Processing-Based Measures of Implicit Statistical Learning of Artificial Grammars

Holly E. Jenkins

Submitted for the degree of Doctor of Philosophy

Institute of Biosciences

Newcastle University

June 2022

General Abstract

Implicit statistical learning, whereby predictable relationships between stimuli are detected without conscious awareness, is important for language acquisition. Although a defining feature of implicit learning is that we are unaware that learning has occurred, implicit statistical learning is often assessed using measures that require explicit reflection (e.g., judgements about the grammaticality of a sequence of stimuli). However, implicit statistical learning can also be assessed without requiring conscious reflection, using 'processing-based' tasks, that instead measure other processes that are facilitated by implicit statistical learning such as reaction times or serial recall. Processing-based measures would be particularly valuable for measuring implicit statistical learning in populations such as children (who may be less adept at explicitly reflecting on implicitly learned knowledge) or individuals with dyslexia (which has been associated with a specific deficit in implicit statistical learning, rather than in explicitly reflecting on implicitly learned information). In this thesis, I developed and tested novel processing-based measures of implicit statistical learning that do not require conscious reflection and combined them with traditional reflection-based measures of learning to investigate the nature of the knowledge acquired by way of implicit statistical learning. I found evidence of implicit statistical learning across a number of experiments, which also suggest that the complexity of the grammar being learned may affect the extent to which implicit and explicit processes are recruited during the tasks. I then applied this novel serial visual recall paradigm to provide evidence that implicit statistical learning abilities are consistent across children aged 8 to 15 years and that serial visual recall may capture differences between children and adults that are not reflected in traditional measures of learning. Finally, I applied these paradigms across a number of experiments to assess differences in implicit statistical learning between individuals with and without dyslexia or other reading difficulties, and found no evidence that dyslexia is associated with a deficit in implicit statistical learning. Overall, these experiments suggest that processing-based measures are a valuable tool for measuring implicit statistical learning across a number of different populations and highlight the importance of using a combination of both processing- and reflection-based tasks to gain a more detailed insight into the nature of the knowledge acquired through implicit statistical learning.

Acknowledgements

This thesis would not have been possible without a tremendous amount of support from colleagues, friends and family.

I must first thank my small army of supervisors: Dr Ben Wilson, Dr Faye Smith, Dr Nick Riches, Dr Kai Alter and Prof Chris Petkov. Without their unwavering support and guidance, this project would not have been possible. Special thanks must go to Ben, who has dedicated immense amounts of time and effort in his support, and whom I have been indescribably lucky to have as my primary supervisor.

I give my thanks to past and present members of the Laboratory of Comparative Neuropsychology, Dr Yuki Kikuchi, Dr Ross Muers, Lab Manager Jen Nacef, Dr Ryan Calmus, Dr Cody McCants, Dr David Howett, Dr Beshoy Agayby, Ben Slater and Zhipeng Zhang, as well as Ysanne de Graaf, Angelle Antoun, Rohini Murugan and Calen MacDonald, whom it has been a pleasure to work with.

Thanks must go to members of the Auditory Cognition Group, Prof Tim Griffiths, Dr Ester Benzaquén, Kate Yukhnovich and Dr Will Sedley, as well as my panel members Prof Adrian Rees and Dr Sinéad Mullally for their invaluable input to the project. I also thank the ESRC NINE DTP for their funding, and Liam, Caitlin and Robin for their company during conferences.

I must also thank my friends, many of whom have already been mentioned, but also Cat, Janire, Beth, Maria, Rachael, Matthew, Chris, Courtney, Charley, Lydia, Hannah, Millie, Tom and Andy for always lifting my mood.

I would also like to thank my family: my Mum and Dad, and my brother Chrissy for their immeasurable emotional and practical support.

Finally, I give my thanks to the participants who took part in my experiments, as their contributions made this project possible.

Table of Contents

Chapter 1: General Introduction to Implicit Statistical Learning Across Development and Language Ability 1
1.1. Implicit Statistical Learning and Language 2
1.1.1. Defining Implicit Statistical Learning 2
1.1.2. Implicit Statistical Learning of Artificial Grammars 3
1.1.3. Neural Processes Underpinning Implicit Statistical Learning 5
1.1.4. Individual Differences in Implicit Statistical Learning 6
1.2. Implicit Statistical Learning in Adults Without Language Difficulties 8
1.2.1. Traditional Approaches to Measuring Implicit Statistical Learning 8
1.2.2. Reaction Times as a Processing-Based Measure of Implicit Statistical Learning ... 9
1.2.3. Serial Recall as a Processing-Based Measure of Implicit Statistical Learning 10
1.3. Implicit Statistical Learning in Children 11
1.3.1. The Importance of Assessing Implicit Statistical Learning in Children 11
1.3.2. The Developmental Trajectory of Implicit Statistical Learning 11
1.3.3. Processing-Based Measures of Implicit Statistical Learning Across Development 12
1.4. Domain-General Theories of Dyslexia 14
1.4.1. The Phonological Deficit Theory of Dyslexia 14
1.4.2. Implicit Statistical Learning in Dyslexia 14
1.4.3. Processing-Based Measures of Implicit Statistical Learning in Dyslexia 16
Conclusions 16
Chapter 2: Processing-Based Measures of Implicit Statistical Learning 18
Abstract 18
Experiment 2.1: Serial Reaction Times as a Measure of Implicit Statistical Learning 19
Introduction 19
Method 20
Results 27
Discussion 29

Experiments 2.2 – 2.6: Serial Visual Recall as a Measure of Implicit Statistical Learning 31
Experiments 2.2 & 2.3: Visual Artificial Grammar Learning Recall Task 35
Introduction....................35
Methods....................35
Results....................44
Discussion....................49
Experiments 2.4 & 2.5: Chunking Visual Recall Task....................52
Introduction....................52
Methods....................53
Results....................63
Discussion....................71
Experiment 5: Online Hybrid Recall Task....................74
Introduction....................74
Methods....................74
Results....................77
Discussion....................80
General Discussion....................81
Conclusion....................83
Chapter 3: Assessing Implicit Statistical Learning in Children....................84
Abstract....................84
Experiment 3.1: Investigating Implicit Statistical Learning in Children Using Processing-Based Measures....................85
Introduction....................85
Methods....................87
Results....................89
Discussion....................100
Chapter 4: Implicit Statistical Learning Deficits in Dyslexia....................103
Abstract....................103

General Introduction 104
Experiments 4.1 & 4.2: Measuring Implicit Statistical Learning in Dyslexia Using Serial Visual Recall 106
Introduction 106
Methods 108
Results 109
Discussion 118
Experiments 4.3 & 4.4: Is Dyslexia Associated with Language-Specific or Domain-General Deficits? 120
Introduction 120
Methods 121
Results 124
Discussion 131
Conclusions 132
Chapter 5: General Discussion 133
References 141
Appendix 153

List of Figures

Figure 2.1. Experiment 1 stimuli 21
Figure 2.2. Experiment 2.1 trial design 25
Figure 2.3. Experiment 2.1. SRT-AGL Sequence Completion and Grammaticality Judgement task performance 28
Figure 2.4. Experiments 2.2 and 2.3 artificial grammar and exposure stimuli. 37

Figure 2.5. Experiments 2.2 and 2.3 procedure and trial design 40
Figure 2.6. Experiment 2.2 in-person Visual Recall and Grammaticality Judgement task performance 45
Figure 2.7. Experiment 2.3 online Visual Recall and Grammaticality Judgement task performance 46
Figure 2.8. Chunking grammar and stimuli 54
Figure 2.9. Experiments 2.4 and 2.5 trial design 59
Figure 2.10. Experiment 2.4 in-person Visual Recall and Grammaticality Judgement task performance 63
Figure 2.11. Experiment 2.5 online Visual Recall and Grammaticality Judgement task performance 65
Figure 2.12. Experiment 2.6 stimuli and recall sequences 71
Figure 2.13. Experiment 2.6 Hybrid Visual Recall and Grammaticality Judgement task performance 73
Figure 3.1. Children's Visual Recall and Grammaticality Judgement task performance 85
Figure 3.2. Correlations between age and performance in the Visual Recall and reflection-based tasks...88
Figure 3.3. Children and adult's Visual Recall Task and reflection-based task performance .92
Figure 4.1. Visual Recall and Grammaticality Judgement task performance for individuals with dyslexia recruited in-person (Experiment 4.1)...104
Figure 4.2. Visual Recall and Grammaticality Judgement task performance for individuals with reading difficulties recruited online (Experiment 4.2)...106
Figure 4.3. Performance of individuals with and without dyslexia across the in-person Visual Recall and reflection-based tasks (Experiment 4.1) 109
Figure 4.4. Performance of individuals with and without reading difficulties across the online Visual Recall and reflection-based tasks (Experiment 4.2) 110
Figure 4.5. In-person nonword and tone auditory AGL performance 120
Figure 4.6. Online nonword and tone auditory AGL performance..121
Figure 4.7. Correlations between Nonword and Tone AGL task performance and standardised cognitive and language tasks across participants with and without dyslexia 123

List of Tables

Table 2.1. Descriptive statistics for the Visual Recall tasks in Experiments 2.2 and 2.3........45
Table 2.2. Experiment 2.4 and 2.5 Grammaticality Judgement task sequences......................58
Table 2.3. Experiment 2.4 and 2.5 Sequence Completion sequences......................................58
Table 2.4. Descriptive Statistics for the Visual Recall Task in Experiments 2.4 and 2.5. 63
Table 2.5. Descriptive Statitsics for the Visual Recall task in Experiment 2.6. 77

Chapter 1: General Introduction to Implicit Statistical Learning Across Development and Language Ability

Our environment is full of rich statistical information in the sense that some events predict others. For example, rain can typically be predicted by the presence of dark clouds. We detect and extract these regularities through the process of implicit statistical learning, which critically occurs without conscious awareness or any intention to learn. This process of acquiring and processing structure based on distributional cues is central to many aspects of cognition, including musical processing (Tillmann & McAdams, 2004), visual search (Baker et al., 2004) and visuomotor learning (Hunt & Aslin, 2001). However, implicit statistical learning is particularly important for language learning, where knowledge of the underlying grammatical structure is often acquired implicitly, without conscious awareness (Kidd, 2012).

Although implicit statistical learning typically occurs outside of conscious awareness, the tasks that are used to measure this learning often require explicit reflection. However, implicit statistical learning can also be measured using 'processing-based' tasks, which do not require explicit reflection, and instead measure other processes that are facilitated by implicit statistical learning. This thesis aims to develop novel processing-based measures of learning and combine them with traditional reflection-based tasks to assess implicit statistical learning. I also used these tasks to study adults with and without dyslexia, and in children, to better understand the processes that occur during implicit statistical learning in these populations.

This introduction aims to provide an overview of the relevant fields of research, and to address a number of key points. First, implicit statistical learning processes underpin the acquisition and processing of many of the different types of relationship found in language (Section 1.1). Second, novel paradigms may reflect implicit statistical learning more accurately than the tasks that have traditionally been used, which may be affected by conscious processing (Section 1.2). Third, these novel paradigms may help address issues in investigating the developmental trajectory of implicit statistical learning, where findings have been contradictory (Section 1.3). Finally, it is unclear whether individuals with developmental dyslexia show impaired implicit statistical learning (Section 1.4). This thesis aims to develop and test novel measures of implicit statistical learning that better reflect these mechanisms and apply these tasks to investigate the implicit/explicit nature of the knowledge acquired (Chapter 2), the developmental trajectory of implicit statistical learning (Chapter 3), and the proposed deficits in implicit statistical learning in dyslexia (Chapter 4).

1.1. Implicit Statistical Learning and Language

1.1.1. Defining Implicit Statistical Learning

The ability to detect and extract patterns from our environment plays an important role in many aspects of cognition (Batterink et al., 2019). Multiple lines of research have investigated this phenomenon using different terms. For example, "implicit learning" (Reber, 1967) refers to the process by which knowledge about the regularities within the environment is acquired outside of conscious awareness. Similarly, the term "statistical learning" (Saffran et al., 1996) refers to the process of detecting and extracting patterns within our environment and is characterised by the lack of instruction or awareness required for learning to occur, along with an inability to report any of the learned information (Conway, 2020). These different fields appear to be investigating a remarkably similar phenomenon (Christiansen, 2019), despite using different paradigms and approaches (Perruchet & Pacton, 2006). Given the similarities between implicit and statistical learning, it is unsurprising that they are often taken to refer to the same underlying mechanism, which has been termed "implicit statistical learning" (Conway et al., 2010; Conway & Christiansen, 2009; Emberson et al., 2011). Throughout this thesis I will use 'implicit statistical learning' to refer to the ability to extract regularities from the environment through automatic learning mechanisms operating outside of immediate awareness (Cleeremans et al., 1998; Reber, 1967; Saffran et al., 1996).

Implicit statistical learning theories of language acquisition originally emerged as a domain-general opposition to more traditional domain-specific accounts of language acquisition (Chomsky, 1965; Reber, 1967). Implicit learning was commonly viewed as a single cognitive process, thought to underlie the acquisition of regularities of a range of stimuli, across a number of different domains. Therefore, implicit statistical learning was traditionally conceptualised as a unitary, domain general mechanism (Bulf et al., 2011). Indeed, many studies have indicated that implicit statistical learning can underpin the learning of regularities within a range of stimuli, such as linguistic (e.g. Pelucchi et al., 2009) and non-linguistic (e.g. Saffran et al., 1999) stimuli, across a number of domains, including visual (e.g. Fiser & Aslin, 2001), auditory (e.g. Saffran et al., 1996) and motor (e.g. Nissen & Buellmer, 1967) domains. Whilst implicit statistical learning underpins the learning of regularities across a range of stimuli and domains, studies have continually shown that patterns of modality and stimulus specificity that cannot be accounted for by unitary theories. For example, there is often limited transfer of learning across modality and learning in one modality can proceed in the presence of interference in another modality (Conway & Christiansen, 2006). Furthermore, qualitative differences in implicit statistical learning patterns have been reported across modalities (Conway & Christiansen, 2005), and performance in tasks measuring implicit statistical learning across modalities are often uncorrelated (Frost et al., 2015).

To account for these findings, there has been increased interest in the development of multimodal theories of implicit statistical learning (Batterink et al., 2019; Conway, 2020; Frost et al., 2019; Frost et al., 2015). Frost et al. (2015) suggested that implicit statistical learning involves a number of neurobiological mechanisms that detect and encode a range of distributional properties within a given modality or type of input. They argue that these mechanisms are not supported by a single unitary system, but instead reflect separate neural networks in distinct cortical areas. When an internal representation is encoded, the process is constrained by the specific properties of the input and the cortical areas these are processed in. For example, auditory cortex exhibits poorer sensitivity to spatial information, and improved sensitivity to temporal information, meaning that adjusting these parameters will lead to different patterns of learning in the auditory domain compared to the visual domain (Emberson et al., 2011). Whilst modality may affect learning at the point of encoding, and at lower levels of processing, other prefrontal cortical areas are often recruited during implicit statistical learning irrespective of modality, which suggests that beyond the initial stages of processing stimuli, implicit statistical learning may operate as a domain general mechanism. These frameworks highlight the importance of studying implicit statistical learning across a range of different modalities (e.g. auditory, visual) in order to gain a greater understanding of the mechanisms that underlie these processes.

1.1.2. Implicit Statistical Learning of Artificial Grammars

Many of the rules that govern language can be learned through implicit statistical learning. For example, identifying word boundaries within a continuous speech stream involves detecting relationships between syllables. In English, the words ‘*prettydolly’* can easily be separated within a speech stream because the phoneme ‘*pre’* is often followed by ‘*tty’*, but ‘*tty’* is rarely followed by ‘*do’*. The frequency with which pairs of phonemes (A and B) co-occur in a corpus of speech can be calculated, and is typically reported as the Transitional Probability (TP_{AB}). This is calculated as:

$$\text{TP of } A \text{ to } B = \text{frequency of } AB \text{ / frequency of } A$$

In the example ‘*prettydolly’*, the transitional probability of the within-word transition is much higher than the between-word transition. In a seminal study, Saffran et al. (1996) designed an experiment to compare learning of high versus low transitional probabilities and demonstrated that 8-month-old infants can detect word boundaries within a continuous speech stream based only on statistical regularities generated by an artificial grammar. In this experiment, the infants were exposed to a speech stream containing syllables (e.g., ‘*pu*’, ‘*ti*’, ‘*ba*’) arranged into four predictable trisyllabic ‘words’ (i.e., ‘*putiba*’). The syllables within a word always occurred together in the same order, whereas the words themselves could appear within the speech stream in any order. As such, the transitional probabilities within a word were higher than those between words. The ability to detect word boundaries based only on statistical regularities within the input has been replicated consistently in both the auditory (Aslin et al., 1998; Saffran, 2002) and visual domain (Kirkham et al., 2002) using artificial grammars, and also when using natural languages (Pelucchi et al., 2009).

These studies are examples of artificial grammar learning paradigms, which are the most common measure of implicit statistical learning. These tasks typically consist of two parts: an exposure phase and a testing phase. In the exposure phase, the participant is presented with structured sequences of stimuli, typically auditory or visual, and the order of these stimuli is governed by certain rules. These rules mean that the ‘grammatical’ sequences contain statistical regularities and predictable transitions between stimuli, in that some stimuli co-occur more frequently than others. In the exposure phase, participants may be asked to attend to the stimuli, but they are not informed about the presence of rules underlying the sequences. The testing phase consists of a task that measures learning that has occurred during the exposure phase and differs depending on the population being tested. For example, Saffran et al. (1996). used a preferential looking time task, which is suitable for infants. In these tasks, infants are presented either with many grammatical sequences of sounds or continuous streams of phonemes generated by the grammar, until they habituate to these stimuli. They are then presented with both grammatical and ungrammatical sequences. If infants show a dishabituation response, indicated by increased looking times towards ungrammatical sequences, then this is taken as evidence that the learning has occurred. In adults and older children, learning is typically assessed using a grammaticality judgement task, in which participants are presented with sequences of stimuli, some of which are grammatical and follow the same rules as those in the exposure phase and some which violate these rules (‘ungrammatical’ sequences). Participants are asked whether a sequence follows the same patterns as those in the exposure phase, and often asked to respond based on their ‘gut feeling’ (Reber, 1967). Different responses to the grammatical sequences compared to the ungrammatical sequences can be taken as evidence that the participant learned (at least some of) the rules during the initial exposure phase. The inability to report the rules of the grammar in any detail has been taken as further evidence that the learning that has occurred is implicit, rather than explicit (Dienes & Berry, 1997).

The flexibility of artificial grammar learning paradigms make them suitable for investigating how many different aspects of syntax can be learned through implicit statistical learning, as language consists of much more than just simple deterministic adjacent relationships, such as those learned in Saffran et al.'s (1996) study. For example, implicit statistical learning is also thought to support the learning of more probabilistic adjacent relationships, such as those that are found between words. A considerable amount of research has demonstrated that the more variable between word relationships within an artificial grammar can be learned by way of implicit statistical learning (Fiser & Aslin, 2002; Gebhart et al., 2009; Gomez & Gerken, 1999; Reber, 1967; Saffran et al., 2008).

Language also contains more complex relationships between word and phrases, such as nonadjacent relationships, where the regularities between words or phrases are separated by some intervening material. For example, in English, nonadjacent relationships can be found in tense agreement (e.g., *is walking*, vs *has walked*). For these relationships to be learned and processed, the first element must be held in memory until further in the sequence. Previous research has indicated that these more complex relationships can also be learned using similar paradigms as adjacent relationships (Gómez & Maye, 2005; Newport & Aslin, 2004), which suggests that implicit statistical learning can also support the acquisition of more complex relationships. Furthermore, phrases within sentences are often structured hierarchically (e.g., *the dog [the cat chased] barked*). Artificial grammars of the form A^nB^n (which generate sequences such as "A[AB]B", in which one AB phrase is embedded within another) have been used to assess learning of these more complex hierarchical relationships (Bahlmann et al., 2008; Friederici et al., 2006; Friederici et al., 2011; Petersson et al., 2012; Uddén et al., 2009). Due to their complexity, previous research has primarily measured learning in adults, however there is some more recent evidence that suggests that children show learning of these artificial grammars as well (Ojima & Okanoya, 2020; Winkler et al., 2018). These various types of relationship are fundamental to language and can be learned by way of implicit statistical learning, which highlights the importance of these processes in language acquisition and processing.

1.1.3. Neural Processes Underpinning Implicit Statistical Learning

Artificial grammar learning paradigms also tap into language-relevant neural processes. For example, processing syntactic violations in natural language is often associated with a number of neural responses, including early left anterior negativity (ELAN), negativity around 400ms following baseline (N400; Hagoort & van Berkum, 2007; Hagoort, 2008; Hagoort et al., 2004) and positivity around 600ms following baseline (P600; Hahne & Friederici, 2002; Osterhout et al., 1994; Friederici et al, 2002). Such responses are also found in response to syntactic violations during artificial grammar learning paradigms (Abla & Okanoya, 2009; Christiansen et al., 2012; Tabullo et al., 2013).

Evidence from neuroimaging studies have also suggested that artificial grammar learning tasks recruit similar brain regions to natural language processing (Christiansen et al., 2012). Namely, the involvement of left inferior frontal gyrus (BA 44/45, or Broca's area; Friederici, 2011; Hagoort et al., 2004; Karuza et al., 2013; Udden et al., 2008), striatum (Turk-Browne et al., 2009) and medial temporal lobe (Schapiro et al., 2014) has been highlighted in implicit statistical learning of artificial grammars. It is well established that the left inferior frontal gyrus plays an important role in language production (Broca, 1861; Friederici, 2002; Hickok & Poeppel, 2007; Sahin et al., 2009; Vigneau et al., 2006) and comprehension (Bogen & Bogen, 1976; Homae et al., 2002; Zurif, 1980) particularly when processing syntactic complexity (Friederici, 2011). Furthermore, the striatum and medial temporal lobe have also been shown to play a substantial role in language processing (Chan et al., 2013; Meyer et al., 2005).

Similar neural processes underlying artificial grammar learning and natural language learning highlights that the implicit statistical learning processes that are being measured in artificial grammar learning paradigms likely play an important role in language acquisition and processing.

1.1.4. Individual Differences in Implicit Statistical Learning

Beyond the similarities between the neural systems recruited by artificial grammar learning paradigms and natural language processing, further evidence that similar cognitive systems underlie these processes comes from the study of individual differences in language and implicit statistical learning.

Previous research has demonstrated that performance on implicit statistical learning tasks is positively correlated with sentence comprehension tasks (Misyak & Christiansen, 2012), and speech perception abilities, even after controlling for other cognitive abilities such as memory and IQ (Conway et al., 2010). These relationships between implicit statistical learning and language ability are also relevant for more complex nonadjacent relationships: performance in implicit statistical learning tasks containing nonadjacent dependencies predicted individual differences in processing sentences containing long-distance dependencies (Conway et al., 2010). Conway and colleagues showed a positive correlation between visual implicit statistical learning and auditory sentence processing, and also between auditory implicit statistical learning and audio-visual sentence processing, demonstrating that across domains, implicit statistical learning is associated with language ability. Such correlations between implicit statistical learning and language ability have also been demonstrated in children aged between 4 and 14 years (Evans et al., 2009; Kidd, 2012; Kidd & Arciuli, 2016). Furthermore, performance on implicit statistical learning tasks has also been shown to relate to reading in both children and adults (Arciuli & Simpson, 2012; von Koss Torkildsen et al., 2019). Based on these findings, implicit statistical learning plays an important role in language acquisition and processing.

However, studies assessing how individual differences in implicit statistical learning relate to language ability have been criticised for their lack of replicability. A number of recent studies have found no correlation between implicit statistical learning and language ability in children and adults (Schmalz et al., 2019; West et al., 2018), whereas others have only reported small correlations (Spencer et al., 2015). Other studies have also found correlations only when using auditory implicit statistical learning tasks (Qi et al., 2019). Siegelman (2020) suggested that poor reliability across measures of implicit statistical learning (Erickson et al., 2016) may account for the lack of replicability of the relationship between individual differences in implicit statistical learning and language ability. There is a clear need for the refinement in the current tasks used to measure implicit statistical learning, with a particular focus on developing tasks that accurately capture variability in these abilities.

