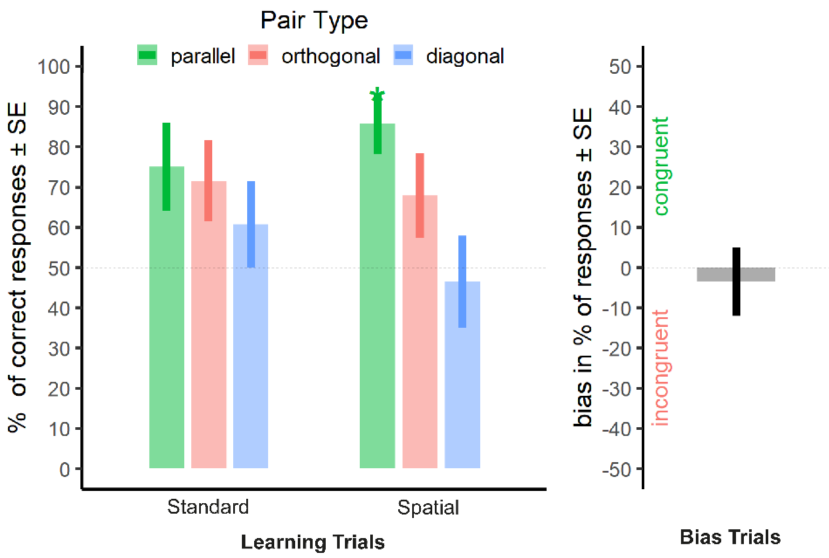

Figure D.3 Experiment 6b Explicit Results. The y-axis represents the explicit participants' mean performance on 2AFC trials. Error bars represent the standard error. The dotted line indicates the chance level of 50% or 0%. Stars represent the significance of the difference from chance. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.005$. The *Standard learning trials* was a standard learning test using one real pair from the training phase and one foil pair created by combining shapes of two real pairs. It measures learning of item co-occurrence. The *Spatial learning trials* showed the same real pair twice. Once in its correct orientation and once rotated by 90°. It measures learning of the spatial arrangement of learned pairs. The *Bias trials* showed the same foil pair twice. Once horizontally and once vertically. There is no correct response, and it measures bias for one of the orientations.

E. stVSL Supplementary Experiment 1: More Training

The experiments testing the role of temporal coherence in the implicit learning of spatial structure in Chapter 4 (Experiments 5a-c) used realized one arbitrary length training. To test whether similar results would emerge for a longer duration of training, supplemental Experiment 1 is a replication of Experiment 5a with twice the number of training scenes.

Participants

90 participants (38 female, mean age 28.9, SD = 8.7) were recruited via prolific.co. The total hourly compensation was £ 2.5. All participants had normal or corrected-to-normal vision. The sample size was chosen to match that of Experiment 5a. 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 identical to Experiment 5a.

Procedure

The procedure was identical to Experiment 5a, except for the amount of training scenes viewed. They were doubled from 72 to 144 scenes, increasing the amount of steps (motion animations) from 216 to 432. This left the relative number of steps along each motion direction and the relative number of partial presentations unchanged, meaning that the conditional and transitional statistics are also the same.

Results

Prior to analysis, three participants were removed for response bias, and 15 participants were removed for acquiring explicit knowledge of the structure of the task.

One-sample t-tests showed that performance for all of the three pair types was significantly different from chance: *parallel* (M(SE) = 55.7(2.0), $t(71) = 2.8$, $p = 0.006$, $d = 0.33$, $BF = 4.8$), *orthogonal* (M(SE) = 56.4(2.1), $t(71) = 3.0$, $p = 0.004$, $d = 0.36$, $BF = 8.0$), *diagonal* (M(SE) = 58.1 (2.3), $t(71) = 3.5$, $p < 0.001$, $d = 0.42$, $BF = 35$).

The results for all three pair types were not significantly different from the results of Experiment 5a: *parallel* ($t(139.99) = -0.45$, $p = 0.657$, $d = 0.07$, $BF = 0.20$), *orthogonal* ($t(139.17) = 0.25$, $p = 0.802$, $d = 0.04$, $BF = 0.19$), *diagonal* ($t(137.52) = 0.20$, $p = 0.845$, $d = 0.03$, $BF = 0.18$)

Discussion

We can see that the participants who received twice the amount of training in Supplementary Experiment 1 showed very similar results to the original participants in Experiment 5a. This suggests that our participants in Experiment 5a were already close to a ceiling effect, and more training did not increase implicit learning.

References

Abel, M., Nickl, A. T., Reßle, A., Unger, C., & Bäuml, K.-H. T. (2023). The role of sleep for memory consolidation: does sleep protect memories from retroactive interference? *Psychonomic Bulletin & Review*, *30*(6), 2296–2304. https://doi.org/10.3758/s13423-023-02264-8

Antovich, D. M., & Graf Estes, K. (2023). Statistical word segmentation: Anchoring learning across contexts. *Infancy*, *28*(2), 257–276. https://doi.org/10.1111/infa.12525

Arato, J., Rothkopf, C. A., & Fiser, J. (2023). Eye movements reflect active statistical learning. In *PsyArXiv*. https://doi.org/10.31234/osf.io/msxhd

Arciuli, J., & Simpson, I. C. (2012). Statistical learning is lasting and consistent over time. *Neuroscience Letters*, *517*(2), 133–135. https://doi.org/10.1016/j.neulet.2012.04.045

Arciuli, J., Torkildsen, J. von K., Stevens, D. J., & Simpson, I. C. (2014). Statistical learning under incidental versus intentional conditions. *Frontiers in Psychology*, *5*, 747. https://doi.org/10.3389/fpsyg.2014.00747

Ashby, F. G., & Maddox, W. T. (2005). Human Category Learning. *Annual Review of Psychology*, *56*(1), 149–178. https://doi.org/10.1146/annurev.psych.56.091103.070217

Ashby, F. G., & Maddox, W. T. (2011). Human category learning 2.0. *Annals of the New York Academy of Sciences*, *1224*, 147–161. https://doi.org/10.1111/j.1749-6632.2010.05874.x

Aslin, R. N., & Newport, E. L. (2012). Statistical Learning. *Current Directions in Psychological Science*, *21*(3), 170–176. https://doi.org/10.1177/0963721412436806