As well as examining individual differences within the general population, the link between implicit statistical learning and language can also be investigated by comparing implicit statistical learning abilities between individuals with and without language disorders. Implicit statistical learning deficits have been found across a number of developmental language disorders, as well as acquired language difficulties. For example, children with Specific Language Impairment (SLI) have been found to perform more poorly in artificial grammar learning tasks (e.g. Evans et al., 2009) and other measures of implicit statistical learning (e.g. Hsu & Bishop, 2014) compared to children without language difficulties. Indeed, a meta-analysis of studies investigating implicit statistical learning deficits in SLI reported a substantial difference in the ability to detect statistical regularities associated with SLI (Lammertink et al., 2017). Whilst the evidence of an implicit statistical learning deficit in dyslexia is less clear (see Section 1.4.2 for a detailed overview of this literature), several studies have reported implicit statistical learning deficits in dyslexic children (e.g., Pavlidou et al., 2009) and adults (Katan et al., 2017). Finally, evidence of implicit statistical learning deficits has been found in individuals with acquired language difficulties, for example agrammatic aphasia (Christiansen et al., 2010).

Taken together, these separate lines of evidence investigating the neural basis of implicit statistical learning, individual differences in these abilities and potential deficits in these abilities in language disorder provide strong evidence of a link between implicit statistical learning and language.

1.2. Implicit Statistical Learning in Adults Without Language Difficulties

1.2.1. Traditional Approaches to Measuring Implicit Statistical Learning

During language acquisition, many of the rules underlying language are learned implicitly (Arciuli, 2018; Arciuli & Simpson, 2012; Arnon, 2019; Dienes, 2012). Indeed, Dienes (2012) highlights that the core grammatical rules that underlie our native language are typically learned at a young age, before we are even aware of the existence of a grammar. Therefore the original artificial grammar learning paradigms pioneered by Reber (1967, 1976) were also designed to elicit implicit learning. According to Dienes and Berry's (1997) framework, knowledge can be considered implicit if participants do not acquire conscious awareness of the acquired knowledge, either because their performance in a grammaticality judgement task is above chance despite their belief they are guessing, or because their confidence in the accuracy of their response is unrelated to their performance. Many studies have used these criteria to demonstrate that the learning that results from artificial grammar learning paradigms is implicit (Batterink, Reber, Neville, et al., 2015; Curran, 1997; Dienes & Altmann, 1997; Reber & Squire, 1994; Scott & Dienes, 2008; Tunney & Altmann, 2001; Willingham & Goedert-Eschmann, 1999). Studies measuring artificial grammar learning in amnesic patients provide further support for the learning in these tasks resulting in implicit knowledge: amnesic patients perform above chance in grammaticality judgement tasks despite declarative memory impairments (Knowlton et al., 1992; Knowlton & Squire, 1994).

However, if the aim of artificial grammar learning paradigms is to assess knowledge that is outside of conscious awareness, then the grammaticality judgement tasks that are typically used to measure learning in these paradigms may not be appropriate. These 'reflection-based' tasks rely on making explicit decisions that require conscious reflection about what has been learned. This presents some issues, as these tasks can only measure learning that an individual can explicitly access, and therefore performance on these tasks may not reflect the implicit knowledge that has been acquired, instead reflecting more explicit, decision-making processes (Christiansen, 2019). Furthermore, these reflection-based tasks are often limited in that they only assess the final outcomes of the complex chain of cognitive processes underlying task performance and provide little information about *how* learning is occurring during the exposure phase.

Alternatively, more implicit, 'processing-based' tasks can be used that do not require explicit decision making and are therefore less confounded by these other cognitive processes. These processing-based tasks often rely on less direct methods of assessing learning, by measuring other variables that are facilitated by implicit statistical learning. These tasks also provide additional benefits over reflection-based tasks in that they can measure learning over the course of the task, as responses are required for every trial, even while learning is taking place. This is in contrast to reflection-based tasks, which involve passive exposure where no data are collected. There are a number of processes that are facilitated through implicit statistical learning that can be measured, which will be discussed in turn in the next section.

1.2.2. Reaction Times as a Processing-Based Measure of Implicit Statistical Learning

One of the reasons that it is adaptive to learn statistical regularities in our environment is because it allows us to better predict upcoming events. Serial Reaction Time (SRT) tasks have exploited this facet of implicit statistical learning and showed that when participants are asked to make sequences of responses (e.g., touch stimuli in different spatial locations), they become faster at producing sequences that contain predictable relationships than randomly ordered sequences (Nissen & Bullemer, 1987; Reber & Squire, 1994; Thomas et al., 2004). In these tasks, faster reaction times demonstrate implicit learning, in this case of sequences of motor action.

However, the learning observed in traditional SRT tasks may be somewhat different to the implicit statistical learning found in artificial grammar learning tasks, in which participants must learn dependencies between sensory stimuli rather than between spatial locations. These tasks are thought to rely on a different cognitive and neural mechanisms than artificial grammar learning tasks (Conway, 2005). However, these SRT tasks have been combined with artificial grammar learning tasks to show that reaction time benefits of implicit statistical learning that are not linked to specific motor responses (Misyak et al., 2009, 2010; Vuong et al., 2011). This removes the reliance on motor learning of action sequences and requires that participants learn the relationships between the nonsense words themselves, rather that specific spatial locations. In Chapter 2, Experiment 2.1, I developed a visual SRT-AGL task along with a number of other reflection-based measures, to assess whether reaction times can be used to measure implicit statistical learning, and if processing- and reflection-based tasks measure similar processes. However, I found no evidence of learning across the processing- and reflection-based tasks, which suggests that other processes that are facilitated by implicit statistical learning, such as recall, may provide a more suitable measure of learning.

1.2.3. Serial Recall as a Processing-Based Measure of Implicit Statistical Learning

In addition to enabling faster responses to predictable events, implicit statistical learning can facilitate the rapid processing of large amounts of incoming information (Christiansen & Chater, 2016). Working memory, particularly for rapidly serially presented information is very limited (Cowan, 2010; Miller, 1956; Vogel et al., 2001). Therefore, when presented with even modest sequences of stimuli the limits of working memory can quickly be exceeded. This cognitive limitation can be overcome by combining frequently co-occurring stimuli into larger "chunks", to reduce working memory demands. This process of chunking incoming information is critical for language learning and processing, to allow sufficiently rapid processing of the large amount of information contained within human language (Christiansen & Chater, 2016). This process of 'chunking' frequently co-occurring information is the mechanism by which the infants in Saffran et al.'s (1996) experiment learned to detect the boundaries between 'words' in a stream of syllables (see Section 1.1.2). Beyond the initial acquisition of language, this process is used throughout life in language *processing*. For example, it allows the processing of rapid speech, as chunks can quickly be formed and passed to higher levels of linguistic representation (from phonemes to words, then to phrases, and finally to sentences): when processing speech we do not have to store every phoneme that is processed in order to recall the semantic meaning of the sentence.

Based on this reasoning, if participants are presented with sequences which contain predictable regularities (e.g., certain stimuli always occur in a fixed order), frequently co-occurring stimuli should be more easily combined into chunks. This in turn should reduce the demands on working memory, and allow for more efficient chunk-and-pass processing (Christiansen & Chater, 2016). We can use this aspect of cognition to our advantage to measure implicit statistical learning: artificial grammars can be used to generate predictable and unpredictable sequences, and because predictable sequences contain more frequently co-occurring elements, these sequences may be more easily chunked and recalled than unpredictable sequences. I conducted a number of experiments using this reasoning to assess the extent to which serial visual recall can be used as a processing-based measure of implicit statistical learning (Experiments 2.2 – 2.6). From these experiments, I have demonstrated that serial visual recall can successful measure implicit statistical learning of both highly predictable and more variable relationships, and that the complexity of the grammar being learned may affect the extent to which explicit processes are recruited during reflection-based tasks.

1.3. Implicit Statistical Learning in Children

1.3.1. The Importance of Assessing Implicit Statistical Learning in Children

Although implicit statistical learning is thought to play an important role in language acquisition, the majority of research in this area has focused on the learning of infants (e.g. Saffran et al., 1996; Gomez & Gerken, 1999, 2000) and adults (Fiser & Aslin, 2002; Reber, 1967; Saffran et al., 1999), rather than older children. Language learning occurs beyond infancy, and examining implicit statistical learning abilities in older children as well as infants and adults is critical to understand the developmental trajectory of these abilities and how they may relate to language acquisition and processing (Arciuli & Conway, 2018). Furthermore, examining implicit statistical learning abilities in children older than infants could provide support for the hypothesis that these abilities play a causal role in language development (Arciuli & Torkildsen, 2012; Conway et al., 2010). Finally, assessing implicit statistical learning abilities in children is important for understanding the nature of potential deficits in implicit statistical learning in developmental language disorders such as dyslexia, which are typically identified during childhood (Deocampo et al., 2018; Saffran, 2018).

1.3.2. The Developmental Trajectory of Implicit Statistical Learning

Previous research has suggested that the ability to acquire certain aspects of language declines with age, with older individuals being less adept in reaching fluency than younger learners (Johnson & Newport, 1989; although see Newport et al., 2001 for a discussion). There is relatively little research focusing on the developmental trajectory of implicit statistical learning abilities, despite the importance of these questions in understanding the mechanisms underlying implicit statistical learning and its role in language acquisition.

There are several possible predictions about how implicit statistical learning abilities may change with age. First, it has been suggested that implicit statistical learning abilities are age-invariant, with some studies showing that performance on measures of implicit statistical learning were unaffected by age (Jost et al., 2015; Raviv & Arnon, 2018; Saffran et al., 1997; Thiessen et al., 2013). Second, like many other cognitive abilities, implicit statistical learning many improve with age. Furthermore, as implicit statistical learning is a domain-general mechanism, it may be supported by other cognitive abilities such as memory, which are known to improve with age (Gathercole et al., 2004). There is some evidence for an improvement in implicit statistical learning across development (Arciuli & Simpson, 2011; Kirkham et al., 2002; Kirkham et al., 2007; Lukács & Kemény, 2015; Thomas et al., 2004; Vaidya et al., 2007). There is also some suggestion that the developmental trajectory of implicit statistical learning is affected by modality, with improvements in visual but not auditory implicit statistical learning being found across age (Raviv & Arnon, 2017). Third, given the importance of implicit statistical learning in language acquisition, it has been argued that infants and children show improved implicit statistical learning compared to adults. Infancy and early childhood are where the majority of language learning typically takes place, and as previously mentioned, there is some suggestion that language learning abilities decline with age (Birdsong, 1999; Newport, 1990). The largest study investigating age-related changes in implicit statistical learning found no differences in such abilities in children aged between 4 and 12 years, however following this there was a decrease in implicit statistical learning abilities which remained constant until aged 60, where there was another decrease (Janacsek et al., 2012). These conflicting findings suggest that we do not have a clear understanding of the developmental trajectory of implicit statistical learning. More research is required in populations without language difficulties to better understand implicit statistical learning as a concept, but also to provide a benchmark to distinguish between deficits and delays in children with language disorders.

1.3.3. Processing-Based Measures of Implicit Statistical Learning Across Development

The few studies assessing the developmental trajectory of implicit statistical learning have typically compared performance on grammaticality judgement tasks across a range of ages. However, as discussed in Section 1.2.1, these traditional tasks require additional abilities such as understanding the task instructions, or decision-making skills, which are unlikely to be as well developed in children as they are in adults (Lammertink et al., 2019). Therefore, when comparing performance on these tasks across children and adults, it is possible that we underestimate the implicit statistical learning abilities of the children. Processing-based measures of learning would be particularly suitable for comparing the developmental trajectory of implicit statistical learning abilities, as both adults and older children could complete the same tasks that would not rely on explicit decision making.

These issues in avoiding explicit judgements when assessing implicit statistical learning have previously been navigated within the literature relating to infant implicit statistical learning. As infants do not have the capacity to provide explicit judgements on their implicitly learned knowledge, other methods have been used to assess learning. The seminal Saffran et al. (1996) study used a preferential-looking paradigm to show the learning of word boundaries in infants, and indeed other studies have demonstrated learning in infants using similar paradigms (for a review, see Saffran & Kirkham, 2018). However, such paradigms are not effective in measuring learning beyond infancy (Wilson et al., 2015), and therefore these tasks would not be suitable for assessing the developmental trajectory of implicit statistical learning. Although neuroimaging methods, such as electroencephalography (EEG), can be used to compare learning across age, these experiments are considerably more time-consuming than typical behavioural experiments. Therefore, behavioural processing-based measures may offer some benefits over these neuroimaging methods in terms of their ease of completion.

Based on the current literature, the developmental trajectory of implicit statistical learning is unclear. However, these studies do suggest that stimulus modality and the tasks used, among other factors, may play an important role in determining performance across age (Conway et al., 2010). Current tasks may be underestimating children's performance relative to adults, as traditional reflection-based grammaticality judgement tasks rely on other cognitive abilities that may be less developed in childhood (Anderson, 2002; Gathercole et al., 2004). It is crucial to examine developmental trajectory of implicit statistical learning further using more intuitive processing-based measures, not only to gain a better understanding of implicit statistical learning as a phenomenon, but also to provide benchmarks for investigating implicit statistical learning deficits in developmental disorders such as dyslexia. To address these issues, in Chapter 3, I conducted an online study with children aged between 8 and 15 years which aimed to measure implicit statistical learning using the processing-based Visual Recall task I had previously developed and tested with adults (Experiments 2.4 and 2.5). There was no difference in performance across the sample of children. Although there were similarities between the performance of children and adults, there was some evidence that the time-course of learning differed between age groups. However, there was some suggestion that this may have been due to differences in attention and/or motivation between children and adults.

1.4. Domain-General Theories of Dyslexia

1.4.1. The Phonological Deficit Theory of Dyslexia

Developmental dyslexia is the most common specific learning disability (Roongpraiwan et al., 2002), characterised by deficits in learning to read and spell (Lyon et al., 2003). Given that the issues with dyslexia primarily relate to difficulties in reading and spelling, the most prominent theory of dyslexia relates to a deficit in phonological processing (Snowling, 1998). Phonological processing involves accessing and manipulating phonemes, and is critical for word identification, and therefore reading (e.g. Vellutino et al., 1996). More specifically, studies investigating the phonological deficit suggest that the deficits associated with dyslexia fit into three key facets: poor phonological awareness, verbal short-term memory and slow lexical access (for a review, see Vellutino et al., 2004). Much of the research has demonstrated impaired performance in dyslexia across tasks that measure these abilities. For example, individuals with dyslexia show poorer performance in tasks that require the rapid and/or precise manipulation of phonemes, such as spoonerism and phoneme deletion tasks (Farquharson et al., 2014; Rispens & Been, 2007), rapid automatized naming tasks (Denckla & Rudel, 1976; Jones et al., 2009) and nonword repetition tasks (Melby-Lervåg et al., 2012). These findings are also reflected in the data from neuroimaging studies, which suggest that individuals with dyslexia show different patterns of activation during phonological processing tasks than controls (see Demb et al., (1999) , for a review).

However, there are a wider constellation of sub-clinical differences associated with dyslexia that cannot be explained by the phonological deficit theory, for example differences in motor skills (Fawcett & Nicolson, 1995), auditory processing (Tallal, 1984; see Farmer & Klein, 1995 for a review), visual processing (Stein, 2001, 2019), and implicit statistical learning (Folia et al., 2008; Gombert, 2003; Menghini et al., 2006; Ullman & Pierpont, 2005). As these differences cannot be explained by an impairment in phonological processing specifically, alternative theories have been proposed which highlight more domain-general deficit associated with dyslexia.

1.4.2. Implicit Statistical Learning in Dyslexia

Given the importance of implicit statistical learning in language acquisition and processing, there has been a considerable amount of research investigating potential implicit statistical learning deficits in dyslexia using a variety of tasks. Thus far, the findings have been mixed (for a review, see Schmalz et al., 2017). Some studies have shown poorer performance for children with dyslexia in artificial grammar learning tasks (Pavlidou et al., 2009), and there is some evidence that adults with dyslexia are impaired in artificial grammar learning tasks, but only those which require the learning of more complex grammars (Katan et al., 2017). These findings have not been consistently replicated across the literature: Nigro et al. (2016) found no evidence of a statistical learning deficit in children, although there may have been some evidence that the children with dyslexia had difficulties in generalising the rules to novel stimuli. Inácio et al. (2018) also found no difference between children with and without dyslexia using a grammaticality judgement task, providing further evidence against a statistical learning deficit in dyslexia. Overall, there is mixed evidence for differences in implicit statistical learning in dyslexia when using artificial grammar learning paradigms.

Although fewer studies have investigated statistical learning deficits in dyslexia using SRT tasks, the resulting findings are as inconsistent as those from artificial grammar learning studies. Support for a statistical learning deficit in dyslexia using SRT tasks is typically evidenced by faster reaction times for predictable sequences over random sequences in individuals without dyslexia, but less so in individuals with dyslexia. This effect has been demonstrated in both adults (Stoodley & Stein, 2006) and children (Stoodley et al., 2008). However, these findings are not unanimous: other studies have found no group differences between dyslexic and control participants (Menghini et al., 2010). In fact, the largest SRT study investigating statistical learning in dyslexia recruited over 400 children, and found no group differences in statistical learning, although children with dyslexia had slower RTs overall (Waber et al., 2003). It was suggested that these differences in speed could be accounted for by attentional difficulties, and indeed once these difficulties had been accounted for, the differences disappeared.

A meta-analysis of studies investigating artificial grammar learning in dyslexia suggested that dyslexic individuals showed impairments in implicit statistical learning, and that these deficits may be more pronounced in children compared to adults (van Witteloostuijn et al., 2017). However, the authors also highlighted that there was evidence of publication bias within the literature, and that unpublished data may null the effect. In a systematic review on implicit statistical learning deficits in dyslexia, Schmalz et al. (2017) highlighted that drawing conclusions about the nature of implicit statistical learning deficits in dyslexia is difficult due to the lack of high-quality data. Indeed, given the conflicting findings within the literature, it is clear that more research is required to understand any differences in implicit statistical learning in individuals with dyslexia.

1.4.3. Processing-Based Measures of Implicit Statistical Learning in Dyslexia

Although previous research has provided invaluable insights into the conditions in which people with dyslexia may show differences in implicit statistical learning, there remains little clarity within the literature relating to the nature of this proposed deficit. One of the reasons for the mixed findings throughout the literature may be due to the tasks that are typically used to measure performance. Previous research has highlighted that tasks measuring implicit statistical learning are poorly defined, and often do not correlate well with one another (Schmalz et al., 2019). These reflection-based measures may be less suitable for addressing the question of whether there is a specific deficit in implicit statistical learning dyslexia, as these tasks are likely measuring explicit decision-making processes as well as implicitly acquired knowledge. Instead, one avenue that is worth exploring is whether processing-based measures, which do not require explicit decision-making, can offer any additional insights into any differences in implicit statistical learning in individuals with and without dyslexia. Whilst SRT tasks are often used to assess implicit statistical learning deficits in dyslexia (see Section 1.4.2), these tasks rely on motor responses, in which individuals with dyslexia may also show differences to individuals without dyslexia (e.g., Fawcett & Nicolson, 1995). In Chapter 4, I assess implicit statistical learning in adults with dyslexia using a range of processing- and reflection-based tasks and compare performance to adults without dyslexia. Across these experiments, there was little evidence of differences in implicit statistical learning between participants with and without dyslexia based on performance across both processing-based and reflection-based tasks, which provides some evidence against differences in implicit statistical learning in dyslexia.

Conclusions

In this introduction I have provided an overview of implicit statistical learning and its relevance in language and language disorders, whilst highlighting the need for more appropriate processing-based measures of learning. Such tasks do not require conscious reflection (e.g., decision-making processes) and therefore are likely to provide a more accurate measure of knowledge that has been acquired through implicit statistical learning. Such measures are particularly relevant when measuring implicit statistical learning across development and in individuals with language difficulties, such as dyslexia. In Chapter 2, I developed and tested novel processing-based measures of learning using reaction times (Experiment 2.1) and visual serial recall in both in-person (Experiments 2.2 and 2.4) and online (Experiments 2.3, 2.5 and 2.6) samples of adults without language difficulties. In Chapter 3, I examine the developmental trajectory of implicit statistical learning across children aged 8 to 15 years using visual serial recall (Experiment 3.1). In Chapter 4, I investigate implicit statistical learning deficits in dyslexia using the visual serial recall task (Experiment 4.1 and 4.2) and directly compare language-specific and domain-general theories of dyslexia using Nonword and tone artificial grammar learning tasks (Experiments 4.3 and 4.4).

Chapter 2: Processing-Based Measures of Implicit Statistical Learning

Abstract

A defining feature of implicit statistical learning is that it occurs without conscious awareness; however, implicit statistical learning is often assessed using measures that require explicit reflection. 'Processing-based' tasks can measure learning without requiring conscious reflection, by measuring processes that are facilitated by implicit statistical learning. In Chapter 2, I aimed to combine a novel processing-based measure of implicit statistical learning with traditional reflection-based tasks to gain further insight into the processes that occur during implicit statistical learning. We first assessed the efficacy of reaction times as a measure of implicit statistical learning, by leveraging the fact that participants should respond more quickly to predictable compared to unpredictable stimuli. However, we found no evidence of learning using the serial reaction time artificial grammar learning task (Experiment 2.1). Therefore, we shifted focus towards the development of a serial visual recall task. We conducted a series of 5 serial visual recall experiments, based on the premise that frequently co-occurring stimuli may be "chunked" into a single cognitive unit, reducing working memory demands and facilitating recall. In these experiments, we predicted that when participants were asked to remember and recreate sequences of serially presented images, they would show improved recall for grammatical sequences, which can be chunked, over ungrammatical sequences that cannot. In Experiments 2.2 and 2.3 using methods based on previous visual artificial grammar learning paradigms we saw no evidence of learning based on serial recall but did replicate prior learning effects using the other reflection-based measures. In Experiments 2.4 and 2.5, we adapted the method away from a traditional exposure-test paradigm to use a blocked design, as well as developing a novel artificial grammar more susceptible to chunking. In these experiments, strong learning effects were observed across tasks. To assess whether this learning was due to the changes in procedure or the new grammar, we conducted one further study using the new methods and the original artificial grammar. Experiment 2.6 also showed substantial learning effects in both the serial recall task and subsequent more reflection-based measures of learning. These data demonstrate that serial recall is a valuable approach to measure implicit statistical learning and highlights some conditions under which this approach is and is not successful.

Experiment 2.1: Serial Reaction Times as a Measure of Implicit Statistical Learning

Introduction

Implicit statistical learning is typically measured using artificial grammar learning paradigms in which participants are exposed to grammatical sequences, and then tested on their learning of the rules underlying these sequences using a grammaticality judgement task. However, as previously discussed, these reflection-based tasks may not accurately reflect implicit statistical learning and may instead be a reflection of the more explicit decision-making processes that are required to make responses. Processing-based measures, which do not rely on explicit-decision-making, typically measure other variables that are facilitated by implicit statistical learning. For example, serial reaction time tasks are often used to demonstrate implicit learning of motor sequences, as participants typically become faster at producing predictable sequences over randomly ordered sequences. (Nissen & Bullemer, 1987; Reber & Squire, 1994; Thomas et al., 2004).

However, the learning observed in traditional SRT tasks may be somewhat different to the implicit statistical learning found in artificial grammar learning tasks, in which participants must learn dependencies between sensory stimuli rather than between spatial locations. These tasks are thought to rely on a different cognitive and neural mechanisms than artificial grammar learning tasks (Conway, 2005). However, these SRT tasks have been combined with artificial grammar learning tasks to create “SRT-AGL” tasks that have shown that the reaction time benefits of implicit statistical learning are not necessarily linked to specific motor responses. For example, Misyak et al. (2009) presented participants with a visual array containing two rows of three nonsense word stimuli (see Figure. 2.1 for a similar design). Participants were then presented with a sequence of three auditory nonsense words, each corresponding to one of the nonsense words on screen and asked to click on the matching nonsense word stimulus as quickly as possible. Unbeknownst to the participants, the first nonsense word of the sequence always predicted the final nonsense word, by way of a nonadjacent dependency. Over the course of the experiment participants got faster at responding to this final predictable nonsense word, and reaction times were slower in a Testing Block when they were presented with randomised sequences containing no predictable dependencies (Misyak et al., 2009). Importantly, in this task although the nonsense word sequences always unfolded from left to right across the screen, the vertical position of the stimuli was randomised, and was not predictable based on the previous nonsense words. This removes the reliance on motor learning of action sequences and requires that the participants learn relationships between the nonsense words themselves.