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of Conditional Probability Statistics by 8-Month-Old Infants. *Psychological Science*, *9*(4), 321–324. https://doi.org/10.1111/1467-9280.00063

Baillargeon, R. (2008). Innate Ideas Revisited: For a Principle of Persistence in Infants' Physical Reasoning. *Perspectives on Psychological Science*, *3*(1), 2–13. https://doi.org/10.1111/j.1745-6916.2008.00056.x

Baker, C. I., Olson, C. R., & Behrmann, M. (2004). Role of Attention and Perceptual Grouping in Visual Statistical Learning. *Psychological Science*, *15*(7), 460–466. https://doi.org/10.1111/j.0956-7976.2004.00702.x

Balas, B. J., Schmidt, J., & Saville, A. (2015). A face detection bias for horizontal orientations develops in middle childhood. *Frontiers in Psychology*, *6*, 772. https://doi.org/10.3389/fpsyg.2015.00772

Ball, F., Spuerck, I., & Noesselt, T. (2021). Minimal interplay between explicit knowledge, dynamics of learning and temporal expectations in different, complex uni- and multisensory contexts. *Attention, Perception & Psychophysics*, *83*(6), 2551–2573. https://doi.org/10.3758/s13414-021-02313-1

Barlow, H. B. (1989). Unsupervised Learning. *Neural Computation*, *1*(3), 295–311. https://doi.org/10.1162/neco.1989.1.3.295

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge University Press.

Batterink, L. J., Mulgrew, J., & Gibbings, A. (2023). Rhythmically Modulating Neural Entrainment during Exposure to Regularities Influences Statistical Learning. *Journal of Cognitive Neuroscience*, 1–21. https://doi.org/10.1162/jocn_a_02079

Batterink, L. J., & Paller, K. A. (2017). Online neural monitoring of statistical learning. *Cortex*, *90*, 31–45. https://doi.org/10.1016/j.cortex.2017.02.004

Bays, B. C., Turk-Browne, N. B., & Seitz, A. R. (2016). Dissociable behavioural outcomes of visual statistical learning. *Visual Cognition*, *23*(9–10), 1–26. https://doi.org/10.1080/13506285.2016.1139647

Behrens, T. E. J., Muller, T. H., Whittington, J. C. R., Mark, S., Baram, A. B., Stachenfeld, K. L., & Kurth-Nelson, Z. (2018). What Is a Cognitive Map? Organizing Knowledge for Flexible Behavior. *Neuron*, *100*(2), 490–509. https://doi.org/10.1016/j.neuron.2018.10.002

Ben-Yosef, G., Kreiman, G., & Ullman, S. (2020). Minimal videos: Trade-off between spatial and temporal information in human and machine vision. *Cognition*, *201*, 104263. https://doi.org/10.1016/j.cognition.2020.104263

Bertels, J., Destrebecqz, A., & Franco, A. (2015). Interacting Effects of Instructions and Presentation Rate on Visual Statistical Learning. *Frontiers in Psychology*, *6*, 1806. https://doi.org/10.3389/fpsyg.2015.01806

Bertels, J., Franco, A., & Destrebecqz, A. (2012). How Implicit Is Visual Statistical Learning? *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *38*(5), 1425–1431. https://doi.org/10.1037/a0027210

Blackburn, S. (2008). Abstraction. In *The Oxford Dictionary of Philosophy*. Oxford University Press.

Bloch, A., Tamir, D., Vakil, E., & Zeilig, G. (2016). Specific Deficit in Implicit Motor Sequence Learning following Spinal Cord Injury. *PloS One*, *11*(6), e0158396. https://doi.org/10.1371/journal.pone.0158396

Block, N. (1995). On a confusion about a function of consciousness. *The Behavioral and Brain Sciences*, *18*(2), 227–247. https://doi.org/10.1017/s0140525x00038188

Bruner, J. S., Goodnow, J. J., & Austin, G. A. (1956). *A study of thinking*. John Wiley and Sons.

Buysse, D. J., Reynolds, C. F., Monk, T. H., Berman, S. R., & Kupfer, D. J. (1989). The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research. *Psychiatry Research*, *28*(2), 193–213. https://doi.org/10.1016/0165-1781(89)90047-4

Castro-Rodrigues, P., Akam, T., Snorasson, I., Camacho, M., Paixão, V., Maia, A., Barahona-Corrêa, J. B., Dayan, P., Simpson, H. B., Costa, R. M., & Oliveira-Maia, A. J. (2022). Explicit knowledge of task structure is a primary determinant of human model-based action. *Nature Human Behaviour*, *6*(8), 1126–1141. https://doi.org/10.1038/s41562-022-01346-2

Chambers, A. M. (2017). The role of sleep in cognitive processing: focusing on memory consolidation. *Wiley Interdisciplinary Reviews. Cognitive Science*, *8*(3). https://doi.org/10.1002/wcs.1433

Champely, S., Ekstrom, C., Dalgaard, P., Gill, J., Weibelzahl, S., Anandkumar, A., Ford, C., Volcic, R., & De Rosario, H. (2017). *pwr: Basic functions for power analysis*. cran.r-project.org/package=pwr

Choi, D., Batterink, L. J., Black, A. K., Paller, K. A., & Werker, J. F. (2020). Preverbal Infants Discover Statistical Word Patterns at Similar Rates as Adults: Evidence From Neural Entrainment. *Psychological Science*, *31*(9), 1161–1173. https://doi.org/10.1177/0956797620933237

Chomsky, N. (1959). A Review of B. F. Skinner's Verbal Behavior. *Language*, *35*(1), 26–58. http://www.biolinguagem.com/ling_cog_cult/chomsky_1958_skinners_verbalbehavior.pdf

Chough, J., & Zinszer, B. D. (2022). Deep in the Trenches: First language performance predicts primacy in statistical learning of two structures. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *44*(44). https://escholarship.org/uc/item/48q5494m

Connell, L. (2018). What have labels ever done for us? The linguistic shortcut in conceptual processing. *Language, Cognition and Neuroscience*, *34*(10), 1–11. https://doi.org/10.1080/23273798.2018.1471512

Covington, N. V., Brown-Schmidt, S., & Duff, M. C. (2018). The Necessity of the Hippocampus for Statistical Learning. *Journal of Cognitive Neuroscience*, *30*(5), 680–697. https://doi.org/10.1162/jocn_a_01228

Dakin, S. C., & Watt, R. J. (2009). Biological "bar codes" in human faces. *Journal of Vision*, *9*(4), 2.1-10. https://doi.org/10.1167/9.4.2

Dale, R., Duran, N. D., & Morehead, J. R. (2012). Prediction during statistical learning, and implications for the implicit/explicit divide. *Advances in Cognitive Psychology*, *8*(2), 196–209. https://doi.org/10.2478/v10053-008-0115-z

Davis, C. P., & Yee, E. (2018). Features, labels, space, and time: factors supporting taxonomic relationships in the anterior temporal lobe and thematic relationships in the angular gyrus. *Language, Cognition and Neuroscience*, *34*(10), 1–11. https://doi.org/10.1080/23273798.2018.1479530

Denis, D., Schapiro, A. C., Poskanzer, C., Bursal, V., Charon, L., Morgan, A., & Stickgold, R. (2020). The roles of item exposure and visualization success in the consolidation of memories across wake and sleep. *Learning & Memory* , *27*(11), 451–456. https://doi.org/10.1101/lm.051383.120

Dewar, M. T., Cowan, N., & Sala, S. D. (2007). Forgetting due to retroactive interference: a fusion of Müller and Pilzecker's (1900) early insights into everyday forgetting and recent research on anterograde amnesia. *Cortex*, *43*(5), 616–634. https://doi.org/10.1016/s0010-9452(08)70492-1

Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews. Neuroscience*, *11*(2), 114–126. https://doi.org/10.1038/nrn2762

Dienes, Zoltán. (2007). Subjective measures of unconscious knowledge. *Progress in Brain Research*, *168*, 49–269. https://doi.org/10.1016/s0079-6123(07)68005-4

Dienes, Zoltán, & Berry, D. (1997). Implicit learning: Below the subjective threshold. *Psychonomic Bulletin & Review*, *4*(1), 3–23. https://doi.org/10.3758/BF03210769

Dienes, Zoltan, & Perner, J. (1999). A theory of implicit and explicit knowledge. *The Behavioral and Brain Sciences*, *22*(5), 735–808. https://doi.org/10.1017/s0140525x99002186

Djonlagic, I., Rosenfeld, A., Shohamy, D., Myers, C., Gluck, M., & Stickgold, R. (2009). Sleep enhances category learning. *Learning & Memory* , *16*(12), 751–755. https://doi.org/10.1101/lm.1634509

Durrant, S. J., Cairney, S. A., & Lewis, P. A. (2013). Overnight Consolidation Aids the Transfer of Statistical Knowledge from the Medial Temporal Lobe to the Striatum. *Cerebral Cortex*, *23*(10), 2467–2478. https://doi.org/10.1093/cercor/bhs244

Durrant, S. J., Cairney, S. A., & Lewis, P. A. (2016). Cross-modal transfer of statistical information benefits from sleep. *Cortex*, *78*, 85–99. https://doi.org/10.1016/j.cortex.2016.02.011

Durrant, S. J., Taylor, C., Cairney, S., & Lewis, P. A. (2011). Sleep-dependent consolidation of statistical learning. *Neuropsychologia*, *49*(5), 1322–1331. https://doi.org/10.1016/j.neuropsychologia.2011.02.015

Eckstein, M. K., & Collins, A. G. E. (2020). Computational evidence for hierarchically structured reinforcement learning in humans. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(47), 29381–29389. https://doi.org/10.1073/pnas.1912330117

Ellenbogen, J. M., Hulbert, J. C., Stickgold, R., Dinges, D. F., & Thompson-Schill, S. L. (2006). Interfering with theories of sleep and memory: sleep, declarative memory, and associative interference. *Current Biology*, *16*(13), 1290–1294. https://doi.org/10.1016/j.cub.2006.05.024

Ellis, K., Wong, C., Nye, M., Sablé-Meyer, M., Morales, L., Hewitt, L., Cary, L., Solar-Lezama, A., & Tenenbaum, J. B. (2021). DreamCoder: bootstrapping inductive program synthesis with wake-sleep library learning. *Proceedings of the 42nd ACM SIGPLAN International Conference on Programming Language Design and Implementation*, 835–850. https://doi.org/10.1145/3453483.3454080

Emberson, L. L., & Rubinstein, D. Y. (2016). Statistical learning is constrained to less abstract patterns in complex sensory input (but not the least). *Cognition*, *153*, 63–78. https://doi.org/10.1016/j.cognition.2016.04.010

Endress, A. D., & Johnson, S. P. (2021). When forgetting fosters learning: A neural network model for statistical learning. *Cognition*, *213*, 104621. https://doi.org/10.1016/j.cognition.2021.104621

Fischer, S., Drosopoulos, S., Tsen, J., & Born, J. (2006). Implicit Learning–Explicit Knowing: A Role for Sleep in Memory System Interaction. *Journal of Cognitive Neuroscience*, *18*(3), 311–319. https://doi.org/10.1162/jocn.2006.18.3.311

Fiser, J., & Aslin, R. N. (2001). Unsupervised Statistical Learning of Higher-Order Spatial Structures from Visual Scenes. *Psychological Science*, *12*(6), 499–504. https://doi.org/10.1111/1467-9280.00392

Fiser, J., & Aslin, R. N. (2002a). Statistical Learning of Higher-Order Temporal Structure From Visual Shape Sequences. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *28*(3), 458–467. https://doi.org/10.1037/0278-7393.28.3.458

Fiser, J., & Aslin, R. N. (2002b). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, *99*(24), 15822–15826. https://doi.org/10.1073/pnas.232472899

Fiser, J., & Aslin, R. N. (2005). Encoding Multielement Scenes: Statistical Learning of Visual Feature Hierarchies. *Journal of Experimental Psychology. General*, *134*(4), 521–537. https://doi.org/10.1037/0096-3445.134.4.521

Fiser, J., & Lengyel, G. (2022). Statistical Learning in Vision. *Annual Review of Vision Science*, *8*(17), 1–26. https://doi.org/10.1146/annurev-vision-100720- 03343