Although beyond the scope of the current experiments, we aimed to design an SRT-AGL task that could in future be used to measure implicit statistical learning in individuals with language difficulties, such as dyslexia, where visually presented nonsense words might not be appropriate (Catts et al., 2005). Moreover, we combined this processing-based task with a range of other reflection-based measures of learning to assess whether these tasks capture different facets of learning. If SRT-AGL task is an effective measure of implicit statistical learning, then we would predict that participants would show faster reaction times to the predictable sequences than to unpredictable sequences. We would also predict that participants would perform above chance in the subsequent reflection-based tasks. If processing-based measures of learning measure different processes than reflection-based measures, then we predict no correlation between performance on the SRT-AGL task and performance across the subsequent reflection-based tasks.

Method

Participants

32 participants (23 female, 9 male; mean age: 21.86) were recruited using both the School of Psychology and Institute of Neuroscience participant pools at Newcastle University. This sample size was similar to the 30 participants recruited by Misyak et al. (2009), which this study was based on. 17 participants completed the adjacent version of the task, and 15 participants completed the nonadjacent version. All participants were native English speakers. Participants were not excluded based on their ability to speak any additional languages.

Stimuli

We assess learning of both adjacent and nonadjacent dependencies (Gomez, 2002; Misyak et al., 2009, 2010). In the nonadjacent task, three elements of the form 'AXB' were generated, where the initial 'A' element (e.g., 'A_1') predicts the final 'B' element ('B_1') forming a nonadjacent dependency ('A_1XB_1') while the intervening 'X' element is not dependent on either the 'A' or the 'B' stimuli. In the adjacent task, we also used a similar grammar of the form 'XAB', where the 'A' elements still predicted the 'B' elements, but these items occurred adjacent to one another in a sequence, rather than being separated by the intervening 'X' element.

We used 28 abstract shapes (2 'A', 2 'B' and 24 'X' stimuli) based on shapes from Fiser and Aslin (2001), shown in Figure 2.1.

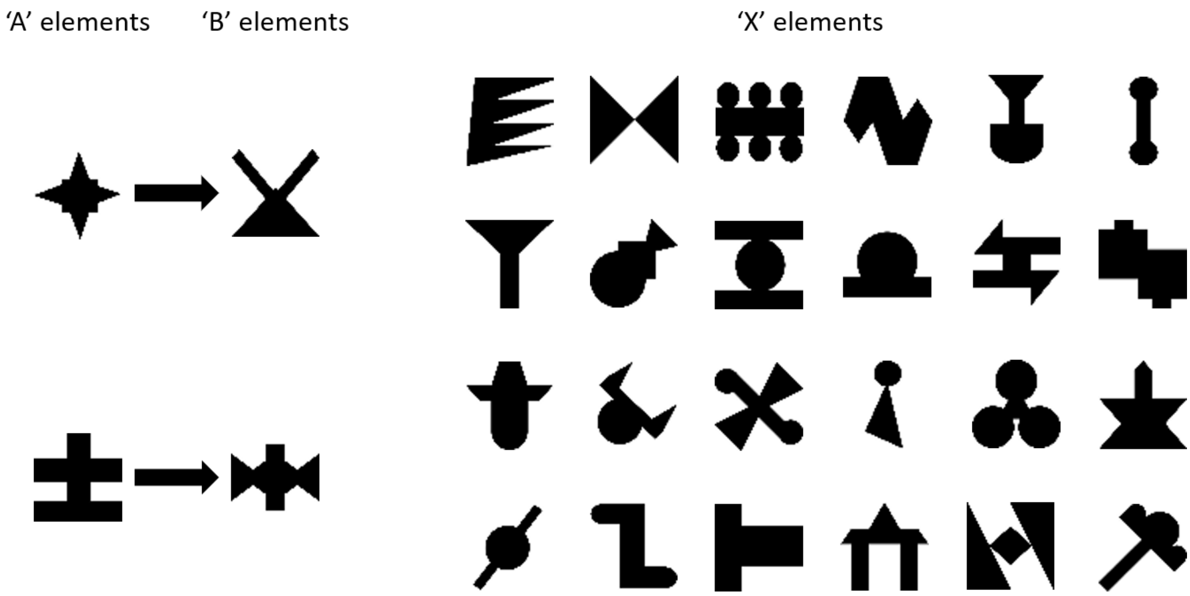


Figure 2.1. Experiment 1 stimuli. The 'A' and 'B' elements always co-occurred, and sequences are presented in the form 'AXB' in the nonadjacent version of the task, and 'XAB' in the adjacent version. There are 2 'A' and 'B' elements, and 24 'X' elements. 24 sequences were presented per block, with each 'X' element presented once per Learning Block. 20 of the sequences in each block were grammatical (e.g., A_1XB_1), and for the 4 ungrammatical sequences, the dependencies between the 'A' and 'B' elements were swapped around (e.g., A_1XB_2; A_2XB_1).

Procedure

This experiment was based on previous Serial Reaction Time-Artificial Grammar Learning (SRT-AGL) experiments (Misyak et al., 2009, 2010; Vuong et al., 2011). The experiment took place in a testing lab within the Institute of Neuroscience at Newcastle University and was coded using MATLAB and Psychtoolbox. Participants were seated approximately 60cm in front of a computer monitor (24-inch Dell U2412M, screen resolution 1920*1200 pixels). Responses were made either with the mouse (in the SRT-AGL and Sequence Completion tasks) or by pressing one of two keys on the keyboard (in the Grammaticality Judgement tasks, see below).

To assess the relationship between implicit statistical learning and language abilities, in Experiment 2.1 we conducted a battery of standardised cognitive and language tests before and after the main experiment. Before the experiment, the TOWRE Words and Nonwords, Backwards Digit Recall and the Recalling Sentences Task were administered (see Appendix 2.1.). After completing the SRT-AGL, Grammaticality Judgement, Sequence Completion and Sequence Generation tasks, the WASI Block Design, Rapid Automatized Naming Digits and Objects tasks were administered.

First, participants completed the SRT-AGL task, followed by the Sequence Completion, Grammaticality Judgement and Sequence Generation tasks.

SRT-AGL task

In the visual SRT-AGL task, participants completed 8 blocks of 24 trials, consisting of 20 grammatical and 4 ungrammatical sequences. In each trial, 6 abstract shapes were presented on the screen (Figure 2.2). For each of the three horizontal positions in the sequence in turn, a cue stimulus matching one of the target shapes appeared between the targets after a 500ms delay (Figure 2.2). Participants were told to click on the matching shape as quickly as possible. The cue remained on the screen until the participant had selected a shape. Feedback on the speed of participant’s responses was given after every trial in the form of happy or unhappy ‘smiley faces’. Participants were provided with the opportunity to take a break half-way through the experiment between blocks 5 and 6.

In the SRT-AGL task the final element of the sequence was predictable based on the previous elements in the sequence (see Experiment 2.1. *Stimuli*). Therefore, if participants had learned the dependency, they should be able to respond more quickly to this predictable element than the preceding, unpredictable stimuli (Misyak et al., 2009, 2010).We calculated the decrease in reaction time to these final elements by subtracting the reaction time in response to the first element (RT_A) from the final element (RT_B). The difference between these reaction times (RT_{A-B}) gives a measure of the speed increase provided by the predictable stimulus. Participants were also presented with ungrammatical sequences in which the final 'B' element in the sequence did not correspond to the 'A' element (e.g., '$A_1X\underline{\mathbf{B_2}}$'). In this experiment we used an oddball design, in which the ungrammatical sequences were infrequently presented throughout the experiment. We predicted that if participants had learned the dependencies (e.g., $A_1 \rightarrow B_1$), they should implicitly anticipate the predictable B stimulus and therefore when they are asked to click on an unexpected element (e.g., B_2) they should show increased reaction times.

In all trials, if participants made an error the trial was omitted from the reaction time analysis (<4% of trials contained an incorrect response). We predicted that participants may also make more errors in their selection of the final 'B' elements for the ungrammatical sequences than for the grammatical sequences, although analysis of the data from this task showed that errors were minimal. To remove outliers, we also omitted trials with reaction times that exceeded the mean reaction time + 3SDs.

Sequence Completion task

The SRT-AGL task was immediately followed by the Sequence Completion task. Each trial in this task began identically to the SRT-AGL task: the participants were presented with an array of 6 visual stimuli, after which the first two cue stimuli were presented and participants responded by clicking on the matching stimuli, as before. However, instead of presenting the final cue stimulus (corresponding to the final element in the sequence), participants were presented with a question mark in place of the final cue stimulus. They were instructed to guess which of the two possible final stimuli they thought completed the sequence. Participants completed 24 trials: each of the 'X' elements was presented once, and half of the sequences contained A_1 (with the correct response being B_1) and the other half of the sequences containing A_2 (with the correct response being B_2).

We predicted that if the participants had learned the sequence dependencies in the earlier part of the task, they should be more likely to select the correct stimulus. As the participants were asked to make an explicit decision in this task, this may rely on more explicit processes, similar to grammaticality judgement tasks. However, unlike the Grammaticality Judgement task, the participants were not informed that there were rules underlying the sequences prior to this task, and therefore this might offer a more implicit alternative to the Grammaticality Judgement task.

Grammaticality Judgement task

In this task, participants were presented with a sequence of three stimuli and asked whether or not this sequence ‘followed the same pattern’ as the sequences they had heard or seen previously, by pressing one of two keys on the keyboard. They were told that if they were not certain they should respond based on their gut feeling. This approach is similar to many other artificial grammar learning studies, particularly testing adult participants. The Grammaticality Judgement task consisted of 24 trials in total, half of which were grammatical and half ungrammatical.

Sequence Generation task

In this task participants were instructed to create their own 3 element long sequences. In each of 24 trials, the participants were presented with 8 elements arranged in a circle on the screen: the 2 ‘A’ and ‘B’ elements were always shown, as well as 4 randomly selected, non-repeating ‘X’ elements. Participants created their sequences by clicking on stimuli in order. Each trial was separated by a 500ms inter-trial interval. No feedback was given.

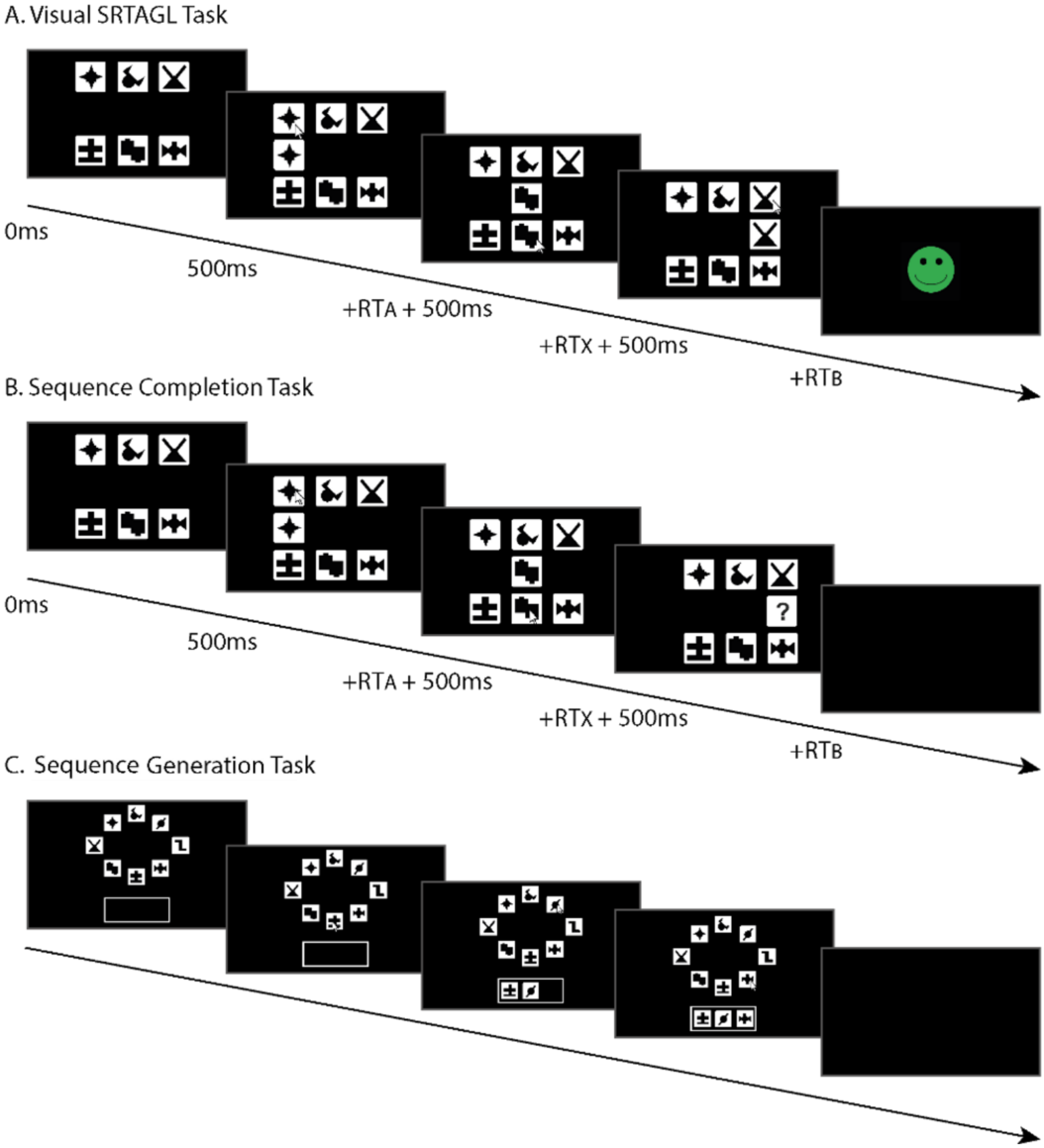


Figure 2.2. Experiment 2.1 trial design. A) Visual SRT-AGL task. On each trial, participants were presented with an array of 6 visual abstract shapes. Every trial contained 2 'A' elements in the left column, 2 randomly selected 'X' elements in the middle column, and 2 'B' items in the right column. The vertical position of the stimuli (upper or lower) was pseudo-randomised so that all items occurred equally frequently in the upper and lower positions, and so that the matching 'A' and 'B' elements occurred in the same row 50% of the time, so the correct responses could not be predicted based on the position of the stimuli. 500ms after the visual stimuli appeared the participants were presented with a visual cue corresponding to one of the two target stimuli in the left column (corresponding to an 'A' element), and the participant clicked on the matching target stimulus as quickly as possible. After a 500ms delay, they were presented with the second auditory cue stimulus ('X' element) and clicked on the matching stimulus in the middle column, then the final cue stimulus corresponding to the predictable 'B' element in the right column. Following their final response, they were given a score corresponding to their total reaction time across the whole sequence. In the adjacent version of

this task, the target stimuli were presented in XAB format. B) Sequence Completion task. As in the SRT-AGL task, participants were presented with a visual array of stimuli and responded to the first two visual cue stimuli as before. However, the final visual cue stimulus was replaced with a question mark, and the participants were asked to guess the shape that they felt completed the sequence. C) Sequence Generation task. In each of the 24 trials, the participants were presented with 8 elements arranged in a circle on the screen: the 2 ‘A’ and ‘B’ elements were always shown, as well as 4 randomly selected, non-repeating ‘X’ elements. Participants created their sequences by clicking on the desired elements in order. Each trial was separated by a 500ms inter-trial interval.

Data Analysis

For the SRT-AGL task, a repeated measures ANOVA was used to compare reaction time differences to grammatical and ungrammatical sequences across blocks. Performance in the Sequence Completion and Grammaticality Judgement task was compared to chance (50%) using one sample *t* tests. To correlate performance on the SRT-AGL task with the reflection-based tasks, Pearson's correlation coefficients were calculated.

Results

In this experiment, there was no evidence of learning of the adjacent or nonadjacent dependencies across any of the tasks. We hypothesised that implicit learning would result in quicker responses to the predictable 'B' element than to the unpredictable 'A' element on grammatical trials relative to ungrammatical trials. Within each block of the SRT-AGL task, for both the non-adjacent and adjacent conditions, we calculated reaction time differences: $RT_{Difference}$ = ungrammatical $RT_{A\text{-}B}$ – grammatical $RT_{A\text{-}B}$. If learning had occurred, then this difference would increase across blocks. We conducted a 2x8 AVOVA, with blocks 1 to 8 as within-subject factors, and the task (nonadjacent or adjacent) as between subject factors. We found no significant effect of run ($F_{4,135} = .498$, $p = .758$) or task ($F_{1,30} = .431$, $p = .516$), or a run*task interaction ($F_{4,135} = 1.133$, $p = .345$), which suggests that there was no decrease in A-B reaction times to grammatical over ungrammatical sequences in either version of the task.

Participants did not perform above chance in any of the explicit tasks. In the non-adjacent condition, participants did not perform significantly above chance in either the Sequence Completion task ($t_{14} = 1.438$, $p = 0.172$) or Grammaticality Judgement task ($t_{14} = 1.00$, $p = .334$). In the adjacent condition, participants performed slightly below chance on the Sequence Completion task ($t_{16} = -2.27$, $p = .037$). In the subsequent Grammaticality Judgement task, participants performed at chance levels ($t_{16} = -.965$, $p = .348$). Only one of the participants who completed the nonadjacent task (shown in green in Figure. 2.3) performed above chance in the reflection-based tasks. This single participant who performed above chance did not show the predicted pattern of performance in the SRT-AGL task, suggesting that any learning occurred after the SRT-AGL task.

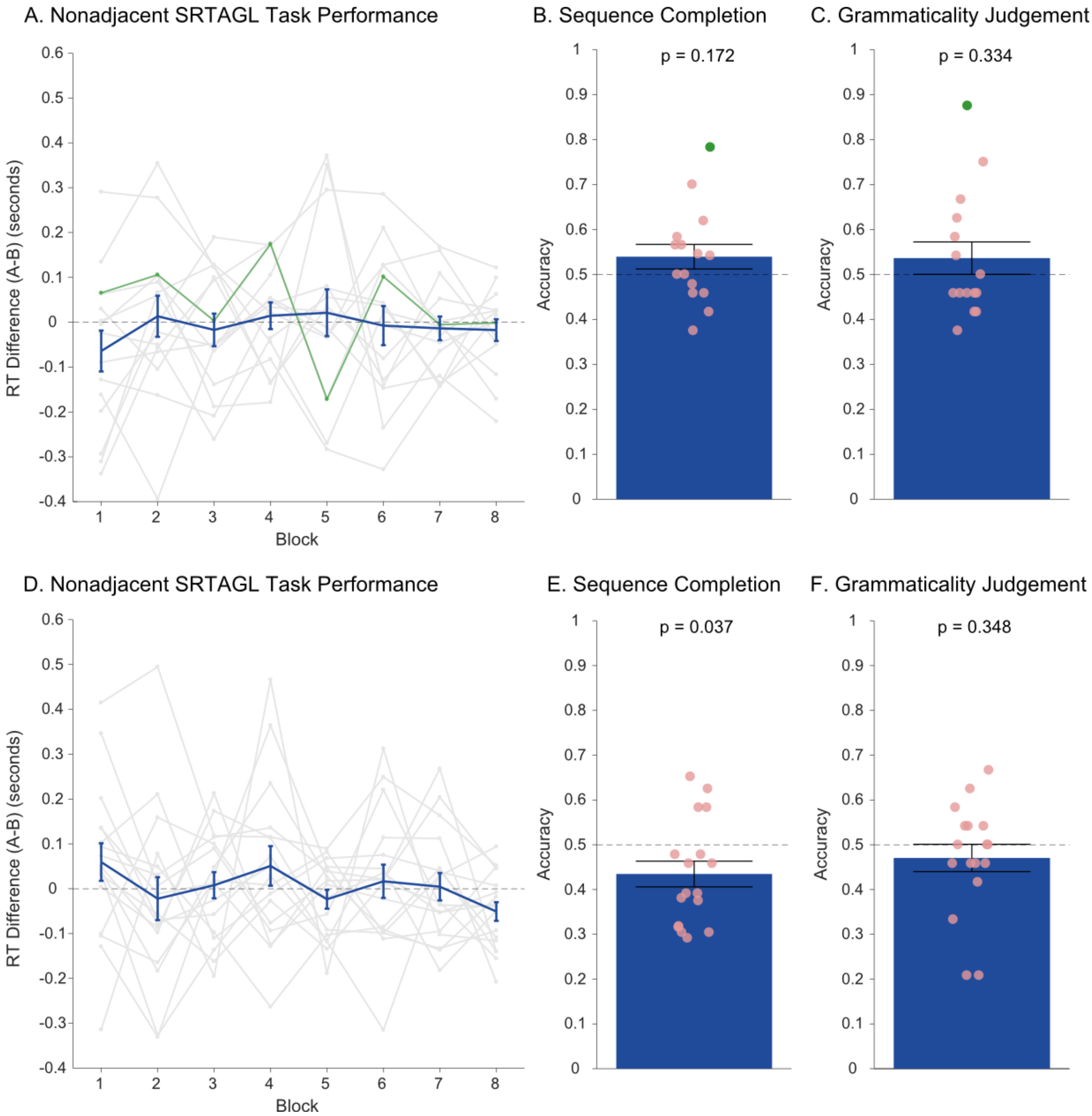


Figure 2.3. Experiment 2.1. SRT-AGL Sequence Completion and Grammaticality Judgement task performance. A) Difference plot showing the difference in mean reaction time differences ($RT_{A\text{-}B}$ ± SEM) between grammatical and ungrammatical sequences in the nonadjacent condition across blocks. Individual data is shown in grey. B) Mean (± SEM) performance in the nonadjacent sequence completion task. Individual performance is shown in circles. Good learners are shown in green (based on individual binomial tests, $p < 0.05$), non-learners are shown in red ($p > 0.05$). C) Mean performance (± SEM) in the nonadjacent Grammaticality Judgement task, including learners and non-learners. D) Difference plot showing the difference in mean reaction time differences ($RT_{A\text{-}B}$ ± SEM) between grammatical and ungrammatical sequences in the adjacent condition across blocks. Individual data is shown in grey. E) Mean (± SEM) performance on the sequence completion task. Individual performance is shown in circles. There were no 'good learners' in this task, therefore non-learners are shown in red ($p > 0.05$). F) Mean performance (± SEM) on the Grammaticality Judgement task.

In the Sequence Generation task, we examined whether participants who showed good performance on the reflection-based tasks were also more likely to produce grammatical sequences in the Sequence Generation task. We found no correlations between performance in the Sequence Generation task and performance in either the Sequence Completion (adjacent: $r = .28$, $p = .269$; nonadjacent: $r = .24$, $p = .395$) or Grammaticality Judgement (adjacent: $r = .16$, $p = .529$; nonadjacent: $r = .479$, $p = .071$) tasks, which was unsurprising given that only one participant showed evidence of learning across the experiment. We also calculated composite measures of SRT-AGL performance based on the $RT_{(A\text{-}B)}$ difference between grammatical and ungrammatical trials between Learning Block 1 and Learning Block 8, and explicit task performance (based on the mean performance in the Sequence Completion and Grammaticality Judgement tasks). These measures were not highly correlated in either the nonadjacent ($r = -.242$, $p = .385$) or adjacent tasks ($r = .102$, $p = .689$), although this is again not surprising given the lack of learning across all tasks.

Discussion

In this experiment we aimed to assess implicit statistical learning using a visual serial reaction time task that did not rely on any prior semantic or linguistic knowledge, with the hope of developing a task that could be used in other groups, including children or people with language difficulties. However, we found little evidence of learning in either the SRT-AGL task or the Sequence Completion, Grammaticality Judgement and Sequence Generation tasks for either the nonadjacent or adjacent versions of the task. Moreover, we only identified a single participant who showed good performance across the reflection-based tasks, and this participant showed no evidence of learning in the SRT-AGL task. This may suggest that participating in the SRT-AGL task may inhibit learning that may have otherwise occurred during a standard artificial grammar learning task.