Foulsham, T., Kingstone, A., & Underwood, G. (2008). Turning the world around: patterns in saccade direction vary with picture orientation. *Vision Research*, *48*(17), 1777–1790. https://doi.org/10.1016/j.visres.2008.05.018

Gallistel, C. R., Fairhurst, S., & Balsam, P. (2004). The learning curve: implications of a quantitative analysis. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(36), 13124–13131. https://doi.org/10.1073/pnas.0404965101

Geambașu, A., Spit, S., van Renswoude, D., Blom, E., Fikkert, P. J. P. M., Hunnius, S., Junge, C. C. M. M., Verhagen, J., Visser, I., Wijnen, F., & Levelt, C. C. (2023). Robustness of the rule-learning effect in 7-month-old infants: A close, multicenter replication of Marcus et al. (1999). *Developmental Science*, *26*(1), e13244. https://doi.org/10.1111/desc.13244

Gepshtein, S., & Kubovy, M. (2000). The emergence of visual objects in space-time. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(14), 8186–8191. https://doi.org/10.1073/pnas.97.14.8186

Gilboa, A., & Marlatte, H. (2017). Neurobiology of Schemas and Schema-Mediated Memory. *Trends in Cognitive Sciences*, *21*(8), 618–631. https://doi.org/10.1016/j.tics.2017.04.013

Gilchrist, I. D., & Harvey, M. (2006). Evidence for a systematic component within scan paths in visual search. *Visual Cognition*, *14*(4–8), 704–715. https://doi.org/10.1080/13506280500193719

Gómez Varela, I., Orpella, J., Poeppel, D., Ripolles, P., & Assaneo, M. F. (2024). Syllabic rhythm and prior linguistic knowledge interact with individual differences to modulate phonological statistical learning. *Cognition*, *245*, 105737. https://doi.org/10.1016/j.cognition.2024.105737

Goujon, A., Didierjean, A., & Poulet, S. (2014). The emergence of explicit knowledge from implicit learning. *Memory & Cognition*, *42*(2), 225–236. https://doi.org/10.3758/s13421-013-0355-0

Hallgató, E., Győri-Dani, D., Pekár, J., Janacsek, K., & Nemeth, D. (2013). The differential consolidation of perceptual and motor learning in skill acquisition. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *49*(4), 1073–1081. https://doi.org/10.1016/j.cortex.2012.01.002

Hazenberg, S. J., Jongsma, M. L. A., Koning, A., & van Lier, R. (2014). Differential Familiarity Effects in Amodal Completion: Support From Behavioral and Electrophysiological Measurements. *Journal of Experimental Psychology. Human Perception and Performance*, *40*(2), 669–684. https://doi.org/10.1037/a0034689

Hazenberg, S. J., & van Lier, R. (2016). Disentangling effects of structure and knowledge in perceiving partly occluded shapes: An ERP study. *Vision Research*, *126*, 109–119. https://doi.org/10.1016/j.visres.2015.10.004

Heald, J. B., Lengyel, M., & Wolpert, D. M. (2021). Contextual inference underlies the learning of sensorimotor repertoires. *Nature*, *600*(7889), 489–493. https://doi.org/10.1038/s41586-021-04129-3

Henin, S., Turk-Browne, N. B., Friedman, D., Liu, A., Dugan, P., Flinker, A., Doyle, W., Devinsky, O., & Melloni, L. (2021). Learning hierarchical sequence representations across human cortex and hippocampus. *Science Advances*, *7*(8), eabc4530. https://doi.org/10.1126/sciadv.abc4530

Hennies, N., Lewis, P. A., Durrant, S. J., Cousins, J. N., & Ralph, M. A. L. (2014). Time- but not sleep-dependent consolidation promotes the emergence of cross-modal conceptual representations. *Neuropsychologia*, *63*, 116–123. https://doi.org/10.1016/j.neuropsychologia.2014.08.021

Himberger, K. D., Finn, A. S., & Honey, C. J. (2019). Reconsidering the automaticity of visual statistical learning. In *PsyArXiv*. https://doi.org/10.31234/osf.io/r659w

Hinton, G. (2014). Where do features come from? *Cognitive Science*, *38*(6), 1078–1101. https://doi.org/10.1111/cogs.12049

Hinton, G. E. (2010). Learning to represent visual input. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *365*(1537), 177–184. https://doi.org/10.1098/rstb.2009.0200

Hinton, G. E., Dayan, P., Frey, B. J., & Neal, R. M. (1995). The "Wake-Sleep" Algorithm for Unsupervised Neural Networks. *Science*, *268*(5214), 1158–1161. https://doi.org/10.1126/science.7761831

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. *Political Analysis: An Annual Publication of the Methodology Section of the American Political Science Association*, *15*(3), 199–236. https://doi.org/10.1093/pan/mpl013

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011). MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. *Journal of Statistical Software*, *42*(8). https://doi.org/10.18637/jss.v042.i08

Hochberg, J. (1968). In the mind's eye. In R. N. Haber (Ed.), *Contemporary theory and research in visual perception*. Holt, Rinehart & Winston.

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics, Theory and Applications*, *6*(2). http://www.jstor.org/stable/4615733

Ingram, K. M., Mickes, L., & Wixted, J. T. (2012). Recollection Can Be Weak and Familiarity Can Be Strong. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *38*(2), 325–339. https://doi.org/10.1037/a0025483

Jiang, Y., Gu, S., Murphy, K., & Finn, C. (2019). Language as an Abstraction for Hierarchical Deep Reinforcement Learning. *Proceedings of the 33rd International Conference on Neural Information Processing Systems*.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, *14*(2), 201–211. https://doi.org/10.3758/bf03212378

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Jun, J., & Chong, S. C. (2016). Visual statistical learning of temporal structures at different hierarchical levels. *Attention, Perception & Psychophysics*, *78*(5), 1308–1323. https://doi.org/10.3758/s13414-016-1104-9

Jun, J., & Chong, S. C. (2018). Visual statistical learning at basic and subordinate category levels in real-world images. *Attention, Perception & Psychophysics*, *80*(8), 1946–1961. https://doi.org/10.3758/s13414-018-1566-z

Jung, Y., Walther, D. B., & Finn, A. S. (2021). Children automatically abstract categorical regularities during statistical learning. *Developmental Science*, *24*(5), e13072. https://doi.org/10.1111/desc.13072

Jurjut, O., Georgieva, P., Busse, L., & Katzner, S. (2017). Learning Enhances Sensory Processing in Mouse V1 before Improving Behavior. *The Journal of Neuroscience*, *37*(27), 6460–6474. https://doi.org/10.1523/JNEUROSCI.3485-16.2017

Kanizsa, G. (1985). Seeing and thinking. *Acta Psychologica*, *59*(1), 23--33.

Kemp, C., & Tenenbaum, J. B. (2008). The discovery of structural form. *Proceedings of the National Academy of Sciences*, *105*(31), 10687–10692. https://doi.org/10.1073/pnas.0802631105

Kim, R., Seitz, A., Feenstra, H., & Shams, L. (2009). Testing assumptions of statistical learning: Is it long-term and implicit? *Neuroscience Letters*, *461*(2), 145–149. https://doi.org/10.1016/j.neulet.2009.06.030

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition*, *83*(2), B35–B42. https://doi.org/10.1016/s0010-0277(02)00004-5

Kirkham, N. Z., Slemmer, J. A., Richardson, D. C., & Johnson, S. P. (2007). Location, Location, Location: Development of Spatiotemporal Sequence Learning in Infancy. *Child Development*, *78*(5), 1559–1571. https://doi.org/10.1111/j.1467-8624.2007.01083.x

Kliegl, O., & Bäuml, K.-H. T. (2020). Buildup and Release from Proactive Interference - Cognitive and Neural Mechanisms. *Neuroscience and Biobehavioral Reviews*, *120*, 264–278. https://doi.org/10.1016/j.neubiorev.2020.10.028

Klinzing, J. G., Niethard, N., & Born, J. (2019). Mechanisms of systems memory consolidation during sleep. *Nature Neuroscience*, *22*(10), 1598–1610. https://doi.org/10.1038/s41593-019-0467-3

Kóbor, A., Horváth, K., Kardos, Z., Nemeth, D., & Janacsek, K. (2020). Perceiving structure in unstructured stimuli: Implicitly acquired prior knowledge impacts the processing of unpredictable transitional probabilities. *Cognition*, *205*, 104413. https://doi.org/10.1016/j.cognition.2020.104413

Kuchibhotla, K. V., Hindmarsh Sten, T., Papadoyannis, E. S., Elnozahy, S., Fogelson, K. A., Kumar, R., Boubenec, Y., Holland, P. C., Ostojic, S., & Froemke, R. C. (2019). Dissociating task acquisition from expression during learning reveals latent knowledge. *Nature Communications*, *10*(1), 2151. https://doi.org/10.1038/s41467-019-10089-0

Kumaran, D., & McClelland, J. L. (2012). Generalization Through the Recurrent Interaction of Episodic Memories. *Psychological Review*, *119*(3), 573–616. https://doi.org/10.1037/a0028681

Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. *Science*, *350*(6266), 1332–1338. https://doi.org/10.1126/science.aab3050

Lee, A. L. F., Liu, Z., & Lu, H. (2021). Parts beget parts: Bootstrapping hierarchical object representations through visual statistical learning. *Cognition*, *209*, 104515. https://doi.org/10.1016/j.cognition.2020.104515

Leeuw, J. R. de. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1–12. https://doi.org/10.3758/s13428-014-0458-y

Lengyel, G., Nagy, M., & Fiser, J. (2021). Statistically defined visual chunks engage object-based attention. *Nature Communications*, *12*(1), 272. https://doi.org/10.1038/s41467-020-20589-z

Lengyel, G., Žalalytė, G., Pantelides, A., Ingram, J. N., Fiser, J., Lengyel, M., & Wolpert, D. M. (2019). Unimodal statistical learning produces multimodal object-like representations. *ELife*, *8*, e43942. https://doi.org/10.7554/elife.43942

Lerner, I., & Gluck, M. (2019). Sleep and the Extraction of Hidden Regularities: A Systematic Review and the Importance of Temporal Rules. *Sleep Medicine Reviews*, *47*, 39–50. https://doi.org/10.1016/j.smrv.2019.05.004

Lewis, P. A., & Durrant, S. J. (2011). Overlapping memory replay during sleep builds cognitive schemata. *Trends in Cognitive Sciences*, *15*(8), 343–351. https://doi.org/10.1016/j.tics.2011.06.004

Lim, A., & Sinnett, S. (2012). Reexamining Visual Orientation Anisotropies: A Bias Towards Simple Horizontal Stimuli on Temporal Order Judgments. *Proceedings of the Annual Meeting of the Cognitive Science Societ, 34(34)*, 1912–1917.

Liu, H., Forest, T. A., Duncan, K., & Finn, A. S. (2023). What sticks after statistical learning: The persistence of implicit versus explicit memory traces. *Cognition*, *236*, 105439. https://doi.org/10.1016/j.cognition.2023.105439

Liu, Y., Dolan, R. J., Kurth-Nelson, Z., & Behrens, T. E. J. (2019). Human Replay Spontaneously Reorganizes Experience. *Cell*, *178*(3), 640-652.e14. https://doi.org/10.1016/j.cell.2019.06.012

Liu, Y., Mattar, M. G., Behrens, T. E. J., Daw, N. D., & Dolan, R. J. (2021). Experience replay is associated with efficient nonlocal learning. *Science*, *372*(6544). https://doi.org/10.1126/science.abf1357

Lupyan, G., & Lewis, M. (2017). From words-as-mappings to words-as-cues: the role of language in semantic knowledge. *Language, Cognition and Neuroscience*, *34*(10), 1–19. https://doi.org/10.1080/23273798.2017.1404114

Lutz, N. D., Diekelmann, S., Hinse-Stern, P., Born, J., & Rauss, K. (2017). Sleep Supports the Slow Abstraction of Gist from Visual Perceptual Memories. *Scientific Reports*, *7*(1), 42950. https://doi.org/10.1038/srep42950

Lutz, N. D., Wolf, I., Hübner, S., Born, J., & Rauss, K. (2018). Sleep Strengthens Predictive Sequence Coding. *The Journal of Neuroscience*, *38*(42), 8989–9000. https://doi.org/10.1523/jneurosci.1352-18.2018