The lack of learning in the reflection-based measures that is typically found during artificial grammar learning tasks could be attributed to the lack of exposure phase that is typically found in artificial grammar learning paradigms. It is possible that completing the SRT-AGL task may act as a distraction, and draw attention away from the dependencies, resulting in a lack of learning and therefore poor performance across tasks. Although the lack nonadjacent dependency learning in this task could be attributed to the more general inconsistency of learning of these dependencies reported in the literature (for a review, see Wilson et al., 2020), the lack of learning of simple adjacent relationships during this task strongly suggests that the complexity of nonadjacent relationships is not responsible for the lack of learning across these experiments. Instead, as outlined above, completing the SRT-AGL task itself may have inhibit learning of both types of dependency.

SRT-AGL tasks aim to use reaction times to provide a processing-based measure of implicit statistical learning which is not confounded by spatial information. Previous research has shown that they can be effective measures of implicit statistical learning (Misyak et al., 2009, 2010; Vuong et al., 2011), and in these experiments we aimed to replicate and adapt the task to be more suitable for testing individuals with language difficulties in the future. We failed to show learning at a group level in all three experiments, despite a small number of participants showing predicted patterns of learning. The findings from these experiments raise the question of why these SRT-AGL tasks do not accurately measure implicit statistical learning for the majority of individuals. In traditional serial reaction time (SRT) tasks, participants are required to respond based on spatial positions (Nissen & Bullemer, 1987), where one location predicts another, which relies on motor learning (Conway, 2005; Heyes & Foster, 2002; Robertson, 2007). However, in SRT-AGL tasks, whilst the participants must still locate the elements on the screen, it is the stimuli that predict one another, not the location. This design was to ensure that participants learned the relationships between the stimuli, however it also means that the participants only need to attend to the shapes enough to identify their key features, which could be less salient. It is possible that matching auditory or visual cues to their respective visual stimuli in the SRT-AGL task did not require participants to take note of the dependencies between stimuli, resulting in a lack of learning in the SRT-AGL tasks compared to more traditional SRT tasks.

Although we found no evidence of learning in the SRT-AGL task, there is still a need for processing-based measures of implicit statistical learning that are not affected by conscious reflection. Recently, serial recall tasks have been used to as a processing-based measure of implicit statistical learning (Isbilen et al., 2017). In the following experiments (Experiments 2.2 - 2.6), we outline a novel serial visual recall task that can be used to measure implicit statistical learning of variable, language-like grammars.

Experiments 2.2 – 2.6: Serial Visual Recall as a Measure of Implicit Statistical Learning

In addition to enabling faster responses to predictable events, implicit statistical learning can facilitate the rapid processing of large amounts of incoming information (Christiansen & Chater, 2016). Working memory, particularly for rapidly serially presented information is very limited (Cowan, 2010; Miller, 1956; Vogel et al., 2001). Therefore, when presented with even modest sequences of stimuli the limits of working memory can quickly be exceeded. This cognitive limitation can be overcome by combining frequently co-occurring stimuli into larger "chunks", to reduce working memory demands. This process of chunking incoming information is also critical for language learning and processing, to allow sufficiently rapid processing of the large amount of information contained within human language (Christiansen & Chater, 2016). These processes are important in artificial grammar learning tasks, for example, in Saffran et al.'s (1996) seminal study, from a rapid stream of syllables, predictable combinations of syllables are chunked to form 'words' based only on statistical regularities. This process makes rapid speech processing more efficient, as chunks can quickly be formed and passed to higher levels of representation: when processing speech we do not have to store every phoneme that is processed in order to recall the semantic meaning of the sentence.

Based on this reasoning, if participants are presented with sequences which contain predictable regularities, frequently co-occurring elements should be more easily combined into chunks. This in turn may reduce the demands on working and allow for more efficient chunk-and-pass processing, in which input is rapidly, incrementally chunked at multiple levels of linguistic structure (from phonemes to words, then to phrases and sentences) in order to process this input before new information arrives (Christiansen & Chater, 2016). We can use this facet of cognition to our advantage to design a processing-based measure of implicit statistical learning. Artificial grammars can be used to generate predictable and unpredictable sequences. Predictable sequences contain more frequently co-occurring elements, therefore these sequences can be chunked and recalled more easily than unpredictable sequences. This has recently been demonstrated in visuo-motor tasks (Conway et al., 2007), in auditory (Isbilen et al., 2017; Kidd et al., 2020) and visual statistical learning tasks (Isbilen et al., 2020) and in natural language (McCauley & Christiansen, 2015). In Isbilen et al.'s (2020) experiment, participants first completed a passive exposure phase, where they were presented with sequences of syllables arranged into predictable trisyllabic 'words', before completing both a novel Serial Implicit Chunking Recall task and a more explicit reflection-based task. In the recall task, participants were presented with a sequence of nonword syllables consisting of either predictably or randomly ordered syllables and asked to verbally recall the sequence. The reflection-based task was a two-alternative forced-choice (2AFC) task, in which participants are presented with both a predictable and random sequences and asked to select the sequence that follows the pattern they had previously heard during the exposure phase. They found that participants had improved verbal recall of syllables in predictable sequences over random sequences and performed above chance in the 2AFC task. However, performance on these two tasks were not correlated, suggesting that the 2AFC task, which relies on conscious decision making, may be measuring different processes to the more implicit recall task.

The Serial Implicit Chunking Recall task has also been adapted to assess implicit statistical learning of nonword stimuli in the visual domain, using written transcriptions of the nonword syllables from the original auditory experiment. As in the auditory version of the task, participants were shown to have improved recall of predictable over random sequences, as well as above chance performance in the reflection-based task. However, in the visual domain, they did find evidence of a positive correlation between performance across the tasks that was not found in the auditory domain. The authors attributed this to the fact that the visual recall task may involve more reflection than the auditory recall task, as participants were able to revisit and revise their responses more easily than in the auditory recall task, where any amendments had to be made from memory. The findings from these tasks demonstrate that processing based-measures of implicit statistical learning work and may be measuring something substantially different from more reflection-based measures.

Serial recall tasks have been shown to effectively measure implicit statistical learning and may prove to be a more reliable measure of learning than more traditional reflection-based tasks (Isbilen et al., 2020). However, it is currently unclear what constraints are placed on measuring learning using these tasks. Both the auditory and visual SICR tasks are based on the seminal statistical learning paradigm introduced by Saffran et al., (1996), in which the stimuli consist of syllables arranged into predictable trisyllabic ‘words’. As the syllables within a word always co-occur together, it is cognitively efficient to ‘chunk’ these items together in working memory, rather than storing each phoneme separately, meaning the representations for these words are different to those of ‘part-words’. This ability to chunk frequently co-occurring stimuli appears to facilitate the serial recall effect in both the auditory and visual SICR tasks (Isbilen et al., 2022; Isbilen et al., 2017, 2020; McCauley & Christiansen, 2015). However, implicit statistical learning has been shown to extend far beyond learning of word boundaries, in fact, similar processes have been shown to underlie learning of more variable relationships between words in artificial grammars (Newport & Aslin, 2004).

In a series of 5 experiments, we aimed to extend previous research to investigate if visual serial recall can be used to measure implicit statistical learning of more variable, language-like grammars. In each experiment, following the serial visual recall task, participants completed a number of reflection-based tasks: a traditional Grammaticality Judgement task, Sequence Generation and Sequence Completion task. The Sequence Generation and Sequence Completion tasks were included to assess the extent to which any sequence knowledge that was obtained was available to consciousness, as the ability to generate and complete sequences would suggest more explicit knowledge of the structure (Destrebecqz & Cleeremans, 2001; Wilkinson & Shanks, 2004). Across all 5 experiments, we predicted that we would see evidence of learning across the both the processing-based tasks and the reflection-based tasks. If processing-based measures of learning are measuring different processes than reflection-based measures, then there will be no correlation between performance in the Visual Recall task and performance across the subsequent reflection-based tasks.

In Experiments 2.2 and 2.3 (in-person and online, respectively), using methods based on previous visual artificial grammar learning paradigms we saw no evidence of learning based on visual serial recall, but did replicate prior learning effects using the subsequent reflection-based measures. In Experiments 2.4 and 2.5 (in-person and online, respectively), we adapted the serial visual recall task away from a traditional exposure-test paradigm to use a blocked design, where participants complete many recall blocks containing predictable sequences, before completing a block containing unpredictable sequences followed by a final block containing predictable sequences. We also developed a novel artificial grammar designed to be more susceptible to chunking. In these experiments, strong learning effects were observed in the serial visual recall task, as well as the subsequent reflection-based measures, and performance across these tasks was positively correlated. To assess whether the facilitation of learning in the visual serial recall task was due to the changes in procedure or the new grammar, in Experiment 2.6 we conducted one further online using the new methods from Experiments 2.4 and 2.5 and the artificial grammar from Experiments 2.2 and 2.3. In Experiment 2.6 we again found strong evidence of learning across all measures of learning, however performance across processing- and reflection-based tasks was not correlated. Taken together, these findings suggest that the success of the visual serial recall task measuring learning is not dependent on the grammar being learned or based on whether the experiment takes place in-person or online. However, the design of the visual serial recall task is an important factor in successfully measuring learning, with blocked designs proving to be effective where oddball tasks are not. Furthermore, these findings suggest that the complexity of the grammar may affect the extent to which explicit processes can be used, and therefore whether processing- and reflection-based measures capture similar mechanisms of learning.

Experiments 2.2 & 2.3: Visual Artificial Grammar Learning Recall Task

Introduction

In Experiments 2.2 & 2.3, we aimed to investigate whether serial visual recall could be used to effectively measure learning of the between-word relationships found in artificial grammars which has previously used in both the auditory and visual modalities in humans and nonhuman primates (Milne et al., 2018; Saffran et al., 2008; Wilson et al., 2015). To do this, we adapted the design and stimuli of a previous AGL study (Milne et al., 2018), and integrated it with a novel Visual Recall task, alongside the existing Grammaticality Judgment task and new Sequence Generation and Sequence Completion tasks. The artificial grammar consists of five stimuli, in this case abstract visual shapes, which were presented serially in sequence. After exposure to "grammatical" sequences generated by the artificial grammar, the participants were presented with novel testing sequences, and, after a brief pause, were asked to recall the sequence by clicking on the visual symbols on the screen in order. In Experiments 2.2 and 2.3 we predicted that participants would show evidence of learning in the Visual Recall task, evidenced by an increase in recall accuracy of grammatical, but not ungrammatical sequences across blocks. We also predicted that participants would perform above chance in the subsequent Grammaticality Judgement task, and that performance in this task would be positively correlated with other reflection-based tasks. Finally, we predicted that there would be no correlation between performance in the visual recall task and performance across the subsequent reflection-based measures.

Methods

Participants

In Experiment 2.2, 22 adult participants (15 female, 7 male, mean age = 30.1) were recruited using the Institute of Neuroscience participant pool at Newcastle University. We aimed to test approximately 40 participants – similar numbers to previous experiments using recall as a measure of implicit statistical learning, which recruited between 26 and 69 participants (e.g. Isbilen et al., 2017) – however due to the impact of COVID-19 we were forced to complete recruitment prematurely. All participants were native English speakers, and had normal or corrected-to-normal vision and hearing. Participants were not excluded based on their ability to speak any additional languages. Ethics was approved by the Faculty of Medical Sciences Ethics Committee at Newcastle University.

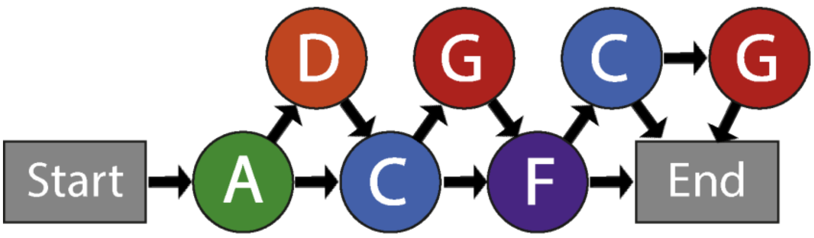
Experiment 2.2 was carried out in-person prior to the pandemic. Due to the COVID-19 lockdown, further in-person data collection was not possible. Therefore, we recoded the experiment to enable them to run online. In Experiment 2.3, 43 participants (26 female, 17 male; mean age = 30.98 years) were recruited using Prolific, an online recruitment platform. As previously mentioned, this sample size was similar to previous experiments using recall as a measure of implicit statistical learning (Isbilen et al., 2017). Participants were pre-screened to include native English speakers, and to exclude participants who had language disorders, as previous research has suggested that there may be deficits in implicit statistical learning in these populations (Folia et al., 2008; Hsu & Bishop, 2014; Obeid et al., 2016). Participants were not excluded based on their ability to speak any additional languages. An additional 7 participants completed the experiment but were excluded from analysis for failing attention checks. Ethics was approved by Emory University IRB.

Stimuli

The sequences for the Visual Recall task in both Experiments 2.2 and 2.3 were generated using an artificial grammar developed by Saffran et al. (1999, 2008), using abstract white shapes inspired by previous artificial grammar stimuli (Conway & Christiansen, 2006; Milne et al., 2018; Osugi & Takeda, 2013; Seitz et al., 2007). The grammar consisted of 5 elements (A, C, D, F, G), each represented by an abstract shape (see Figure. 2.4). In all phases of the task these stimuli were presented as white shapes (200*200) pixels on a black background (see Figure. 2.4). The recall task itself was split into exposure and testing phases. The exposure phases consisted of 8 different grammatical sequences presented 8 times, totalling 64 sequences (see Figure 2.4). This included all possible grammatical sequences, except those which were 5 elements long, which were not presented during the exposure phase so they would remain novel to the participants in the testing phase.

In the testing phase, participants were presented with 5-element-long grammatical and ungrammatical sequences, none of which had previously appeared in the exposure phase. There were 4 grammatical sequences and 8 ungrammatical sequences (see Figure. 2.4.), so each grammatical sequence was presented twice per phase to ensure the number of grammatical and ungrammatical sequences was balanced. Each ungrammatical sequence contained at least one illegal transition, that is, a transition between elements that had not occurred during the exposure phase and therefore had a TP of 0

A. Artificial Grammar



Exposure Sequences	Mean TP
ACF	0.40
ACFC	0.48
ACGF	0.54
ACGFCG	0.55
ADCF	0.60
ADCFCG	0.59
ADCGFC	0.64
ADCGFCG	0.63

B. Testing Sequences

Testing Sequence	Mean TP
ACFCG	0.48
ADCFC	0.52
ACGFC	0.48
ADCGF	0.60
ADCFG	0.45
ADFCG	0.42
ADGCF	0.24
ADGFC	0.37
AGCFG	0.15
AGFGC	0.15
AGDCF	0.34
AGFDC	0.35

Figure 2.4. Experiments 2.2 and 2.3 Artificial Grammar and Exposure Stimuli. A) Illustration of the artificial grammar and the exposure sequences used in Experiments 2.2 and 2.3. Sequences are produced by following the arrows from the start to the end. The grammar contains 5 elements which are represented by abstract shapes. B) The testing sequences consisted of 4 grammatical sequences (shown in blue), each of which was repeated twice per block, and 8 ungrammatical sequences (shown in red), each presented once per block. The mean TP of the ungrammatical sequences was lower than the grammatical sequences, meaning ungrammatical sequences contain fewer frequently co-occurring transitions and are therefore harder to chunk.

Procedure

Experiment 2.2 took place in-person, in a testing lab with the same set up as Experiment 2.1. Responses were made either with the mouse (in the Visual Recall, Sequence Generation and Sequence Completion tasks) or by pressing one of two keys on the keyboard (in the Grammaticality Judgement tasks). To assess the relationship between implicit statistical learning and language abilities, we conducted the same battery of standardised cognitive and language tests before and after the main experiment as in Experiment 2.1. Before the experiment, the TOWRE Words and Nonwords, Backwards Digit Recall and the Recalling Sentences Task were administered (see Appendix 2.1.). After completing the Visual Recall, Grammaticality Judgement, Sequence Generation and Sequence Completion tasks, the WASI Block Design, Rapid Automatized Naming Digits and Objects tasks were administered.

In Experiment 2.3 the Visual Recall, Grammaticality Judgement, Sequence Generation and Sequence Completion tasks were adapted from MATLAB to PsychoPy (version 2021.2.3) to enable them to run online through Pavlovia. Participants completed the experiment on their own desktop or laptop computer. No standardised cognitive and language tasks were conducted as part of this experiment, as these require in-person contact.

As in traditional artificial grammar learning paradigms, in both Experiments 2.2 and 2.3 the Visual Recall task was split into two phases: exposure and testing. In both of these phases, each sequence was presented serially across the screen (Figure 2.5). Each element was presented on the screen for 450ms before being removed, and the elements were separated by an inter-element interval of 300ms. In both exposure and testing phases, each sequence was separated by an inter sequence interval of 1000ms.

Exposure phase

In the exposure phase, the participants were asked to pay attention to the sequences but were not asked to make any responses. Participant were not told about the presence of any patterns in the sequences. In the first exposure phase (following the baseline recall test) participants were presented with 64 grammatical sequences, consisting of all 8 possible grammatical sequences repeated 8 times. This phase lasted approximately 5 minutes. Subsequent exposure phases were designed to refamiliarize the participants with the grammatical sequences, so were shorter, presenting 24 sequences and lasting approximately 2 minutes.

In the online version of the task (Experiment 2.3), attention checks to the exposure phase to ensure participants were paying attention to the sequences. One in eight of the exposure sequences was randomly selected to be an attention check sequence. In an attention check sequence, one element in the sequence was randomly selected and replaced with a star shape. Participants were instructed to attend to the sequences as in Experiment 2.2, and to press the “space” key whenever they saw a star within a sequence.

Testing phase

In the testing phase, each testing sequence was presented in the same way as in the exposure phase. After the sequence was presented, there was a 1000ms retention period. Following this, the stimulus elements were presented simultaneously in a circle on the screen (see Figure. 2.5). The position of each element was randomised on each trial, so that participants could not rely on positional cues or motor sequence learning. The participant was asked to recreate the sequence by clicking on the appropriate elements in the correct order. No feedback was given. An inter-trial interval of 1200ms separated the participant’s response from the presentation of the next sequence. Participants completed 4 testing phases in total. The experiment began with a baseline testing phase to assess working memory in each participant. This phase was identical to the other testing phases, except that it was not preceded by an exposure phase, and therefore we would predict no differences in recall accuracy between the ‘grammatical’ and ‘ungrammatical’ sequences. All subsequent testing phases were separated by a short exposure phase.

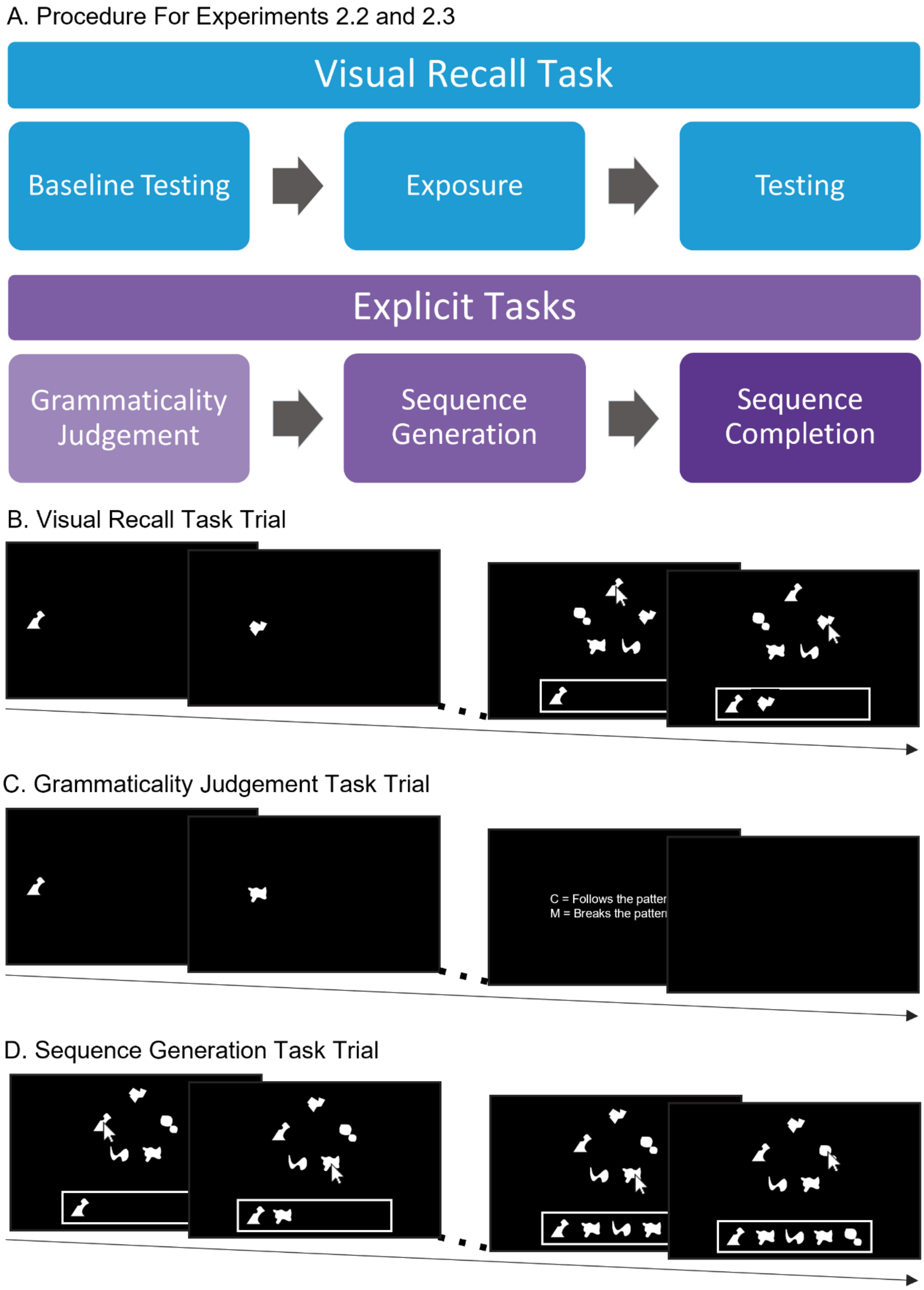


Figure 2.5. Experiments 2.2 and 2.3 procedure and trial design. A) Procedure. In each trial, a sequence of 5 shapes was presented serially across the screen. Each shape was presented for 450ms, and each shape was separated by a 300ms inter-stimulus interval. After the sequence had been displayed, there was a 1000ms retention period. Following this, participants were presented with all 5 possible stimuli simultaneously on the screen. Participants were asked to

recreate the sequence by clicking on the desired elements in order. Each trial was separated by a 1500ms inter-trial interval. B) Grammaticality Judgment task trial. Participants were presented with grammatical and ungrammatical sequences. Following this, participants pressed one of two keys on the keyboard to indicate whether they felt that the sequence followed the same pattern as the sequence they had seen previously or not. Each trial was separated by a 1500ms inter-trial interval. C) Sequence Generation task. In each of the 8 trials, participants were presented with all 5 possible elements arranged in a circle on the screen. Unlike the other tasks, no sequence was presented for recall or completion. Instead, participants created their own 5-element-long sequences by clicking on the desired elements in order. Each trial was separated by a 1500ms inter-trial interval. D) Sequence Completion task. In each of the 24 trials, participants were presented with a 5-element long sequence in which one of the elements was replaced by a question mark. Participants were asked to complete the sequence by clicking on the desired shapes to fill in the gap. Each trial was separated by a 1500ms inter-trial interval.

Grammaticality Judgement Task

After the Visual Recall task, the participants then completed the Grammaticality Judgement task. Prior to this task participants were re-familiarised with the grammatical sequences through a short exposure phase lasting 2 minutes, as described above. The participants were then informed that the sequences that they had just seen followed a pattern, and that they would see new sequences, some of which follow the same pattern and some that would not. The same 8 grammatical and 8 ungrammatical sequences that were used in the testing phase of the recall task were presented in a random order. For this task, once the sequence was presented, participants were asked to judge if the sequence followed the pattern or not by pressing one of two keys on a keyboard. Participants completed two runs of the Grammaticality judgement task, separated by a short 2-minute exposure phase.

Sequence Generation Task

In the Sequence Generation task, all elements were displayed on the screen, as shown in Figure 2.5. In each trial, participants were asked to create their own 5 element long sequences that fit the pattern they had seen previously. Participants created the sequences by clicking on the desired elements in order, in the same way that sequences were recalled in the testing phase of the Visual Recall task. Once the sequence had been created, it remained on screen for 1000ms, after which participants were presented with a black screen for 1000ms between trials. The Sequence Generation task consisted of 8 trials.

Sequence Completion Task

In the Sequence Completion task, participants were presented with a 5-element-long grammatical sequence with one element missing and instructed to try to select the appropriate element to fill in the gap. The Sequence Completion task consisted of 20 trials, and the sequences were counterbalanced so that for each of the grammatical test sequences, the missing element occurred once in each of the 5 positions within the sequence.

The Sequence Generation task provides more information regarding the extent to which participants are consciously aware of any implicitly acquired knowledge, as the ability to generate sequences relies on gaining explicit access to knowledge of the dependencies (Destrebecqz & Cleeremans, 2001; Wilkinson & Shanks, 2004). As in Experiment 2.1, this task was included to assess the extent to which participants had access to the information they had (presumably) learned implicitly, in order to make more explicit decisions. Although at face value both the processing-based Visual Recall task and reflection-based Sequence Generation task both involve creating sequences, the completion of these task relies on different processes.

The Visual Recall task is a simple working memory task: participants were asked to remember and recreate sequences, which does not require participants to be aware of any of the rules underlying the sequences they are recalling, and indeed participants are not told about the presence of rules when completing the task. In the Visual Recall task, implicit statistical learning is assessed indirectly by comparing recall of grammatical and ungrammatical sequences, with the prediction that if learning has taken place, that recall of grammatical sequences should be improved relative to ungrammatical sequences. Conversely, prior to the Sequence Generation task, participants are told about the presence of rules underlying the sequences they have seen and asked to create sequences that follow this pattern. Success in this task requires participants to have at least some awareness of the rules, and therefore is likely to rely on different, more explicit processes compared to the Visual Recall task.

Data Analysis

In the Visual Recall task, there were two measures of performance. The first method was to score a trial as correct if the participant recalled the entire sequence correctly (henceforth *absolute correct* score). The second method was to calculate the proportion of each sequence that was correctly recalled (henceforth *proportion correct* score). In the Grammaticality Judgement task, a trial was scored as correct if the participants successfully classified it as grammatical or ungrammatical, and performance on this task was compared to chance levels (50%). In the Sequence Generation task, a trial was marked correct if the participant generated a sequence that was completely correct. In the Sequence Completion task, a trial was scored as correct if the participant chose the correct element to complete the sequence. This task cannot be solved based on exclusion alone (if 4 shapes are present, it must be the remaining shape) as the grammar allows for some repetition. As only one element out of each 5-element long sequence was missing, chance performance was 20%.

In the online version of the task, we calculated the percentage of stars correctly identified in the exposure phase. As we expected the majority of learning to occur in the longer initial exposure phase, it was particularly important to ensure that participants were paying attention in this block. Therefore, any participants who failed to score above 75% in responding to the attention checks within the initial exposure phase were excluded from the analysis. In the subsequent shorter exposure phases, any participants who did not correctly identify over 60% of stars in more than one block were excluded from the analysis.

Results

In both the in-person and online Visual Recall tasks, we predicted that recall accuracy would improve across testing blocks for the grammatical sequences relative to the ungrammatical sequences, as participants learned the statistical relationships between the elements leading to higher levels of chunking in these more predictable sequences. We first analysed the data based on absolute correct scores (2x4 repeated measures ANOVA with factors: Condition (grammatical and ungrammatical) and Run (4 runs). For the in-person experiment, there was a main effect of run ($F_{3, 63}$ = 47.389, p < .001), indicating an improvement in recall accuracy of both grammatical and ungrammatical sequences over the course of the experiment. Post-hoc tests (Bonferroni corrected) indicated significant differences in recall accuracy between the baseline run and subsequent testing runs (p < .001 in all cases), and between testing run 1 and the final testing run (p = .006), but not between testing run 1 and testing run 2 (p = .101). There were no significant differences in recall accuracy between other runs. Moreover, there was a significant main effect of condition, however this indicated that recall of grammatical strings was poorer than ungrammatical strings ($F_{1, 21}$ = 6.73, p = .017). We also found a significant interaction between condition and run ($F_{3, 63}$ = 6.09, p = .001), indicating that recall of grammatical sequence improved across runs to a greater extent than recall of ungrammatical sequences. These findings were also reflected in the absolute correct scores in the online experiment: there was a main effect of run ($F_{3, 87.58}$ = 41.38, p < .001). Bonferroni corrected post-hoc tests indicated significant differences in recall accuracy between baseline run and subsequent testing runs (p < .001), between testing run 1 and testing run 3 (p = .013) but not testing run 2 (p = .454). There were no significant differences in recall accuracy between other runs. There was no main effect of condition ($F_{1, 42}$ = .056, p = .814), and a significant interaction between condition and run ($F_{3, 126}$ = 3.77, p = .012). A similar pattern of results was observed when using proportion correct scores in the in-person experiment: there was a main effect of run ($F_{2.23, 46.74}$ = 41.10, p < .001). Post-hoc tests (Bonferroni corrected) indicated similar significant differences in recall accuracy between the baseline run and subsequent testing runs (p < .001), and between testing run 1 and the final testing run (p = .004), but between testing run 1 and testing run 2 (p = .053). There were no significant differences in recall accuracy between other runs. There was no main effect of condition ($F_{1,21}$ = 2.07, p = .165), and an interaction between run and grammaticality ($F_{2.17, 45.47}$ = 6.79, p = .002). In the online experiment, there was a main effect of run ($F_{3,80}$ = 34.76, p < .001). Post-hoc tests (Bonferroni corrected) showed significant differences in recall accuracy between the baseline run and subsequent testing runs (p < .001), but no significant differences between other testing runs.

There was no main effect of condition ($F_{1, 42}$ = .088, p = .768), and no interaction between run and condition ($F_{3, 126}$ = 1.88, p = .136).

Although there is some evidence of an interaction between condition and run in both in-person and online experiments, this effect was driven by particularly poor recall accuracy of grammatical sequences in the baseline block (Figure. 2.6), before any exposure to grammatical sequences had taken place. This could be due to the participants, prior to seeing the exposure sequences, wanting to avoid repetition of elements within a sequence. In these experiments, grammatical sequences are more likely to contain repeating elements than ungrammatical sequences, which means that if before exposure participants are avoiding repetition, this will negatively impact performance on grammatical but not ungrammatical strings. Following exposure, where participants have seen many examples of repetition within a sequence, they may be less likely to avoid repetition, and indeed, in the remaining runs of both experiments, there was no difference in recall accuracy between grammatical and ungrammatical sequences, as we would predict if learning had occurred.

Table 2.1. Descriptive statistics for the Visual Recall tasks in Experiments 2.2 and 2.3

	Experiment 2.2 (In-person)				Experiment 2.3 (Online)			
	Grammatical		Ungrammatical		Grammatical		Ungrammatical	
	Mean	SEM	Mean	SEM	Mean	SEM	Mean	SEM
Baseline	0.12	0.03	0.33	0.05	0.27	0.04	0.36	0.04
Run 1	0.54	0.07	0.55	0.05	0.59	0.05	0.58	0.04
Run 2	0.61	0.05	0.67	0.06	0.65	0.04	0.62	0.05
Run 3	0.70	0.06	0.66	0.06	0.71	0.04	0.69	0.05

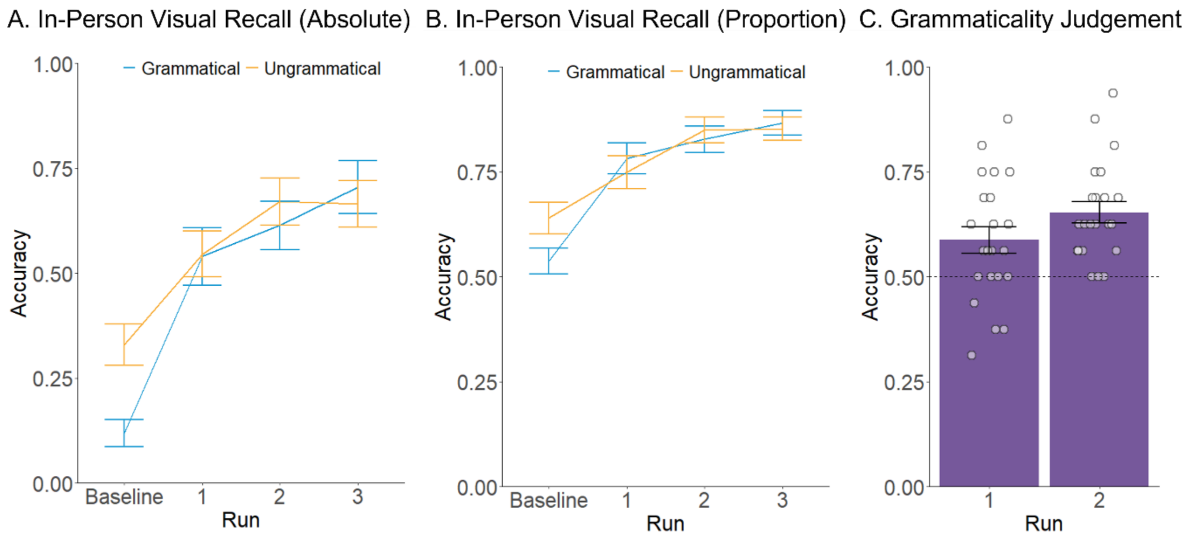


Figure 2.6. Experiment 2.2 in-person Visual Recall and Grammaticality Judgement task performance. In all panels, error bars represent +/- 1 SEM. A) Mean absolute recall accuracy for grammatical and ungrammatical sequences across testing blocks 1 to 3. B) Mean proportion correct recall accuracy for grammatical and ungrammatical sequences across testing blocks. In both A and B there was no evidence of a difference in recall accuracy between the grammatical and ungrammatical sequences across runs. Performance on the Grammaticality Judgement task showed above chance performance (indicated by the dashed line) in both runs. Individual performance is shown in white.

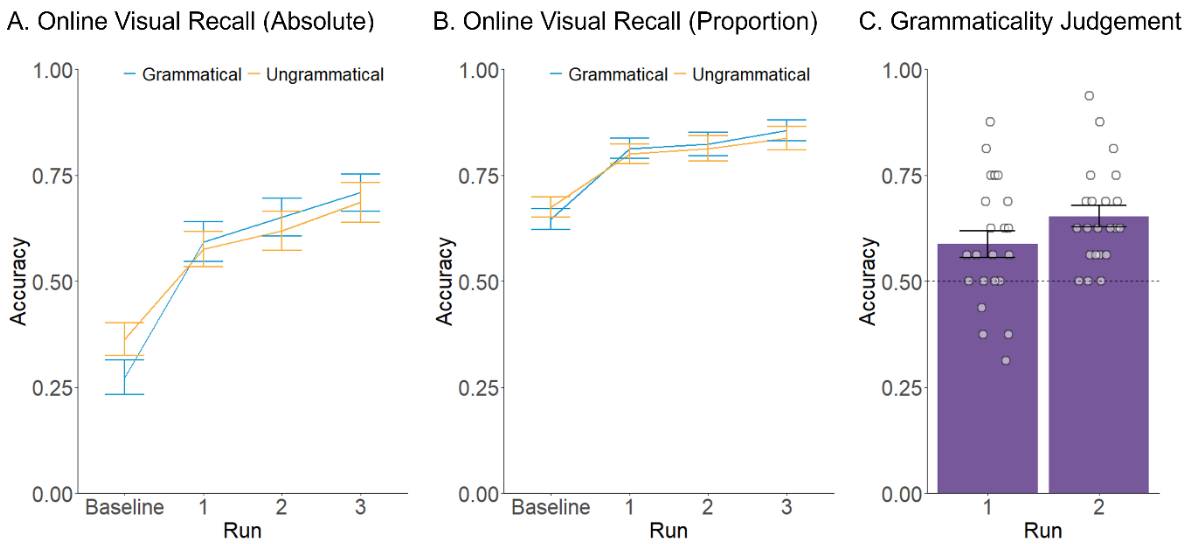


Figure 2.7. Experiment 2.3 online Visual Recall and Grammaticality Judgement task performance. In all panels, error bars represent +/- 1 SEM. A) Mean absolute recall accuracy for grammatical and ungrammatical sequences across testing blocks 1 to 3. B) Mean proportion correct recall accuracy for grammatical and ungrammatical sequences across testing blocks. In both A and B there was no evidence of a difference in recall accuracy between the grammatical and ungrammatical sequences across runs. Performance on the Grammaticality Judgement task showed above chance performance (indicated by the dashed line) in both runs. Individual performance is shown in white.

In Experiment 2.2, participants were significantly better than chance at correctly classifying the testing sequences as grammatical or ungrammatical across both runs of the Grammaticality Judgement task (run 1: $M = 0.59$, SEM = 0.031; $t_{21} = 2.796$, $p = .011$; run 2: $M = 0.65$, SEM = 0.025; $t_{21} = 6.156$, $p < .001$). There was some indication that performance improved across runs, however this difference did not reach significance ($t_{21} = -1.879$, $p = .074$). In the online experiment, participants were close to performing above chance in run 1, however this did not reach significance ($M = 0.55$, SEM = 0.03; $t42 = 2.01$, $p = .051$). In run 2, participants did perform significantly better than chance ($M = 0.58$, SEM = 0.03; $t_{42} = 2.93$, $p = .005$). Similarly to Experiment 2.3, whilst performance in run 2 was better than run 1 in the online task, this difference was not significant ($t_{42} = 1.15$, $p = .257$). These findings suggest that across both in-person and online experiments, there may be some improvement in performance once participants have been informed about the presence of rules and been given an opportunity to look for these rules during exposure.

In the Sequence Generation task, we examined whether participants who showed good performance on the explicit tasks were more likely to create grammatical sequences in the Sequence Generation task. In Experiment 2.2, performance on the Sequence Generation task was positively correlated with performance on the second run of the Grammaticality Judgement task ($r = .539$, $p = .009$), but not the first run ($r = .148$, $p = .511$). In the Sequence Completion task, participants performance was compared to chance (20%) using a one sample t test. Participants were significantly more likely to choose the correct element to fill in the gap (t_{21}= 18.835, $p < .001$). Similarly to the Sequence Generation task, performance in the Sequence Completion task was positively correlated with performance in the second run of the Grammaticality Judgement task ($r = .531$, $p = .011$), but not the first run ($r = .145$, $p = .519$). Performance in the Sequence Completion task was also positively correlated with performance on the Sequence Generation task ($r = .727$, $p < .001$), which we would expect given the similarity between these tasks. These findings were replicated in the online experiment: we found a positive correlation between performance in run 2 of the Grammaticality Judgement task and both the Sequence Generation ($r = .428$, $p = .004$), and Sequence Completion ($r = .635$, $p < .001$) tasks, but no correlation between performance in run 1 of the Grammaticality Judgement task and the Sequence Generation ($r = .203$, $p = .191$) and Sequence Completion ($r = .214$, $p = .169$) tasks. Performance on the Sequence Completion task was above chance (t_{42} = 26.33, $p < .001$), and there was a positive correlation between performance on the Sequence Generation and Sequence Completion tasks ($r = .452$, $p = .002$). Correlations between performance in run 2, but not run 1 of the Grammaticality Judgement task with the other reflection-based measures may suggest that being told about the presence of rules underlying the sequences after run 1 of the Grammaticality Judgement task may result in the recruitment of more explicit processes during the second run of the Grammaticality Judgement task, similar to those participants are likely to use when generating their own sequences.

To assess whether the processing-based visual recall task and subsequent reflection-based tasks are measuring similar processes, we calculated composite measures of Visual Recall task performance based on the difference in recall accuracy for grammatical and ungrammatical trials between blocks 2 and 4. We correlated this with composite reflection-based task performance (based on the mean performance in the Grammaticality Judgement, Sequence Generation, and Sequence Completion tasks). There was no correlation between these measures either in-person (r = -.175, p = .436) or online (r = .165, p = .277) which is unsurprising given that the Visual Recall task failed to measure any learning. We also examined whether performance in the Visual Recall task and subsequent reflection-based tasks was correlated with performance on the battery of language tasks, however we did not find any consistent correlations (see Appendix 2.2.), although this may also be in part due to the failure of the Visual Recall task in measuring learning.

To compare performance between the in-person and online versions of the task, we added the between-subjects factor of task (in-person or online) to the previous ANOVAs using absolute and proportion correct scores. We found no differences in performance based on whether the tasks were completed in-person or online. In the Visual Recall task, we found no main effect of task when using either absolute correct ($F_{1, 63}$ = .391, p = .534) or proportion correct ($F_{1, 63}$ = .336, p = .564) scores, and no interactions between task and the other variables. We also compared in-person and online performance in the Grammaticality Judgement task using a 2x2 mixed ANOVA, with run (2 runs) as a within-subjects factor and task (in-person or online) as a between-subjects factor. We found no main effect of task ($F_{1, 63}$ = .2.20, p = .143), and no additional interactions. Independent t tests show no differences between in-person and online performance in either the Sequence Generation task (t_{63} = 1.08, p = .2.82) or Sequence Completion task (t_{63} = 1.21, p = .230).

Discussion

In this experiment we saw no evidence of implicit statistical learning in the visual recall task, despite good performance across subsequent reflection-based measures. Performance in the second run of the Grammaticality Judgement task was positively correlated with performance in both the Sequence Generation and Sequence Completion tasks, further suggesting that learning has taken place during the experiment. Despite learning having occurred, it is clear that the Visual Recall task was not an effective measure of learning in this experiment.

There are two possible explanations for why there is evidence of learning in the explicit tasks, but not the Visual Recall task: either learning occurred during the Visual Recall task that was not reflected in improved recall performance, or learning did not occur until after the Visual Recall task. It is likely that learning did occur during the Visual Recall task, because, similarly to traditional AGL paradigms, the task consisted of exposure and testing phases, meaning participants were given similar opportunities to learn the regularities as in traditional AGL experiments where learning has previously been found (e.g., Reber, 1967; Saffran et al., 1996). Furthermore, participants show above chance performance in the first run of the Grammaticality Judgement task, and at this stage of the task, they have not had the chance to complete an exposure phase whilst knowing about the presence of rules. Therefore, any learning that has occurred at this point is more likely to be a reflection of learning that has occurred during the preceding exposure phases by implicit statistical learning, than by explicit processes seeking of rules during the exposure phases. From this, we can conclude that learning likely occurred during the Visual Recall task, but this was not reflected in improved recall of grammatical over ungrammatical sequences.

Although not significant, participants show a trend towards performing better in the second run of the Grammaticality Judgement task than the first. Furthermore, performance in the second run, but not the first run, was positively correlated with performance across the other reflection-based tasks. Both findings may be explained by the fact that the second run of the Grammaticality Judgement task is completed after the participant has had the chance to complete an exposure phase after being explicitly told about the presence of rules underlying the sequences. Being able to adopt more explicit processes to aid in rule learning during this exposure phase may facilitate performance in the second run. Furthermore, it is possible that this more explicit, top-down approach to rule learning may account why there is a positive correlation between performance in the second run of the Grammaticality Judgement task and performance in the other explicit tasks, but not the first run.

There are several methodological reasons that may explain why recall was not an effective measure of learning in this experiment. First, ungrammatical sequences were interspersed within each Recall Block, which may have interfered with learning of the grammar. Second, in the (Saffran et al., 2008) grammar being learned in these experiments, the majority of transitions were more variable than the within-word transitions from previous recall tasks, which had a TP of 1 (e.g., Isbilen et al., 2017). In Experiments 2.3 and 2.4, only one transition (element "D" to element "C") had a TP of 1, which means it is possible that recall did not reflect learning because the grammatical sequences could not be reliably chunked. Furthermore, the ungrammatical sequences used in this experiment only contain subtle violations, rather than consisting of random transitions, as in previous serial auditory recall tasks (Isbilen et al., 2022; Isbilen et al., 2017, 2020; Kidd et al., 2020). Therefore, these ungrammatical sequences consist primarily of legal transitions that can be chunked, potentially providing a memory benefit for these sequences in addition to the grammatical sequences. This may explain why we do not see any differences in recall of grammatical and ungrammatical sequences in this experiment. It is possible that recall may only be an effective measure of learning when the ungrammatical sequences consist primarily of illegal transitions. Furthermore, these ungrammatical sequences were interspersed within each Recall Block, which may have interfered with the learning of the regularities: participants may have been more engaged with the task during the Recall Blocks compared to the exposure phases, which means that some learning of ungrammatical transitions may have been learned. In Experiments 2.4 and 2.5, we made changes to the design of the Visual Recall task to understand which conditions were best for measuring learning.

Experiments 2.4 & 2.5: Chunking Visual Recall Task

Introduction

In Experiments 2.2 and 2.3, using methods based on previous visual artificial grammar learning paradigms, we saw no evidence of learning based on visual serial recall, but did replicate prior learning effects using the subsequent reflection-based measures. In Experiments 2.4 and 2.5, we made some changes to the design of the Visual Recall task to facilitate learning. In Experiments 2.2 and 2.3, ungrammatical sequences were included in each Recall Block. The aim of this design was to gain more information on the time-course of learning across the Visual Recall task, as opposed to more traditional grammaticality judgement tasks, where learning is assessed at the end of the experiment following exposure to grammatical sequences. However, it is possible that interspersing ungrammatical sequences throughout the experiment may have interfered with learning of the regularities. Therefore, in Experiments 2.4 and 2.5, we adopted a blocked design: participants completed an initial baseline block, followed by 6 Recall Blocks consisting of only grammatical trials. They then completed a Testing Block of random, ungrammatical trials, before finishing with a final "Recovery" Block of grammatical trials.

We also introduced a novel grammar to be learned during the experiment. Isbilen et al.'s (2017) task involved learning the tri-syllabic nonwords, where the transitional probabilities within a "word" was 1, but the transitions between words were more variable (0.33). In Experiments 2.3 and 2.4, we examined whether recall tasks were also effective at measuring learning of between word relationships that are typically found in artificial grammars. However, as the Visual Recall task in Experiments 2.2 and 2.3 showed no evidence of learning, it is possible that recall tasks are only effective measures of implicit statistical learning of highly regular relationships. Therefore, in Experiment 2.4, we developed a novel artificial grammar that was specifically designed to contain elements that can be chunked together to investigate if Visual Recall tasks are able to measure implicit statistical learning of less variable regularities. Finally, in Experiments 2.4 and 2.5, we used randomly generated sequences containing many illegal transitions as ungrammatical sequences, as opposed to ungrammatical sequences containing more subtle violations to the grammar. Isbilen et al.'s (2017) SICR task found differences in recall accuracy between grammatical and random nonword triplets, where the difference in TP between the grammatical and random sequences was much larger than in our Experiments 2.2 and 2.3. By including ungrammatical sequences with considerably lower TPs than the grammatical sequences, we aimed to emphasise the legal transitions and facilitate learning.

We predicted predict 3 key findings in the Visual recall task: recall accuracy would improve across Recall Blocks; recall accuracy would be significant higher in the final Recall Block compared to the Testing Block and finally, recall accuracy would be significantly higher in the Recovery Block compared to the Testing Block. Similarly to Experiments 2.2 and 2.3, following the Visual Recall task, participants completed Grammaticality Judgement, Sequence Generation and Sequence Completion tasks, where we predict similar findings to Experiments 2.2 and 2.3.