Maheu, M., Meyniel, F., & Dehaene, S. (2022). Rational arbitration between statistics and rules in human sequence processing. *Nature Human Behaviour*, *6*(8), 1087–1103. https://doi.org/10.1038/s41562-021-01259-6

Makowski, D. (2018). The psycho package: An efficient and publishing-oriented workflow for psychological science. *Journal of Open Source Software*, *3*(22), 470. https://doi.org/10.21105/joss.00470

Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, *283*(5398), 77–80. https://doi.org/10.1126/science.283.5398.77

Mareschal, D., & French, R. M. (2017). TRACX2: a connectionist autoencoder using graded chunks to model infant visual statistical learning. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *372*(1711), 20160057. https://doi.org/10.1098/rstb.2016.0057

Mattar, M. G., & Daw, N. D. (2018). Prioritized memory access explains planning and hippocampal replay. *Nature Neuroscience*, *21*(11), 1609–1617. https://doi.org/10.1038/s41593-018-0232-z

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why There Are Complementary Learning Systems in the Hippocampus and Neocortex: Insights From the Successes and Failures of Connectionist Models of Learning and Memory. *Psychological Review*, *102*(3), 419–457. https://doi.org/10.1037/0033-295x.102.3.419

McDevitt, E. A., Zhang, J., MacKenzie, K. J., Fiser, J., & Mednick, S. C. (2022). The effect of interference, offline sleep, and wake on spatial statistical learning. *Neurobiology of Learning and Memory*, *193*, 107650. https://doi.org/10.1016/j.nlm.2022.107650

Meijman, T. F., de Vries-Griever, A. H., & de Vries, G. (1988). *The evaluation of the Groningen Sleep Quality Scale*. Heymans Bulletin.

Melton, A. W. (1963). Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior*, *2*(1), 1–21. https://doi.org/10.1016/s0022-5371(63)80063-8

Moreau, C. N., Joanisse, M. F., Mulgrew, J., & Batterink, L. J. (2022). No statistical learning advantage in children over adults: Evidence from behaviour and neural entrainment. *Developmental Cognitive Neuroscience*, *57*, 101154. https://doi.org/10.1016/j.dcn.2022.101154

Morgan, M. J., Findlay, J. M., & Watt, R. J. (1981). Aperture Viewing: A Review and a Synthesis. *The Quarterly Journal of Experimental Psychology*, *34*(2), 211–233. https://doi.org/10.1080/14640748208400837

Németh, D., Gerbier, E., Born, J., Rickard, T., Diekelmann, S., Fogel, S., Genzel, L., Prehn-Kristensen, A., Payne, J., Dresler, M., Simor, P., Mazza, S., Hoedlmoser, K., Ruby, P., Spencer, R. M. C., Albouy, G., Vékony, T., Schabus, M., & Janacsek, K. (2023). Optimizing the methodology of human sleep and memory research. *Nature Reviews Psychology*, 1–15. https://doi.org/10.1038/s44159-023-00262-0

Nemeth, D., Janacsek, K., Londe, Z., Ullman, M. T., Howard, D. V., & Howard, J. H., Jr. (2010). Sleep has no critical role in implicit motor sequence learning in young and old adults. *Experimental Brain Research*, *201*(2), 351–358. https://doi.org/10.1007/s00221-009-2024-x

Orbán, G., Fiser, J., Aslin, R. N., & Lengyel, M. (2008). Bayesian learning of visual chunks by human observers. *Proceedings of the National Academy of Sciences*, *105*(7), 2745–2750. https://doi.org/10.1073/pnas.0708424105

O'Reilly, R. C., Bhattacharyya, R., Howard, M. D., & Ketz, N. (2014). Complementary Learning Systems. *Cognitive Science*, *38*(6), 1229–1248. https://doi.org/10.1111/j.1551-6709.2011.01214.x

Piaget, J. (1954). *The Construction of Reality in the Child*. Basic Books.

Pöhlchen, D., & Schönauer, M. (2020). Sleep-dependent memory consolidation in the light of rapid neocortical plasticity. *Current Opinion in Behavioral Sciences*, *33*, 118–125. https://doi.org/10.1016/j.cobeha.2020.02.001

Pothos, E. M. (2007). Theories of Artificial Grammar Learning. *Psychological Bulletin*, *133*(2), 227–244. https://doi.org/10.1037/0033-2909.133.2.227

Preston, A. R., & Eichenbaum, H. (2013). Interplay of Hippocampus and Prefrontal Cortex in Memory. *Current Biology*, *23*(17), R764–R773. https://doi.org/10.1016/j.cub.2013.05.041

Quentin, R., Fanuel, L., Kiss, M., Vernet, M., Vékony, T., Janacsek, K., Cohen, L. G., & Nemeth, D. (2021). Statistical learning occurs during practice while high-order rule learning during rest period. *Npj Science of Learning*, *6*(1), 14. https://doi.org/10.1038/s41539-021-00093-9

Rasch, B., & Born, J. (2013). About Sleep's Role in Memory. *Physiological Reviews*, *93*(2), 681–766. https://doi.org/10.1152/physrev.00032.2012

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, *6*(6), 855–863. https://doi.org/10.1016/s0022-5371(67)80149-x

Reber, A. S. (1989). Implicit Learning and Tacit Knowledge. *Journal of Experimental Psychology. General*, *118*(3), 219–235. https://doi.org/10.1037/0096-3445.118.3.219

Richler, J. J., & Palmeri, T. J. (2014). Visual category learning. *Wiley Interdisciplinary Reviews. Cognitive Science*, *5*(1), 75–94. https://doi.org/10.1002/wcs.1268

Robertson, E. M., Pascual-Leone, A., & Press, D. Z. (2004). Awareness Modifies the Skill-Learning Benefits of Sleep. *Current Biology*, *14*(3), 208–212. https://doi.org/10.1016/j.cub.2004.01.027

Rogers, L. L., Park, S. H., & Vickery, T. J. (2021). Visual statistical learning is modulated by arbitrary and natural categories. *Psychonomic Bulletin & Review*, 1–8. https://doi.org/10.3758/s13423-021-01917-w

Rolls, E. T. (2012). Invariant Visual Object and Face Recognition: Neural and Computational Bases, and a Model, VisNet. *Frontiers in Computational Neuroscience*, *6*, 35. https://doi.org/10.3389/fncom.2012.00035