Methods

Participants

30 participants (23 female, 7 male; mean age = 26.83), were recruited using the Newcastle University Neuroscience Participant Pool and the School of Psychology Student Participant Pool. Previous studies using serial recall tasks to a measure implicit statistical learning report effect sizes between 0.4 and 1 (Isbilen et al., 2017, 2020, 2022) from which we predicted an average effect size of approximately 0.7. Based on this expected effect size, we conducted a power analysis, which indicated that 30 participants would be appropriate. All participants were native English speakers, and had normal or corrected-to-normal vision and hearing. Participants were not excluded based on their ability to speak any additional languages. Ethics was approved by the Faculty of Medical Sciences Ethics Committee at Newcastle University.

Experiment 2.4 was carried out in-person in May 2019 - March 2020 and was disrupted due to the COVID-19 lockdown. As such, the task was recoded to allow for online data collection (Experiment 2.5). In Experiment 2.5, 36 participants (17 female, 19 male; mean age = 25.74 years) were recruited for this study from Prolific. This sample size was selected as it was similar to the in-person version of the task and previous experiments (Isbilen et al., 2017, 2020, 2022). As in Experiment 2.3, we pre-screened participants to include native English speakers and to exclude individuals with language disorders. In this experiment we also ensured not to recruit participants who had participated in Experiment 2.3. Participants were not excluded based on their ability to speak any additional languages. An additional 3 participants were excluded from the analysis for failing attention checks present throughout the task. Ethics was approved by Emory University IRB.

Methodological Changes

In an attempt to improve performance in the Visual Recall task, we made a number of methodological changes. First, we changed the artificial grammar that was being learned during the experiment. In Experiments 2.2 and 2.3, we used an artificial grammar consisting of 5 stimuli, containing a wide variety of possible transitions between these stimuli, resulting in sequences containing a range of transitional probabilities. By contrast, previous successful serial recall tasks (and many other implicit statistical learning tasks, e.g. Saffran et al., 1996) have used grammars with more predictable structure, in which certain elements consistently occur in 'chunks', with a transitional probability of 1. Chunking facilitates recall because it is more efficient to chunk elements that always occur together rather than processing each element separately. However, if the elements co-occur with more variability, as in Experiments 2.2 and 2.3, then chunking may not occur. It is possible that previous serial recall experiments have successfully measured implicit statistical learning because the artificial grammars being learned contained more predictable transitions. Therefore, we used a new artificial grammar (see *Artifical Grammar*) that contained more predictable transitions. However, between word transitions remain an important feature of natural language, and therefore we designed this novel artificial grammar to contain both highly predictable chunks, and more variable between chunk transitions.

Second, we made changes to the stimuli used across the tasks. In Experiments 2.2 and 2.3, participants recalled 5-element-long sequences of abstract shapes. However, as the novel artificial grammar consisted of several two-element-long chunks, we could only generate sequences that were an even number in length to avoid violating the rules of the grammar. Therefore, we chose to generate 6-element-long sequences in Experiments 2.4 and 2.5. However, performance in the Visual Recall task in Experiments 2.2 and 2.3 was relatively poor with 5-element-long sequences, and therefore we were concerned that creating 6-element-long sequences of abstract shapes may result in floor effects. As such in Experiments 2.4 and 2.5, we decided to use pictures of animals instead of abstract shapes, as we expected memory of familiar objects to be better than abstract shapes.

Finally, we changed the procedure of the Visual Recall task. In Experiments 2.2 and 2.3, each block of the Visual Recall task contained both grammatical and ungrammatical sequences. However, it is possible that including ungrammatical sequences may have interfered with learning of the grammatical transitions. Therefore, in Experiments 2.4 and 2.5, we changed the Visual Recall task to use a blocked design, similar to traditional SRT tasks. In these experiments, participants complete many blocks of grammatical sequences only, before completing a 'testing' block containing only ungrammatical sequences, before completing a final grammatical 'recovery' block. By adopting this new design we aimed to avoid any interference from ungrammatical sequences in the learning of the structure.

Artificial Grammar

As previously mentioned, it is possible that recall tasks are not an effective measure of implicit statistical learning when the grammar being learned contains relatively variable transitions between stimuli. Therefore, we designed a novel grammar that would be more susceptible to chunking (hereafter *Chunking grammar*; see Figure. 2.8). The Chunking grammar consists of 8 elements (A-H), pairs of which always co-occur (i.e., AB, CD, EF, GH), resulting in 4 chunks. The transitional probabilities between the elements in a chunk are 1.0, (i.e., element 'A' can only be followed by element 'B'), and the elements within a chunk can only appear in this order. In addition to these within chunk relationships, there are also relationships between the 4 chunks. There are two types of possible grammatical transitions between chunks: one high TP transition (e.g., AB→CD; TP = 0.66) and one low TP transition (e.g., AB→GH; TP = 0.33).

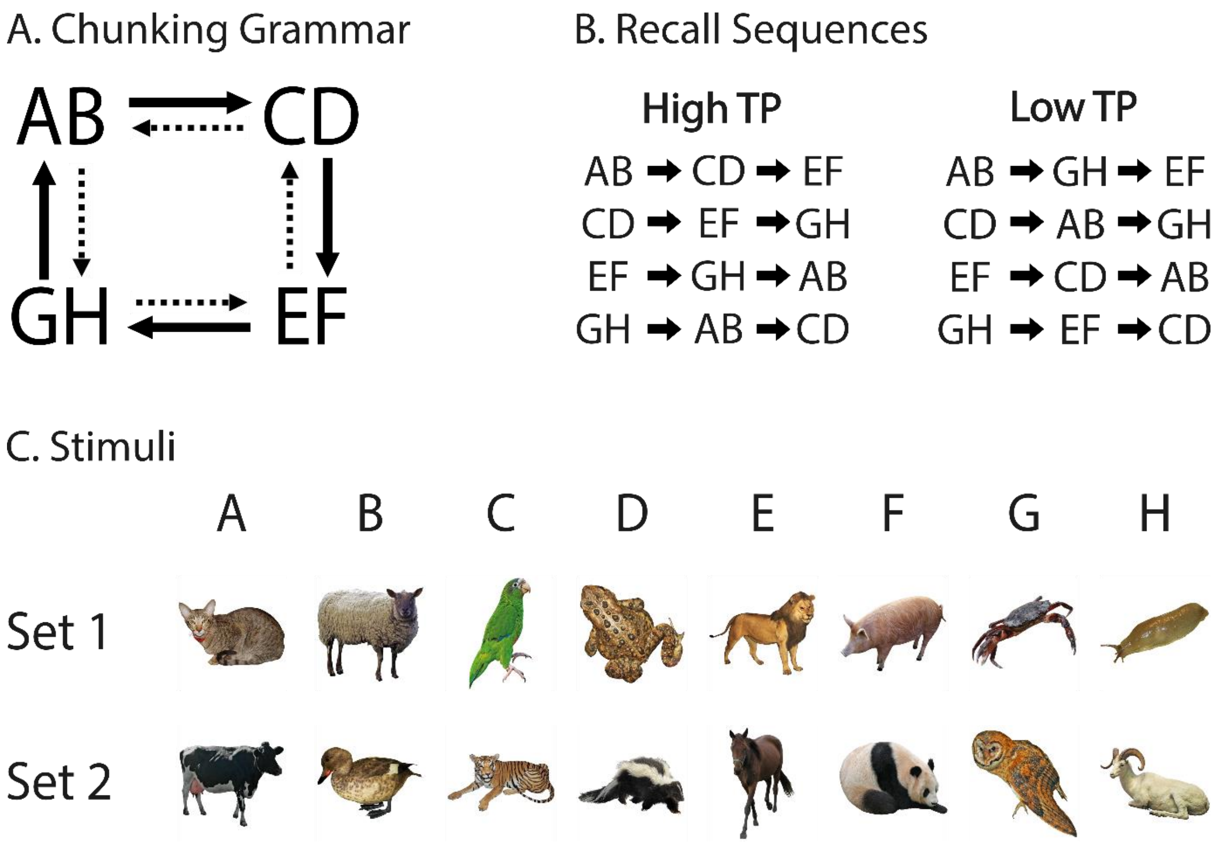


Figure 2.8. Chunking grammar and stimuli. A) Illustration of the artificial grammar used in Experiments 2.4 and 2.5. Sequences are generated by following the arrows to produce 6-element long sequences. B) The recall sequences were grammatical sequences. To establish the high and low TP sequences, the high TP sequences were repeated twice per block, whereas the low TP sequences were only presented once per block, totalling 12 sequences per block. C) The grammar contains 8 elements which are represented by images of animals. A different set of animals was used in the initial random block to familiarise participants with the task without affecting their learning of the stimuli or the grammar.

The structure of the grammar allows us to examine the learning of the different types of relationships underlying the grammar separately: we can investigate whether the within chunk relationships are learned (e.g., the relationship between ‘A’ and ‘B’), if the between chunk relationships are learned (e.g., that ‘AB’ is more likely to be followed by ‘CD’ than ‘GH’, and is never followed by ‘EF’), or if both of these relationships are learned. As the transitional probabilities within elements in a chunk are considerably higher than the TPs between chunks, we would predict that the within chunk relationships will be learned before the between chunk relationships, if the between chunk relationships are learned at all.

Stimuli

The stimuli in this task consisted of two sets of 8 images of animals (Figure 2.8), each 200x200 pixels. The first set was used in the initial training block, whereas the second set was used for the remainder of the experiment. In the Visual Recall task, in order to create high TP and low TP grammatical sequences, 6-element-long high TP sequences (TP = .66) were presented twice per grammatical block. 4 low TP sequences (TP = .33) were presented once per block, resulting in 12 sequences per grammatical block. 12 randomly generated sequences were presented in the random blocks.

Because the Visual Recall task was comparing grammatical with random sequences, in order to further examine types of rules participants may pick up on, for the subsequent explicit tasks we designed ungrammatical sequences with more specific violations (similarly to the ungrammatical testing sequences from Experiments 2.2 and 2.3). For example, if a participant has correctly learned that certain elements are chunked together, they should perform better at judging sequences with violations to these chunks as ungrammatical. For the Grammaticality Judgement task, a total of 48 6-element long sequences were created, half of which were grammatical. There are only 8 possible 6-element long grammatical sequences that do not include repetition of chunks within a sequence (the same grammatical sequences from the Visual Recall task). These 8 sequences were repeated 3 times to equal the number of ungrammatical sequences. The ungrammatical sequences contained equal numbers of three types of violation: within chunk violations (e.g., AB C**F** GH), between chunk violations (e.g., AB CD **GH**) and both within and between chunk violations (e.g., AB **DC** EF). The sequences were balanced so that violations were equally likely to occur in each position and with each element (Table 2.1).

Table 2.2. Experiment 2.4 and 2.5 Grammaticality Judgement task sequences. Grammatical, within violation, between violation, and both within and between chunk violation sequences used in the Grammaticality Judgement task. Illegal transitions are shown in bold. To balance the number of grammatical and ungrammatical sequences, the grammatical sequences were repeated three times.

	Grammatical Sequences	Sequences With Within Chunk Violations	Sequences With Between Chunk Violations	Sequences With Both Violations
High TP Sequences	AB CD EF	AB C**F** GH	AB CD **GH**	AB **DC** EF
	CD EF GH	CD E**H** AB	EF GH **CD**	CD **FE** GH
	EF GH AB	EF G**B** CD	CD AB **EF**	EF **HG** AB
	GH AB CD	GH A**D** EF	GH EF **AB**	GH **BA** CD
Low TP Sequences	AB GH EF	AB G**F** CD	CD GH **AB**	AB **HG** EF
	CD AB GH	CD A**H** EF	GH CD **EF**	CD **BA** GH
	EF CD AB	EF C**B** GH	AB EF **CD**	EF **DC** AB
	GH EF CD	GH E**D** AB	EF AB **GH**	GH **FE** CD

Similarly to the Grammaticality Judgement task, the Sequence Completion task can provide more detailed information regarding the types of relationships that have been learned. Depending on the location of the gaps in each sequence, knowledge of the different relationships between elements can be tested (see Table 2.2). Eight sequences were included to test knowledge of each type (within chunk, between chunk and both within and between chunk) of relationship, totalling 24 sequences in the Sequence Completion task.

Table 2.3. Experiment 2.4 and 2.5 Sequence Completion sequences. These sequences were used to test knowledge of within, between and both within and between chunk relationships.

	Within Chunk Sequences	Between Chunk Sequences	Both Sequences *
High TP Sequences	A_ C_ E_	AB CD __	AB __ EF
	C_ E_ G_	EF GH __	CD __ GH
	_F _H _B	__ EF GH	EF __ AB
	_H _B _D	__ AB CD	GH __ CD
Low TP Sequences	G_ E_ C_	__ GH EF	__ CD __
	E_ C_ A_	__ CD AB	__ EF __
	_D _B _H	CD AB __	__ GH __
	_B _H _F	GH EF __	__ AB __

**Note: there are no high or low sequences for the "both" condition, both high and low TP sequences can be made in these trials depending on the elements that the participant choses.*

Procedure

In Experiment 2.4, the lab set up and procedure were identical to Experiment 2.2: the same cognitive tests were administered, and participants completed the Visual Recall, Grammaticality Judgement, Sequence Generation and Sequence Completion tasks in the same order as Experiment 2.2. In Experiment 2.5, because the experiment took place online, participants completed the computer tasks on their own desktop computer or laptop, and no standardised cognitive or language tasks were completed. In both in-person and online computer tasks, 6-element-long sequences were presented serially across the computer screen. Each image was displayed on the screen for 450ms, before being removed. The inter-stimulus interval was 300ms.

Visual Recall Task

The recall task consisted of two types of blocks: recall blocks, where the sequences were structured based on the Chunking Grammar, and random blocks, where the elements in the sequences were randomly shuffled. The recall task consisted of 9 blocks, with each block containing 12 sequences. The participants were not aware of the transitions between blocks. In each block, after the sequence was presented, there was a 1000ms retention period. Following this, the elements were presented simultaneously on the screen (see Figure 2.9). The position of each element was randomised on each trial, so that participants could not rely on positional cues or motor sequence learning. The participant was asked to recreate the sequence by clicking on the appropriate elements in the correct order. No feedback was given. An inter-trial interval of 1500ms separated the participant's response from the presentation of the next sequence.

The task began with an initial random block, using the first set of stimuli, was included to familiarise participants with the task without affecting their learning of either the stimuli or the grammar. The second set of stimuli were then used for the remainder of the experiment. The initial random block was followed by 6 blocks of sequences structured in accordance with the chunking grammar (Recall Blocks), followed by a Testing Block consisting of random sequences. The final Recovery Block consisted of sequences in which the structure was restored. In the online version of the task used in Experiment 2.5, participants were offered two opportunities to take a short break during the visual recall task, between Recall Blocks 2 and 3 and Recall Blocks 5 and 6.

Grammaticality Judgement Task

The Grammaticality Judgement task was identical to Experiments 2.2 and 2.3.

Sequence Generation Task

In this task participants were instructed to create their own 6-element-long sequences. In each of the 8 trials, the participants were presented with all 8 elements arranged in a circle on the screen (Figure 2.9). Participants created their sequences by clicking on stimuli in the desired order. Each trial was separated by a 1500ms inter-trial interval. No feedback was given.

Sequence Completion Task

In the Sequence Completion task, participants were presented with incomplete 6-element-long sequences, where the missing elements were replaced by question marks. As in the Sequence Generation task, all 8 possible elements were also arranged in a circle on the screen. In each of the 24 trials, participants were asked to fill in the gaps in the sequence by clicking on the stimuli in the desired order.

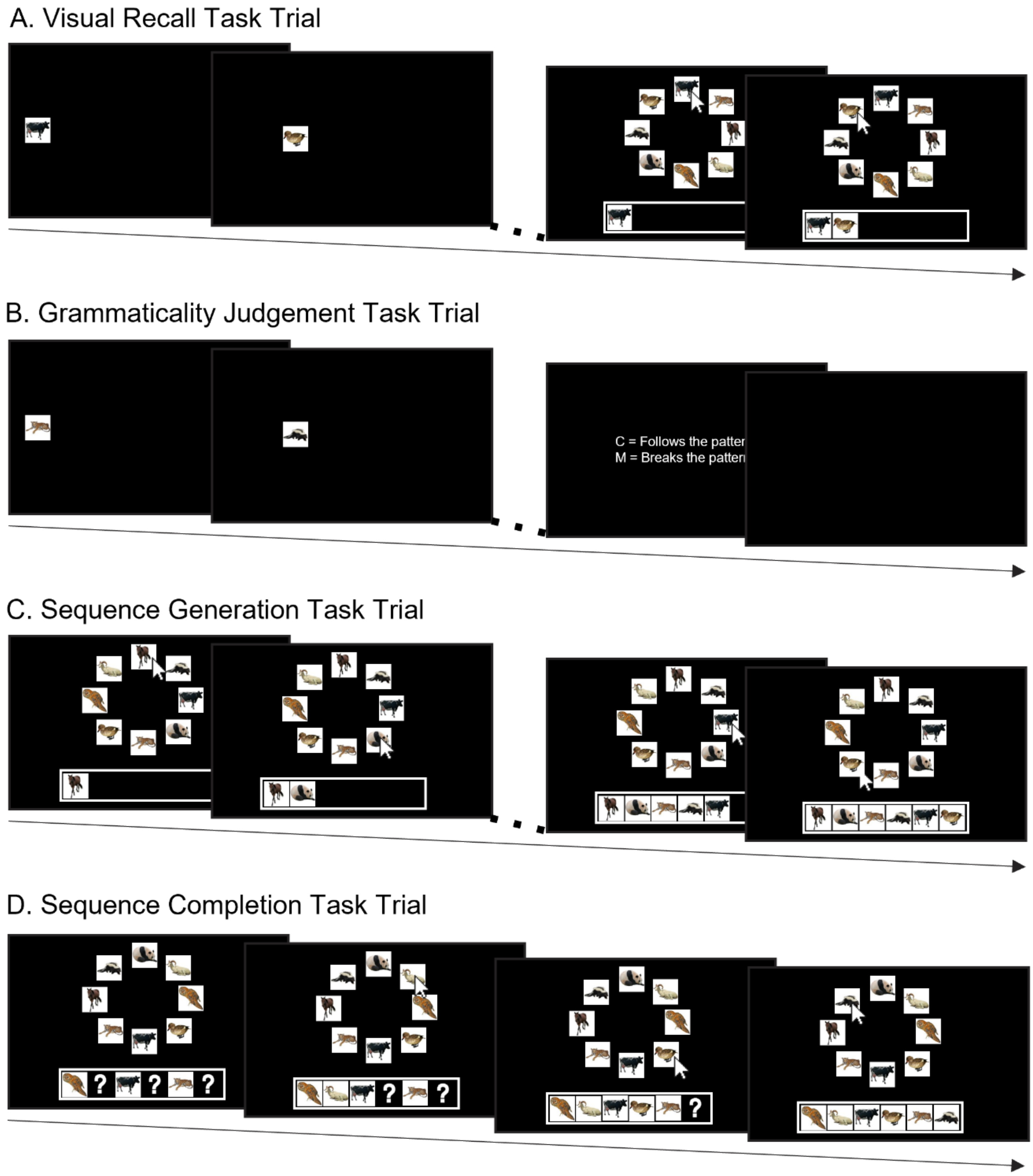


Figure 2.9. Experiments 2.4 and 2.5 trial design. A) Visual Recall task. In each trial, participants were presented with a 6-element-long sequence: each element was displayed on screen for 450ms before being removed, with an inter-stimulus interval of 300ms. After the sequence had been displayed, there was a 1000ms retention period. Following this, participants were presented with all 8 possible stimuli simultaneously on the screen. Participants were asked to recreate the sequence by clicking on the desired elements in order. Each trial was separated by a 1500ms inter-trial interval. B) Grammaticality Judgment task trial. Participants were presented with grammatical and ungrammatical sequences and then pressed one of two keys on the keyboard to indicate whether they felt that the sequence followed the same pattern as the sequence they had seen previously or not. Each trial was separated by a 1500ms inter-trial interval. C) Sequence Generation task. Participants were presented with all 8 possible elements arranged in a circle on the screen. Participants created their own 6-element-long sequences by clicking on the desired elements in order. Each trial was separated by a 1500ms inter-trial interval. D) Sequence Completion task. In each of the 24 trials, participants were presented

with all 8 possible elements and were a partially completed sequence, where some elements were replaced by question marks. Participants were tasked with filling in the gaps in these sequences by clicking on the desired stimuli in order. Each trial was separated by a 1500ms inter-trial interval.

Data Analysis

In Experiments 2.4 and 2.5, the data were analysed in the same way as Experiments 2.2 and 2.3. In both experiments, we also compared performance of high and low transitional probability sequences across the Visual Recall, Grammaticality Judgement, Sequence Generation and Sequence Completion tasks. As Experiment 2.5 was completed online, to ensure participants were completing the task properly, we included several quality checks prior to analysis. If a participant created two or more sequences within a block that consisted of only one element repeatedly, then they were excluded from the analysis. We also used participants responses times to ensure quality: for incorrectly recalled sequences, if participants took longer than 10 seconds to begin recreating the sequence, or if they took a break of longer than 5 seconds whilst recreating the sequence then this sequence was flagged. Participants who had more than three flags per block were excluded from the analysis.

Results

The results from Experiments 2.4 and 2.5 provide strong evidence that learning had occurred in both the Visual Recall tasks and subsequent more explicit tasks. In-person, when using absolute correct scores, recall accuracy was significantly higher in Recall Block 6 was than in Recall Block 1 ($t_{29} = 6.570$, $p < .001$). Recall accuracy was also significantly higher in Recall Block 6 compared to the Testing Block ($t_{29} = 3.916$, $p = .001$). Finally, recall accuracy significantly improved between the Testing and Recovery Blocks ($t_{29} = 4.094$, $p < .001$). This pattern was also reflected in the proportion correct scores: recall accuracy was significantly higher in Recall Block 6 compared to Recall Block 1 ($t_{29} = -7.53$, $p < .001$); Recall accuracy was significantly higher in Recall Block 6 compared to the Testing Block ($t_{29} = 4.48$, $p = .001$). Finally, recall accuracy significantly improved between the Testing and Recovery Blocks ($t_{29} = -4.659$, $p < .001$). This same pattern was observed in the online experiment when using both absolute (Recall Block 6 > Recall Block 1: $t_{35} = 4.72$, $p < .001$; Recall Block 6 > Testing Block: $t_{35} = 5.04$, $p < .001$; Recovery Block > Testing Block: $t_{35} = 4.85$, $p < .001$) and proportion correct (Recall Block 6 > Recall Block 1: $t_{35} = 4.01$, $p < .001$; Recall Block 6 > Testing Block: $t_{35} = 4.86$, $p < .001$; Recovery Block > Testing Block: $t_{35} = 5.12$, $p < .001$) scores. These findings strongly suggest that implicit statistical learning has occurred during the Visual Recall task, and that the Visual Recall task is an effective measure of this learning.

Table 2.4. Descriptive Statistics for the Visual Recall Task in Experiments 2.4 and 2.5.

	Experiment 2.4 (In-Person)				Experiment 2.5 (Online)			
	Absolute Correct		Proportion Correct		Absolute Correct		Proportion Correct	
	Mean	SEM	Mean	SEM	Mean	SEM	Mean	SEM
Baseline	0.20	0.03	0.58	0.03	0.26	0.03	0.62	0.03
Recall 1	0.23	0.04	0.62	0.03	0.26	0.04	0.63	0.03
Recall 2	0.32	0.06	0.66	0.04	0.34	0.05	0.69	0.03
Recall 3	0.43	0.06	0.73	0.04	0.38	0.05	0.71	0.03
Recall 4	0.44	0.07	0.76	0.04	0.41	0.06	0.74	0.03
Recall 5	0.49	0.07	0.78	0.04	0.46	0.06	0.75	0.04
Recall 6	0.53	0.06	0.79	0.04	0.55	0.06	0.79	0.04
Testing	0.29	0.05	0.66	0.03	0.30	0.04	0.66	0.03
Recovery	0.52	0.07	0.79	0.04	0.51	0.06	0.78	0.03

Participants also showed evidence of learning in the Grammaticality Judgement task, both in-person (M = 0.59, SEM = 0.02; t_{29}= 4.13, p < .001) and online (M = 0.59, SEM = 0.02; t_{35} = 3.58, p = .001). Performance on the Grammaticality Judgment task can also be broken down by condition to compare performance across grammatical and ungrammatical sequences containing within chunk, between chunk and both within and between chunk violations (see Table 2.1). In the in-person experiment (Experiment 2.4), participants perform above chance on the grammatical sequences only (M = 0.66, SEM = 0.04; t_{29} = 4.211, p < .001), but not in any of the ungrammatical conditions (Within: M = 0.56, SEM = 0.05; t_{29}= 1.191, p = .243; Between: M = 0.42, SEM = 0.05; t_{29} = -1.561, p = .129; Both: M = 0.57, SEM = 0.05; t_{29} = 1.426, p = .165). In the online experiment (Experiment 2.5), participants performed above chance for the Grammatical condition (M = 0.69, SEM = 0.05; t_{35} = 4.05, p < .001), and significantly below chance in the Between condition (M = 0.33, SEM = 0.05; t_{35} = -3.48, p = .001). Participant's performance in the Within and Both conditions did not differ from chance (Within: M = 0.58, SEM = 0.05; t_{35} = 1.55, p = .131; Both: M = 0.57, SEM = 0.05; t_{35} = 1.22, p = .232). Repeated measure ANOVAs with Condition (Grammatical, Within, Between or Both) as a within-subjects factor indicated that there was a main effect of Condition in the in-person experiment ($F_{3,87}$ = 4.14, p = .016), with post-hoc (Bonferroni corrected) tests indicating significant differences in performance between the Grammatical and Between conditions (p = .045), but not between any other conditions. There was also a main effect of condition in the online experiment ($F_{3,105}$ = 9.03, p < .001), with post-hoc (Bonferroni corrected) tests indicating significant differences in performance in the Between condition compared to the Grammatical (p = .002), Within (p = .013) and Both (p = .031) conditions, but not between other conditions (p > 0.05, in all cases).

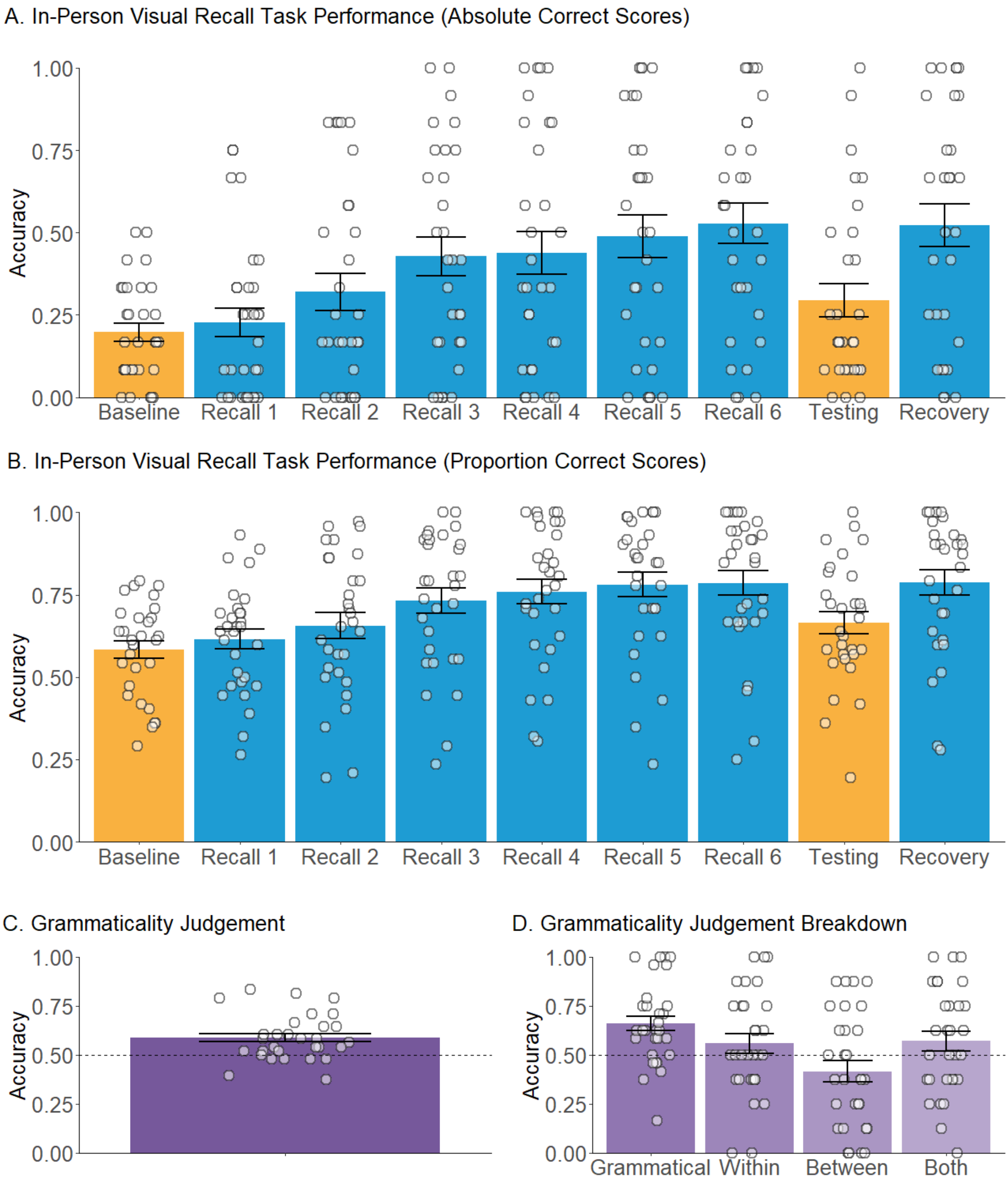


Figure 2.10. Experiment 2.4 in-person Visual Recall and Grammaticality Judgement task performance. A) Mean recall (+/- SEM) accuracy in the Visual Recall task, based on absolute correct scores. Orange bars indicate blocks of random sequences, whereas blue bars indicate blocks of grammatical sequences. Recall increases across grammatical Recall Blocks, before significantly decreasing in the Testing Block. In the final Recovery Block, performance is significantly higher than in the Testing Block. This pattern of results provides clear evidence that implicit statistical learning has taken place. B) The same pattern was observed when using proportion correct scores. C) Performance on the Grammaticality Judgement task showed above chance performance (indicated by the dashed line). Individual performance is shown as white circles. D) Breakdown of performance across the different types of sequences in the Grammaticality Judgement task. Participants performance above chance (shown by the dashed line) at correctly classifying the grammatical sequences only.

Performance on both the Sequence Generation and Sequence Completion tasks was positively correlated with performance on the Grammaticality Judgement task in the in-person (Sequence Generation: $r = .512$, $p = .004$; Sequence Completion: $r = .624$, $p < .001$) and online (Sequence Generation $r = .695$, $p < .001$; Sequence Completion: $r = .775$, $p < .001$) experiments, which may indicate that participants who perform better in the Grammaticality Judgement task have some explicit access to the knowledge of the structure that is needed to create their own sequences. Performance on the Sequence Generation task was also positively correlated with performance on the Sequence Completion task in both in-person ($r = .647$, $p < .001$) and online experiments ($r = .724$, $p < .001$).

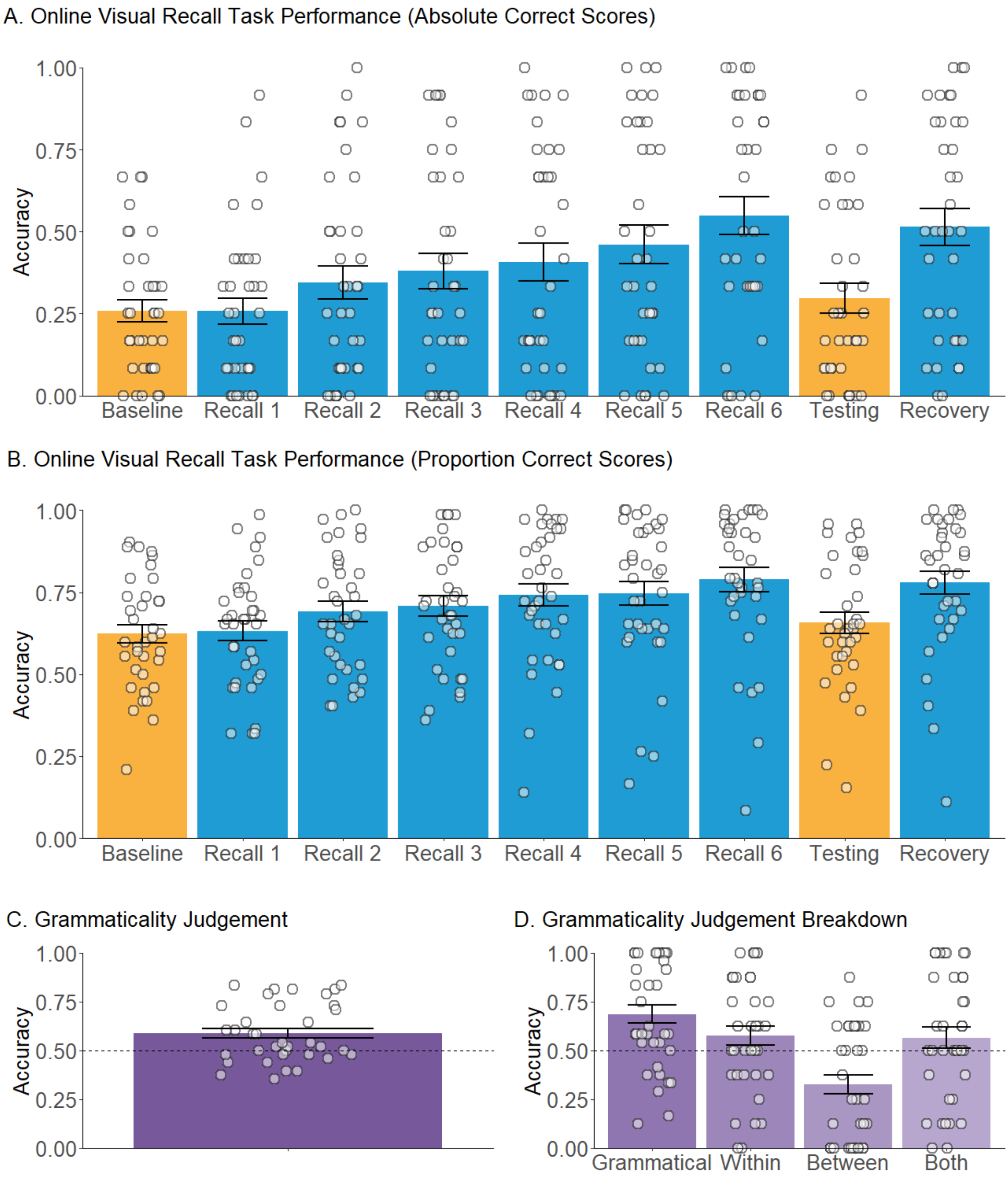


Figure 2.11. Experiment 2.5 online Visual Recall and Grammaticality Judgement task performance. A) Mean recall (+/- SEM) accuracy in the Visual Recall task, based on absolute correct scores. Orange bars indicate blocks of random sequences, whereas blue bars indicate blocks of grammatical sequences. Recall increases across grammatical Recall Blocks, before significantly decreasing in the Testing Block. In the final Recovery Block, performance is significantly higher than in the Testing Block. This pattern of results provides clear evidence that implicit statistical learning has taken place. B) The same pattern was observed when using proportion correct scores. C) Performance on the Grammaticality Judgement task showed above chance performance (indicated by the dashed line). Individual performance is shown as white circles. D) Breakdown of performance across the different types of sequences in the Grammaticality Judgement task. Participants performance above chance (shown by the dashed line) at correctly classifying the grammatical sequences only.

In order to investigate the extent to which participants gained knowledge of the differing transitional probabilities of the between chunk transitions, where possible we compared performance between high and low transitional probability sequences in each of the tasks. We conducted a 2x7 repeated measures ANOVA with within subject factors of run (only runs containing grammatical sequences, and therefore both high and low TP sequences, were included in this analysis, resulting in 7 runs in total) and the TP of the sequence (high and low TP) to determine whether participant's recall of high TP sequences was improved compared to low TP sequences. In Experiment 2.4 we found no differences in recall accuracy of high compared to low transitional probability sequences across the Visual Recall task when using absolute correct scores ($F_{1,29} = 3.61$, $p = .067$). There was a significant main effect of run ($F_{6,174} = 15.73$, $p < .001$), with Bonferroni corrected post-hoc tests indicating significant differences in recall accuracy between Recall Block 1 and all other blocks ($p < .05$) except Recall Block 2 ($p = .309$), and between Recall Block 2 and Recall Blocks 5 and 6 and the Recovery Block ($p < .05$). There were no other significant differences in recall accuracy between blocks ($p > .05$ in all cases). There was no interaction between Run and the TP of the sequence ($F_{6,174} = 1.06$, $p = .387$).

There was no significant difference in recall accuracy of high and low TP sequence when using proportion correct scores ($F_{1,29} = 0.80$, $p = .378$). There was a significant main effect of run ($F_{6,174} = 23.50$, $p < .001$), with Bonferroni corrected post hoc tests indicating significant differences in recall accuracy between Recall Block 1 and all other blocks ($p < .05$) except Recall Block 2 ($p = 1.00$), and between Recall Block 2 and all subsequent blocks ($p < .05$). There was no significant difference in recall accuracy between the other runs ($p > .05$ in all cases). There was no significant interaction between Run and the TP of the sequence ($F_{6,174} = .022$, $p = .883$).

In Experiment 2.5 we also found no difference in recall accuracy for high and low TP sequences in the Visual Recall task when using absolute correct scores ($F_{1,35} = 0.22$, $p = .644$). There was a main effect of run ($F_{6,210} = 9.05$, $p < .001$). Bonferroni corrected post hoc tests indicated that there were significant differences in recall accuracy between the Recovery Block and other Recall Blocks ($p < .05$) except Recall Blocks 5 and 6 ($p > .05$). There were also differences in recall accuracy between Recall Block 6 and Recall Blocks 1, 2, 3 and 4 ($p < .05$). There were no other significant differences in recall accuracy between blocks ($p > .05$ in all cases). There was no interaction between Run and the TP of the sequence ($F_{6,210} = 1.55$, $p = .143$).

We found similar results when using proportion correct scores: there was no difference in recall accuracy for high and low TP sequences ($F_{1,35}$ = 1.27, p = .267). There was a main effect of Run ($F_{6,210}$ = 6.45, p < .001). Post hoc tests (Bonferroni corrected) indicated significant differences in recall accuracy between Recall Block 1 and Recall Block 6 and the Recovery Block only (p < .05). There were no other significant differences in recall accuracy between blocks (p > .05 in all cases). There was no interaction between Run and the TP of the sequence ($F_{6,210}$ = 1.93, p = .078).

In Experiment 2.4, conducted in-person, we found no difference in performance for high and low TP sequences in the Grammaticality Judgement task (t_{35} = 0.76, p = 0.455). However in Experiment 2.5, conducted online, participants performed significantly better at classifying the high TP grammatical sequences compared to the low TP grammatical sequences (t_{35} = 3.02, p = 0.005), which suggests that participants completing the online task learned more detailed information regarding the frequency of the between chunk transitions.

This pattern was similar in the Sequence Generation task: in Experiment 2.4, a repeated measures ANOVA (with transition type: high TP, low TP and illegal) showed that participants did not differ in the mean number of high TP, low TP, or illegal between chunk transitions that they created per sequence ($F_{2,58}$ = 2.45, p = 0.095). However, in Experiment 2.5, there was a significant difference in the mean number of times each transition type was created per sequence ($F_{2,70}$ = 5.35, p = 0.007). Post-hoc tests (Bonferroni corrected) indicated that participants generated significantly more high probability transitions than low probability transitions (p = 0.034) or illegal transitions (p = 0.011). However, there were no differences in the mean number of low probability transitions made compared to violation transitions (p > .05).

In the Sequence Completion task, the Within and Between chunk sequences were designed to include sequences with both high and low probability between chunk transitions (see Table 2.3). Furthermore, sequences in the Both condition were designed to test knowledge of Both Within and Between chunk transitions (see Table 2.3) and allowed participants to create either high or low transitional probability sequences. In both Experiments 2.4 and 2.5, there was no difference in performance for high and low TP sequences testing knowledge of Within chunk transitions (in-person: t_{29} = 0.53, p = 0.602; online: t_{35} = 1.64, p = 0.110). However, participants showed improved performance on high TP sequences testing knowledge of Between chunk transitions compared to low TP sequences (in-person: t_{29} = 2.65, p = 0.013; online: t_{35} = 2.57, p = 0.015), and were more likely to complete sequence testing knowledge of Both Within and

Between chunk transitions with a high TP chunk compared to a low TP chunk (in-person: t_{29} = 2.57, p = 0.016; online: t_{35} = 2.63, p = 0.013).

Taken together, these findings suggest that in the Visual Recall task, participants are not aware of the more subtle variations in the transitional probabilities of the grammatical sequences. However, in the reflection-based tasks, there is some evidence that participants have gained improved knowledge of the varying transitional probabilities within the grammatical sequences (e.g., that high TP sequences are more common than low TP sequences). This may be due to increased familiarity with the grammar, as participants complete the reflection-based tasks later in the experiment. Additionally, when the participants are completing the reflection-based tasks, they are told that there are rules underlying the grammar, and this awareness may encourage participants to seek out rules they may otherwise not have gained awareness of.

Similarly to Experiments 2.2 and 2.3, we calculated composite measures of Visual Recall task by averaging performance across the 3 key indicators of learning (the difference between Recall Block 6 and Recall Block 1; the difference between Recall Block 6 and the Testing Block; the difference between the Recovery Block and the Testing Block), meaning that participants with more positive composite Visual Recall task scores show more learning across this task that participants with negative composite scores. Composite explicit task performance was calculated based on the mean performance in the Grammaticality Judgement, Sequence Generation, and Sequence Completion tasks. These measures were positively correlated in both the in-person (r = .485, p = .006) and online (r = .603, p < .001) experiments, which may indicate that the processing-based Visual Recall task and subsequent reflection-based tasks are measuring similar processes. As in Experiment 2.2, in Experiment 2.4 we also examined whether performance in the Visual Recall task and subsequent explicit tasks was correlated with performance on the battery of language tasks. However, we did not find any consistent correlations between performance on the language tasks and composite performance in the Visual Recall or reflection-based tasks (see Appendix 2.3.)

To compare performance between the in-person and online versions of the task, we conducted two mixed 2x9 ANOVAs based on either absolute or proportion correct scores in the Recall task. In both ANOVAs we included run (9 runs) as a within-subjects factor and task (in-person or online) as a between-subjects factor. When using absolute correct scores, we found a significant main effect of run ($F_{3.7,241.6}$ = 24.505, p < .001). Bonferroni corrected post hoc tests indicated that there were significant differences in recall accuracy between the Baseline Block and all other blocks (p < .05) except the Testing Block (p > .05). Additionally, there were further significant differences in recall accuracy between Recall Block 1 and all other blocks

($p < .05$) except the Testing Block. There were significant differences between Recall Block 2 and Recall Block 5 and 6 only ($p < .05$). There were significant differences in performance between Recall Block 6 and all Recall Blocks ($p < .05$) except Recall Block 5 ($p < .05$), and between the Testing Block and Recall Blocks 4, 5, 6 and the Recovery Block ($p < .05$). Finally, there were significant differences in recall accuracy between the Recovery Block and all other blocks ($p < .05$) except for Recall Blocks 5 and 6 ($p > .05$). There were no other significant differences between blocks ($p > .05$ in all cases). There was no main effect of task ($F_{1, 64}$ = .002, $p = .967$) or interaction between run and task ($F_{3.7,241.6}$ = .568, $p = .676$), indicating that there was no difference in performance between participants completing the in-person and online versions of the task.

These findings were also reflected in the proportion correct scores: we found a significant main effect of run ($F_{4.2, 270}$ = 26.191, $p < .001$), with Bonferroni corrected post hoc tests indicated that there were significant differences in recall accuracy between the Baseline Block and all other blocks ($p < .05$) except for Recall Block 1, Recall Block 2 and the Testing Block ($p > .05$). Additionally, there were further significant differences in recall accuracy between Recall Block 1 and all other blocks ($p < .05$) except the Baseline block, Recall block 2 and the Testing Block ($p > .05$). There were further significant differences between Recall Block 2 and Recall Blocks 4, 5 and 6, as well as the Testing Block ($p < .05$), and between Recall Block 3 and Recall Block 6, the Testing Block and the Recovery Block ($p < .05$). Finally, there were significant differences between the Testing Block and all other blocks ($p < .05$) except the Baseline Block and Recall Blocks 1, 2 and 3 ($p > .05$). There were no other significant differences between blocks ($p > .05$ in all cases). There was no main effect of task ($F_{1, 64}$ = .001, $p = .990$) or interaction between run and task ($F_{4.2, 270}$ = .875, $p = .484$).

We also compared performance between in-person and online versions of the Grammaticality Judgement, Sequence Generation and Sequence Completion tasks using *t*-tests. We found no significant difference between performance on the in-person and online versions of the Grammaticality Judgement (t_{64} = .007, $p = .995$) and Sequence Completion Tasks (t_{64} = -1.22, $p = .227$). However, participants completing the online version of the Sequence Generation tasks performed significantly better than those completing the tasks in-person (t_{64} = -2.62, $p = .011$).

Discussion

In Experiments 2.4 and 2.5, we found evidence of learning in both the processing-based Visual Recall task and across subsequent reflection-based measures, indicating that visual serial recall is an effective measure of implicit statistical learning. Positive correlations between the processing- and reflection-based measures suggest that these tasks may be measuring similar mechanisms. We found no difference in performance between Experiment 2.4 (in-person) and Experiment 2.5 (online), which suggests that performance across the Visual Recall and subsequent reflection-based measures is not affected by completing the task online as opposed to in-person. The learning effect found in the Visual Recall task also shows that the efficacy of recall paradigms to assess learning extends beyond the use of nonword stimuli and therefore, the Visual Recall task could be an appropriate measure to investigate the proposed deficits in implicit statistical learning in individuals with language difficulties (for a review, see Schmalz et al., 2017), where nonword stimuli would not be appropriate.

Previous studies have shown that serial recall paradigms, both in the auditory and visual domain, can effectively measure implicit statistical learning (Isbilen et al., 2022; Isbilen et al., 2017, 2020; Kidd et al., 2020). The findings from Experiments 2.4 and 2.5 provide further support for using serial visual recall as a processing-based measure of learning. Although some previous recall tasks have found no correlations between performance on recall tasks and reflection-based tasks (Isbilen et al., 2017), the fact that performance correlated across these tasks may suggest that in Experiments 2.4 and 2.5, participants have gained some explicit knowledge of the structure which aids in explicit decision-making processes during the reflection-based tasks. Some studies using visual stimuli have previously reported positive correlations between processing- and reflection-based tasks (Dale et al., 2012; Isbilen et al., 2020), which may suggest that serial recall in the visual domain requires more reflection than in the auditory domain (Isbilen et al., 2020). Despite findings evidence of learning across both the Visual Recall and more explicit tasks, we found no reliable correlations between task performance and performance across the standardised cognitive and language tasks. However, implicit statistical learning is thought to play a role in many facets of language learning and processing (Conway & Christiansen, 2005; Turk-Browne et al., 2005), including word segmentation (Saffran et al., 1996) and syntax acquisition (Gómez & Gerken, 2000). Although there is less research specifically investigating individual differences in implicit statistical learning and variations in language ability, previous studies have found evidence of correlations between performance on implicit statistical learning tasks and language (Conway et al., 2010; Evans et al., 2009; Isbilen et al., 2022; Kidd, 2012; Misyak & Christiansen, 2012). However, questions have been raised regarding the replicability of these findings: two recent studies found no correlation between implicit statistical learning and language ability in children and adults (Schmalz et al., 2019; West et al., 2018). Further studies have noted small correlations (Spencer et al., 2015) or correlations only when using specific implicit statistical learning tasks (Qi et al., 2019). These mixed findings have previously been attributed to the lack of reliability across measures of implicit statistical learning (Siegelman, 2020), which may explain the lack of consistency within the literature.