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*(5), 356–374. https://doi.org/10.1016/j.jmp.2012.08.001

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. https://doi.org/10.3758/pbr.16.2.225

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928. http://doi.org/10.1126/science.274.5294.1926

Schapiro, A. C., Gregory, E., Landau, B., McCloskey, M., & Turk-Browne, N. B. (2014). The Necessity of the Medial Temporal Lobe for Statistical Learning. *Journal of Cognitive Neuroscience*, *26*(8), 1736–1747. https://doi.org/10.1162/jocn_a_00578

Schapiro, A. C., Kustner, L. V., & Turk-Browne, N. B. (2012). Shaping of Object Representations in the Human Medial Temporal Lobe Based on Temporal

Regularities. *Current Biology*, *22*(17), 1622–1627. https://doi.org/10.1016/j.cub.2012.06.056

Schapiro, A. C., McDevitt, E. A., Chen, L., Norman, K. A., Mednick, S. C., & Rogers, T. T. (2017). Sleep Benefits Memory for Semantic Category Structure While Preserving Exemplar-Specific Information. *Scientific Reports*, *7*(1), 14869. https://doi.org/10.1038/s41598-017-12884-5

Schapiro, A. C., McDevitt, E. A., Rogers, T. T., Mednick, S. C., & Norman, K. A. (2018). Human hippocampal replay during rest prioritizes weakly learned information and predicts memory performance. *Nature Communications*, *9*(1), 3920. https://doi.org/10.1038/s41467-018-06213-1

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

Schapiro, A. C., Turk-Browne, N. B., Norman, K. A., & Botvinick, M. M. (2016). Statistical learning of temporal community structure in the hippocampus. *Hippocampus*, *26*(1), 3–8. https://doi.org/10.1002/hipo.22523

Sherman, B. E., Aljishi, A., Graves, K. N., Quraishi, I. H., Sivaraju, A., Damisah, E. C., & Turk-Browne, N. B. (2023). Intracranial Entrainment Reveals Statistical Learning across Levels of Abstraction. *Journal of Cognitive Neuroscience*, *35*(8), 1312–1328. https://doi.org/10.1162/jocn_a_02012

Sherman, B. E., Turk-Browne, N. B., & Goldfarb, E. V. (2024). Multiple Memory Subsystems: Reconsidering Memory in the Mind and Brain. *Perspectives on Psychological Science*, *19*(1), 103–125. https://doi.org/10.1177/17456916231179146

Siegelman, N., Bogaerts, L., Kronenfeld, O., & Frost, R. (2018). Redefining “Learning” in Statistical Learning: What Does an Online Measure Reveal About the Assimilation of Visual Regularities? *Cognitive Science*, *42*, 692–727. https://doi.org/10.1111/cogs.12556

Simor, P., Zavecz, Z., Horváth, K., Éltető, N., Török, C., Pesthy, O., Gombos, F., Janacsek, K., & Nemeth, D. (2019). Deconstructing Procedural Memory: Different Learning Trajectories and Consolidation of Sequence and Statistical Learning. *Frontiers in Psychology*, *9*, 2708. https://doi.org/10.3389/fpsyg.2018.02708

Singh, D., Norman, K. A., & Schapiro, A. C. (2022). A model of autonomous interactions between hippocampus and neocortex driving sleep-dependent memory consolidation. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(44), e2123432119. https://doi.org/10.1073/pnas.2123432119

Sloutsky, V. M., & Deng, W. (sophia). (2017). Categories, concepts, and conceptual development. *Language, Cognition and Neuroscience*, *34*(10), 1–14. https://doi.org/10.1080/23273798.2017.1391398

Squire, L. R., Genzel, L., Wixted, J. T., & Morris, R. G. (2015). Memory Consolidation. *Cold Spring Harbor Perspectives in Biology*, *7*(8), a021766. https://doi.org/10.1101/cshperspect.a021766

Stärk, K., Kidd, E., & Frost, R. L. A. (2022). Close encounters of the word kind: Attested distributional information boosts statistical learning. *Language Learning*, *73*(2), 341–373. https://doi.org/10.1111/lang.12523

Stone, J. V. (1998). Object recognition using spatiotemporal signatures. *Vision Research*, *38*(7), 947–951. https://doi.org/10.1016/s0042-6989(97)00301-5

Sun, W., Advani, M., Spruston, N., Saxe, A., & Fitzgerald, J. E. (2023). Organizing memories for generalization in complementary learning systems. *Nature Neuroscience*, *26*(8), 1438–1448. https://doi.org/10.1038/s41593-023-01382-9

Sweegers, C. C. G., Takashima, A., Fernández, G., & Talamini, L. M. (2014). Neural mechanisms supporting the extraction of general knowledge across episodic memories. *NeuroImage*, *87*, 138–146. https://doi.org/10.1016/j.neuroimage.2013.10.063

Sznabel, D., Land, R., Kopp, B., & Kral, A. (2023). The relation between implicit statistical learning and proactivity as revealed by EEG. *Scientific Reports*, *13*(1), 15787. https://doi.org/10.1038/s41598-023-42116-y

Tandoc, M. C., Bayda, M., Poskanzer, C., Cho, E., Cox, R., Stickgold, R., & Schapiro, A. C. (2021). Examining the effects of time of day and sleep on generalization. *PloS One*, *16*(8), e0255423. https://doi.org/10.1371/journal.pone.0255423

Tatler, B. W., & Vincent, B. T. (2008). Systematic tendencies in scene viewing. *Journal of Eye Movement Research*, *2*(2). https://doi.org/10.16910/jemr.2.2.5

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to Grow a Mind: Statistics, Structure, and Abstraction. *Science*, *331*(6022), 1279–1285. https://doi.org/10.1126/science.1192788

Thiessen, E. D. (2017). What's statistical about learning? Insights from modelling statistical learning as a set of memory processes. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *372*(1711), 20160056. https://doi.org/10.1098/rstb.2016.0056

Tolman, E. C. (1948). Cognitive Maps in Rats and Men. *Psychological Review*, *55*(4), 189–208.

Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., Witter, M. P., & Morris, R. G. M. (2007). Schemas and Memory Consolidation. *Science*, *316*(5821), 76–82. https://doi.org/10.1126/science.1135935

Tummeltshammer, K., Amso, D., French, R. M., & Kirkham, N. Z. (2017). Across space and time: infants learn from backward and forward visual statistics. *Developmental Science*, *20*(5), e12474. https://doi.org/10.1111/desc.12474

Turk-Browne, N. B., Jungé, J. A., & Scholl, B. J. (2005). The Automaticity of Visual Statistical Learning. *Journal of Experimental Psychology. General*, *134*(4), 552–564. https://doi.org/10.1037/0096-3445.134.4.552

Turk-Browne, N. B., & Scholl, B. J. (2009). Flexible Visual Statistical Learning: Transfer Across Space and Time. *Journal of Experimental Psychology. Human Perception and Performance*, *35*(1), 195–202. https://doi.org/10.1037/0096-1523.35.1.195

Uji, M., & Tamaki, M. (2023). Sleep, learning, and memory in human research using noninvasive neuroimaging techniques. *Neuroscience Research*, *189*, 66–74. https://doi.org/10.1016/j.neures.2022.12.013

Ullman, T. D., Stuhlmüller, A., Goodman, N. D., & Tenenbaum, J. B. (2018). Learning physical parameters from dynamic scenes. *Cognitive Psychology*, *104*, 57–82. https://doi.org/10.1016/j.cogpsych.2017.05.006

van Lier, R., & Gerbino, W. (2015). *Perceptual completions* (J. Wagemans, Ed.). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199686858.013.040

Van Renswoude, D. R., Johnson, S. P., Raijmakers, M. E. J., & Visser, I. (2016). Do infants have the horizontal bias? *Infant Behavior & Development*, *44*, 38–48. https://doi.org/10.1016/j.infbeh.2016.05.005

Wagner, U., Gais, S., Haider, H., Verleger, R., & Born, J. (2004). Sleep inspires insight. *Nature*, *427*(6972), 352–355. https://doi.org/10.1038/nature02223

Wallis, G., & Rolls, E. T. (1997). Invariant Face And Object Recognition In The Visual System. *Progress in Neurobiology*, *51*(2), 167–194. https://doi.org/10.1016/s0301-0082(96)00054-8

Winocur, G., Moscovitch, M., & Bontempi, B. (2010). Memory formation and long-term retention in humans and animals: Convergence towards a transformation account of hippocampal–neocortical interactions. *Neuropsychologia*, *48*(8), 2339–2356. https://doi.org/10.1016/j.neuropsychologia.2010.04.016

Witkowski, S., Schechtman, E., & Paller, K. A. (2020). Examining sleep's role in memory generalization and specificity through the lens of targeted memory reactivation. *Current Opinion in Behavioral Sciences*, *33*, 86–91. https://doi.org/10.1016/j.cobeha.2020.01.007

Wittkuhn, L., Chien, S., Hall-McMaster, S., & Schuck, N. W. (2021). Replay in minds and machines. *Neuroscience and Biobehavioral Reviews*, *129*, 367–388. https://doi.org/10.1016/j.neubiorev.2021.08.002

Wixted, J. T. (2004). The Psychology and Neuroscience of Forgetting. *Annual Review of Psychology*, *55*(1), 235–269. https://doi.org/10.1146/annurev.psych.55.090902.141555

Xu, Z., Theeuwes, J., & Los, S. A. (2023). Statistical learning of spatiotemporal regularities dynamically guides visual attention across space. *Attention, Perception & Psychophysics*, *85*(4), 1054–1072. https://doi.org/10.3758/s13414-022-02573-5

Yan, C., Ehinger, B. V., Pérez-Bellido, A., Peelen, M. V., & de Lange, F. P. (2023). Humans predict the forest, not the trees: statistical learning of spatiotemporal structure in visual scenes. *Cerebral Cortex* , *33*(13), 8300–8311. https://doi.org/10.1093/cercor/bhad115

Yang, J., & Li, P. (2012). Brain Networks of Explicit and Implicit Learning. *PloS One*, *7*(8), e42993. https://doi.org/10.1371/journal.pone.0042993

Yonelinas, A. P. (2002). The Nature of Recollection and Familiarity: A Review of 30 Years of Research. *Journal of Memory and Language*, *46*(3), 441–517. https://doi.org/10.1006/jmla.2002.2864

Yu, R. Q., & Zhao, J. (2018). Object representations are biased toward each other through statistical learning. *Visual Cognition*, *26*(4), 1–15. https://doi.org/10.1080/13506285.2018.1435596

Yun, X., Hazenberg, S. J., & van Lier, R. (2018). Temporal properties of amodal completion: Influences of knowledge. *Vision Research*, *145*, 21–30. https://doi.org/10.1016/j.visres.2018.02.011

Zander, T., Volz, K. G., Born, J., & Diekelmann, S. (2017). Sleep increases explicit solutions and reduces intuitive judgments of semantic coherence. *Learning & Memory* , *24*(12), 641–645. https://doi.org/10.1101/lm.044511.116

Zettersten, M., Potter, C. E., & Saffran, J. R. (2020). Tuning in to non-adjacencies: Exposure to learnable patterns supports discovering otherwise difficult structures. *Cognition*, *202*, 104283. https://doi.org/10.1016/j.cognition.2020.104283

Zhao, J., Ngo, N., McKendrick, R., & Turk-Browne, N. B. (2011). Mutual Interference Between Statistical Summary Perception and Statistical Learning. *Psychological Science*, *22*(9), 1212–1219. https://doi.org/10.1177/0956797611419304

Zhao, J., & Yu, R. Q. (2016). Statistical regularities reduce perceived numerosity. *Cognition*, *146*, 217–222. https://doi.org/10.1016/j.cognition.2015.09.018

Zhuang, C., Yan, S., Nayebi, A., Schrimpf, M., Frank, M. C., DiCarlo, J. J., & Yamins, D. L. K. (2021). Unsupervised neural network models of the ventral visual stream.

Proceedings of the National Academy of Sciences of the United States of America, *118*(3), e2014196118. https://doi.org/10.1073/pnas.2014196118

Zolnai, T., Dávid, D. R., Pesthy, O., Nemeth, M., Kiss, M., Nagy, M., & Nemeth, D. (2022). Measuring statistical learning by eye-tracking. *Experimental Results*, *3*. https://doi.org/10.1017/exp.2022.8