In Experiments 2.2 and 2.3, we found no evidence of learning in the Visual Recall task, therefore in Experiments 2.4 and 2.5 we made a number of changes to the design of the task. We utilised a blocked design whereby ungrammatical sequences were not interspersed throughout Recall Blocks to ensure learning of the regularities was not disrupted. We also developed a novel artificial grammar designed to facilitate chunking of elements, to whether the lack of learning in Experiments 2.2 and 2.3 was due to the variability within the artificial grammar and generated ungrammatical sequences containing random transitions rather than ungrammatical sequences with subtle violations to ensure that chunking could not be used to aid recall of ungrammatical as well as grammatical sequences. Given that in Experiments 2.4 and 2.5 we found evidence of learning in the Visual Recall task, it is clear that at least one of these changes was important to ensure learning was reflected in improved recall for grammatical over ungrammatical sequences. However, based on the data from Experiments 2.4 and 2.5, we cannot conclude which of these design changes were necessary for the Visual Recall task to successfully measure learning. Therefore, in Experiment 2.6 we aimed to see if the Visual Recall task from Experiments 2.4 and 2.5 would be able to measure learning the artificial grammar from Experiments 2.2 and 2.3.

Experiment 5: Online Hybrid Recall Task

Introduction

In Experiments 2.2 and 2.3, we found no evidence of learning in the Visual Recall task. However, following several design changes, in Experiments 2.4 and 2.5 the recall task showed strong evidence of learning of a novel artificial grammar containing pairs of stimuli which always co-occurred as chunks. It is not clear from these findings whether the success of the recall task in Experiments 2.4 and 2.5 was due to the changes in design, or due to the novel chunking grammar. In Experiment 2.6, we aimed to assess whether the efficacy of the Visual Recall task extends beyond the learning of grammars that are designed to be susceptible to chunking. To test this, we used the recall task design that successfully measured learning in Experiments 2.4 and 2.5 and used it to measure learning of the more variable artificial grammar from Experiments 2.2 and 2.3. We predicted similar performance across the visual recall task and subsequent reflection-based measures as in the previous experiments.

Methods

Participants

40 participants (20 female, 20 male; mean age = 30.46 years) were recruited from Prolific. This sample size was selected as it was similar to previous experiments (Isbilen et al., 2017, 2020, 2022; Experiments 2.3 and 2.5). As in Experiments 2.3 and 2.5, participants were pre-screened via Prolific to include native English speakers only, and to exclude individuals with language disorders, and participants who has completed Experiments 2.3 or 2.5. Participants were not excluded based on their ability to speak any additional languages. No participants were excluded from this experiment for failing attention checks.

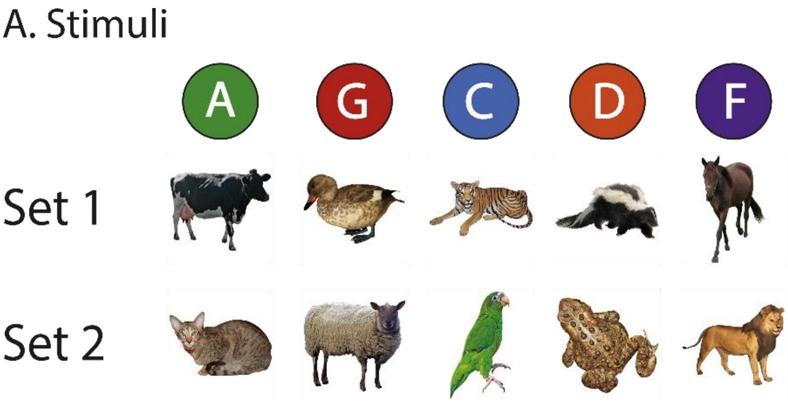
Methodological Changes

In Experiment 2.6, we made some methodological changes based on the results of our previous experiments. In Experiments 2.2 and 2.3, we used Saffran et al.'s (1996) grammar, which contained more variable transitions, and measures learning of this grammar using a Visual Recall task which contained both grammatical and ungrammatical sequences of abstract shapes in every Recall Block. As we did not see evidence of learning in these experiments, in Experiments 2.4 and 2.5, we developed a novel grammar that contained more predictable transitions, and measured recall of sequences of animals using a Visual Recall task with a blocked design. To understand why we found evidence of learning in Experiments 2.4 and 2.5 but not Experiments 2.2 and 2.3, in Experiment 2.6, we measuring learning of the same artificial grammar from Experiments 2.2 and 2.3 (Saffran et al., 1996), but using the blocked design and images of animals from Experiments 2.4 and 2.5. In Experiment 2.6 we aimed to determine whether the success of Experiments 2.4 and 2.5 was due to the use of a more predictable grammar, or due to the various other changes in design.

Stimuli

The grammar was the same as in Experiments 2.2 and 2.3. Each element was represented by an image of an animal drawn from those used in Experiments 2.4 and 2.5 (Figure 2.12. A). Similarly to Experiments 2.4 and 2.5, the first set of images were used for the initial baseline block, and the second set were used for the remainder of the experiment. The grammatical sequences included in the Learning and Recovery Blocks consisted of three 4-element-long and three 6-element-long sequences (Figure 2.12. B). The Baseline and Testing Blocks consisted of randomly generated sequences: three 4 elements long and three 6 elements long. In all blocks, each sequence was presented twice, totalling 12 sequences per block.

The Grammaticality Judgement task was identical to Experiments 2.2 and 2.3, except instead of being split into two runs, separated by an exposure phase, all 32 sequences were presented in one run. The sequences in the Sequence Generation and Sequence Completion tasks were identical to Experiments 2.2 and 2.3.



B. Recall Sequences

Sequence	Mean TP
ACFC	0.52
ACGF	0.60
ADCF	0.64
ADCFCG	0.63
ADCGFC	0.63
ACGFCG	0.60

Figure 2.12. Experiment 2.6 stimuli and recall sequences. A) Stimuli used in Experiment 2.6. As in Experiments 2.4 and 2.5, a different set of animals was used in the initial random block to familiarise participants with the task without affecting their learning of the stimuli or the grammar. B) The recall sequences were 3- and 4-element-long grammatical sequences. In the ungrammatical blocks, the same number of 3- and 4-element-long sequences were randomly generated, and therefore did not follow the rules of the Saffran et al. (2008) grammar. In all blocks, each sequence was presented twice, totalling 12 sequences per block.

Procedure

The procedure was identical to Experiment 2.5.

Data Analysis

The visual recall task was analysed in the same way as in Experiments 2.4 and 2.5, and the Grammaticality Judgement, Sequence Generation and Sequence Completion tasks were analysed in the same way as Experiments 2.2 and 2.3. As this study was conducted online, we applied the same quality checks as in Experiment 2.5.

Results

We found strong evidence of learning in the visual recall task. Based on absolute scores, recall accuracy was significantly higher in Learning block 6 compared to Learning Block 1 (t_{39} = 7.76, $p < .001$), recall accuracy was significantly higher in Learning Block 6 compared to the Testing Block (t_{39} = 6.92, $p < .001$), and recall accuracy was significantly higher in the Recovery Block than the Testing Block (t_{39} = 5.25, $p < .001$). When using proportion correct scores, we find a similar pattern: recall accuracy was significantly higher in Learning Block 6 compared to Learning Block 1 (t_{39} = 6.05, $p < .001$), recall accuracy was significantly higher in Learning Block 6 compared to the Testing Block (t_{39} = 6.19, $p < .001$), and recall accuracy was significantly higher in the Recovery Block compared to the Testing Block (t_{39} = 5.37, $p < .001$). These findings strongly suggest learning has occurred during the recall task, and that visual recall can measure learning of artificial grammars containing more variable transitions.

Table 2.5. Descriptive Statistics for the Visual Recall task in Experiment 2.6.

	Absolute Correct		Proportion Correct	
	Mean	SEM	Mean	SEM
Baseline	0.48	0.03	0.73	0.02
Recall 1	0.57	0.03	0.81	0.02
Recall 2	0.64	0.04	0.84	0.02
Recall 3	0.71	0.04	0.87	0.02
Recall 4	0.74	0.04	0.88	0.02
Recall 5	0.74	0.04	0.89	0.02
Recall 6	0.79	0.03	0.91	0.02
Testing	0.59	0.04	0.80	0.02
Recovery	0.76	0.03	0.90	0.02

Participants also performed significantly above chance (50%) in the Grammaticality Judgement task (*M* = 0.62, SEM = 0.02; $t_{39} = 5.60$, $p < .001$), and above chance (20%) in the Sequence Completion task (*M* = 0.79, SEM = 0.03; $t_{39} = 23.10$, $p < .001$) indicating that the implicitly learned information was available for conscious processing. Performance across the reflection-based tasks were highly correlated: There was a positive correlation between performance in the Grammaticality Judgement task and performance in the Sequence Generation (r = .408, p = .009) and Sequence Completion (r = .321, p = .043) tasks, and a positive correlation between performance in the Sequence Generation and Sequence Completion tasks (r = .434, p = .005). These findings suggest that participants who had learned the grammar during the recall task were able to access this information more explicitly to create their own sequences.

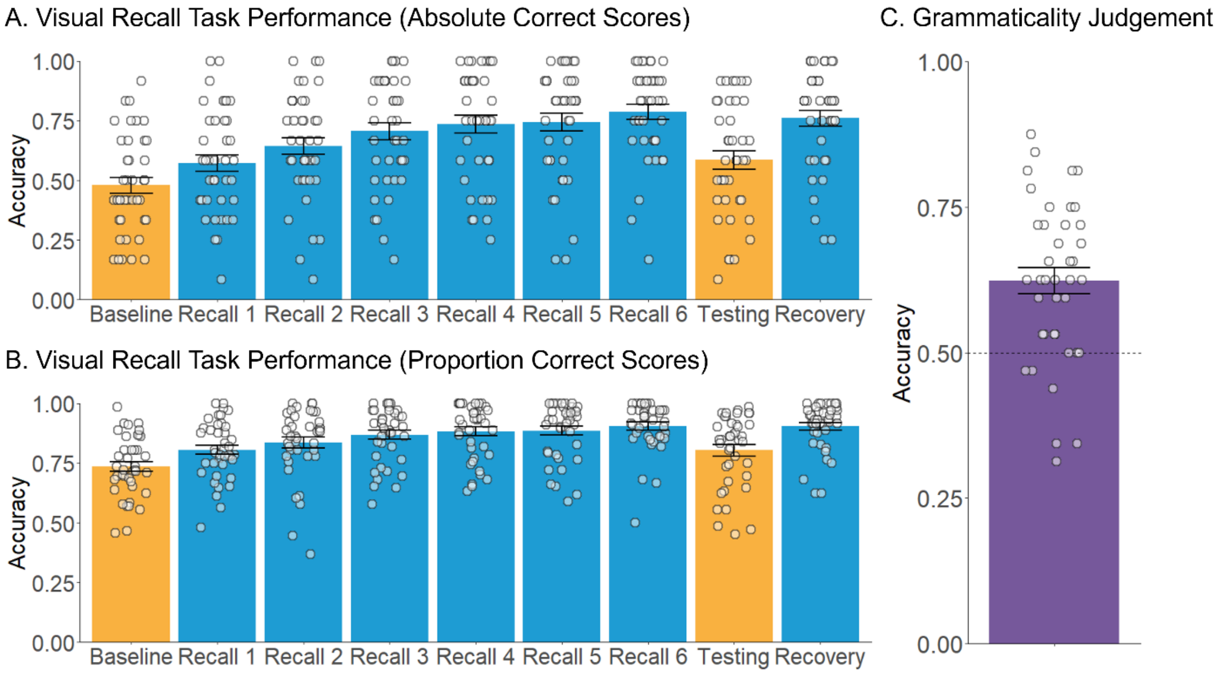


Figure 2.13. Experiment 2.6 Hybrid Visual Recall and Grammaticality Judgement task performance. In all panels, error bars represent +/- 1 SEM. A) Mean recall accuracy in the Hybrid Visual Recall task, based on absolute correct scores. Orange bars indicate blocks of random sequences, whereas blue bars indicate blocks of grammatical sequences. Recall increases across grammatical Recall Blocks, before significantly decreasing in the Testing Block. In the final Recovery Block, performance is significantly higher than in the Testing Block. This pattern of results provides clear evidence that implicit statistical learning has taken place. B) The same pattern was observed when using proportion correct scores. C) Performance on the online Grammaticality Judgement task showed above chance performance (indicated by the dashed line). Individual performance is shown as white circles.

In Experiments 2.4 and 2.5, there was a positive correlation between performance in the Visual Recall task and performance across the subsequent reflection-based tasks. However, in this experiment performance in the reflection-based tasks was not correlated with performance in the Visual Recall task based on either absolute correct (r = .163, p = .315) or proportion correct (r = -.134, p = .408) scores, which may suggest that in this experiment, the tasks are not measuring the same underlying processes.

Discussion

In Experiment 2.6, we aimed to investigate why we found evidence of learning in the Visual Recall task in Experiments 2.4 and 2.5 but not in Experiments 2.2 and 2.3, by using the Visual Recall task from Experiments 2.4 and 2.5 to measure learning of the grammar from Experiments 2.2 and 2.3. We found strong evidence of learning across both the Visual Recall and subsequent reflection-based tasks, which indicates that the Visual Recall task is not limited to measuring learning of highly regular chunks.

In this experiment, unlike Experiments 2.4 and 2.5, we found no correlation between performance in the Visual Recall task and performance across the reflection-based tasks. Previous research has suggested that the lack of correlation between processing- and reflection-based measures suggests that the tasks are measuring different processes, with reflection-based tasks tapping into more explicit-decision making processes as opposed to implicit statistical learning (Isbilen et al., 2017). The findings from Experiment 2.6 suggest that in this case, the Visual Recall and subsequent reflection-based tasks may be capturing different aspects of learning, which we did not find in Experiment 2.5, using a different grammar. This may suggest that the complexity of the grammar affects the extent to which processing- and reflection- based tasks measure similar processing (see General Discussion).

Previous research using serial recall as a measure of implicit statistical learning has demonstrated learning highly regular relationships in both the auditory and visual domains (Isbilen et al., 2022; Isbilen et al., 2017, 2020; Kidd et al., 2020; Experiments 2.4 and 2.5, Chapter 2), however there has been little research examining whether serial recall can effectively measure learning of grammars containing more variable transitions. The findings from Experiment 2.6 demonstrate that recall can provide a more implicit method of measuring more variable transitions as well as highly regular relationships. This finding has relevance for how regularities are learned in both artificial grammars and in natural language. The Visual Recall task relies on the learning of chunks to facilitate recall, and therefore evidence of learning of more variable transitions in this task supports the idea that artificial grammar learning relies at least in part on the learning of chunks (Perruchet & Pacton, 2006). Furthermore, whilst natural languages do contain many highly predictable relationships, the majority of regularities are more variable. As the Visual Recall task can successfully measure learning of more variable transitions, this suggests that in future, serial recall tasks can be used to measure the learning of relationships with varying degrees of predictability, as in natural language.

General Discussion

The aim of these experiments was to develop a processing-based measure of implicit statistical learning and combine it with more traditional reflection-based tasks to investigate the nature of the knowledge acquired during implicit statistical learning. Although we did not effectively measure learning using reaction times in the SRT-AGL task (Experiment 2.1), the findings from the subsequent experiments (Experiments 2.2 – 2.6) suggest that serial visual recall tasks can effectively measure implicit statistical learning of both highly predictable and more variable relationships without requiring conscious reflection or explicit decision-making. However, based on the design changes made between Experiments 2.2 and 2.3 and Experiments 2.4 and 2.5, we can highlight some key factors that affect learning. First, the findings from these experiments suggest that serial visual recall can be used to measure learning of grammars irrespective of the variability of the transitions they contain. Previous research (Isbilen et al., 2022; Isbilen et al., 2017, 2020; Kidd et al., 2020), and the findings from Experiments 2.4 and 2.5 have demonstrated that serial recall can be used to measure learning of highly predictable relationships. The findings from Experiment 2.6 indicate that this extends to the learning of more variable transitions, such as those typically found in artificial grammar learning paradigms. Second, the design of the Visual Recall task plays an important role in whether learning is measuring successfully, with blocked designs consisting of active exposure being more effective than oddball designs. Finally, these experiments suggest that performance does not differ based on whether the task is completed in-person or online, and therefore online artificial grammar learning experiments may provide a useful alternative to in-person testing.

Across these 5 experiments, we find mixed evidence of a correlation between processing- and reflection-based tasks. In particular, we found a positive correlation between these tasks in Experiments 2.4 and 2.5, using the more simple and predictable Chunking Grammar, but not in Experiment 2.6, using the more variable Saffran et al. (2008) grammar. It is possible that the complexity of the grammar affects the processing that occurs during the tasks. When learning a simple grammar (as in Experiments 2.4 and 2.5), participants are more likely to gain some explicit knowledge of the structure during the Visual Recall task, which can be used to aid explicit decision-making processes in the in the reflection-based task, resulting in highly correlated performance across tasks. However, when the grammar contains more variable transitions, and is therefore more complex (as in Experiment 2.6), participants are less likely to gain explicit knowledge of the rules. Therefore, they are less able to rely on this explicit knowledge to aid in decision-making processes in the reflection-based tasks, which results in uncorrelated performance across processing- and reflection-based tasks. Indeed, previous research has shown that if explicit knowledge is available, then participants are more likely to rely on this than implicit knowledge (Batterink, Reber, & Paller, 2015). These findings suggest that the complexity of the grammar may affect the extent to which explicit knowledge of the regularities is acquired during the Visual Recall task, and therefore whether more explicit learning processes are recruited during this task.

The Visual Recall task using a blocked design appears to provide some additional benefits to learning over an oddball design. When using a blocked design, the participant is exposed to grammatical sequences in an active task where they are required to attend and interact with the stimuli, meaning they are constantly engaged throughout the learning process. This is in contrast to typical artificial grammar learning paradigms (including the Visual Recall task from Experiments 2.2 and 2.3 using an oddball design), in which the exposure phase, when learning occurs, is entirely passive. Furthermore, when using oddball designs, following passive exposure, participants completed the Visual Recall task, where they recalled both grammatical and ungrammatical sequences. Implicit statistical learning of regularities occurs because legal transitions occur more frequently than illegal transitions. Therefore, if legal and illegal transitions occur equally frequently, as they did in the Visual Recall tasks with the oddball design, then we may not see differences in recall of grammatical and ungrammatical sequences, particularly if participants are more attentive and engaged in this phase of the task compared to passive exposure. This may suggest that passive exposure is less important than active exposure when both are included in the experiment and explain why we found evidence of learning when using a blocked design but not an oddball design. This also highlights the benefits of using processing-based measures of learning: it is not possible to use active exposure within grammaticality judgement tasks, as it would require all testing sequences to be grammatical, and therefore provide no measure of learning.

Processing-based tasks also provide additional benefits over traditional paradigms in that they allow for the measurement of learning throughout a task, rather than after learning has occurred during an exposure phase. Post-exposure measures of learning, such as grammaticality judgement tasks, can be problematic as they only provide information on *what* is ultimately learned, as opposed to *how* learning occurs (Lammertink et al., 2019; Siegelman, 2020). For example, processing-based measure provide insight into other interesting aspects of learning, such as speed. The findings from the experiments where the Visual Recall task has shown evidence of learning (Experiments 2.4, 2.5 and 2.6) indicate that learning improves consistently across grammatical blocks, suggests that learning is gradual and occurs throughout the Visual Recall task. This provides a benefit over the reflection-based measures: whilst these measures indicate that learning has taken place, they provide little information regarding the speed or trajectory of learning throughout exposure.

Conclusion

These experiments demonstrate that serial visual recall is a valuable approach for measuring implicit statistical learning without requiring conscious reflection and highlights some of the conditions under which this approach is and is not successful. Furthermore, these findings suggest that the complexity of the grammar being learning plays an important role in the processes that underlie performance in these tasks.

Chapter 3: Assessing Implicit Statistical Learning in Children

Abstract

Examining the developmental trajectory of implicit statistical learning is not only important to better understand implicit statistical learning as a phenomenon, but also to provide benchmarks for investigating implicit statistical learning deficits in developmental disorders such as dyslexia. Previous research has typically compared children and adult's performance using reflection-based tasks. These tasks may not accurately reflect implicit statistical learning in children, as they rely on additional cognitive abilities that are likely less developed than in adults. Processing-based measures would provide an intuitive method for measuring implicit statistical learning irrespective of age, without relying on more conscious decision-making processes. In this experiment, we conducted an online study with 89 children aged between 8 and 15 years which aimed to measure implicit statistical learning using the processing-based Visual Recall task we had previously developed and tested with adults. Children showed evidence of implicit statistical learning across the tasks, and there was no difference in performance based on age across the sample of children. There were similarities in performance of children and adults across the Visual Recall task, which may indicate that implicit statistical learning is an age-invariant mechanism. However, the processing-based Visual Recall task provided some suggestion that the time-course of learning may differ between children and adults, which was not captured by traditional reflection-based tasks. These findings highlight the importance of measuring learning using both processing- and reflection-based measures.

Experiment 3.1: Investigating Implicit Statistical Learning in Children Using Processing-Based Measures

Introduction

Implicit statistical learning is critical for language acquisition, and therefore a considerable amount of research has investigated implicit statistical learning in infancy, when the majority of language learning takes place. However, despite this, there is considerably less research investigating implicit statistical learning in older children and adolescents. Furthermore, there are surprisingly few studies investigating the developmental trajectory of implicit statistical learning abilities across development and into adulthood. These studies have found mixed outcomes: there is some suggestion that implicit statistical learning is age-invariant (Raviv & Arnon, 2018; Saffran et al., 1997; Thiessen et al., 2013), while other studies have shown that implicit statistical learning abilities improve alongside other cognitive abilities (Arciuli & Simpson, 2011). It has also been suggested that implicit statistical learning is better in children compared to adults (Jost et al., 2015). It is currently unclear why there are conflicting findings within the literature, although many of the studies use different methods of assessing implicit statistical learning, some of which may not be appropriate for measure learning in children.

Most research investigating implicit statistical learning in children and across development have used artificial grammar learning paradigms that consist of an exposure phase followed by a testing phase, typically a grammaticality judgement task. Although the issues with these traditional reflection-based measures have been highlighted in previous chapters (See General Introduction and Chapter 2), there are limitations that are associated with these tasks that are specific to measuring implicit statistical learning in children. Specifically, it is widely acknowledged that the additional cognitive abilities that these tasks rely on (e.g., understanding task instructions, decision-making skills) improve across development (Lammertink et al., 2019). Furthermore, whilst ‘yes’ biases (in this case, specifically a “grammaticality” bias: a preference for categorising sequences in a grammaticality judgement task as grammatical) are sometimes found in adult samples (e.g., (Dienes et al., 1991), they are more commonly reported in artificial grammar learning tasks with children (Ambridge & Lieven, 2011; Gillis et al., 2022; Lammertink et al., 2020; van der Lely et al., 2011). This may suggest that children may require a lower threshold for classifying sequences as grammatical or pick up on irrelevant features of the sequences as use these to determine grammaticality. For example, children may think that most sequences are grammatical because they contain the same stimuli as in the exposure phase, or the sequences are a similar length, which is unrelated to knowledge of the dependencies between stimuli that grammaticality judgement tasks are meant to be probing. Therefore, when using these paradigms to assess implicit statistical learning in children, it is possible that these tasks do not accurately reflect children's abilities.

Processing-based measures of implicit statistical learning may provide a more intuitive method for measuring learning that is more suitable for children. As previously discussed, (see General Introduction for more detail), processing-based measures typically measure other variables that are facilitated by implicit statistical learning, and therefore these tasks do not require explicit reflection on what has been learned. For this reason, processing-based tasks may allow for more appropriate comparison of implicit statistical learning abilities between adults and children. Processing-based tasks also offer additional advantages over reflection-based tasks, as they can be used to measure learning 'online' - that is, over the course of the task - rather than post-exposure. This provides information about the time-course of learning, which may also differ across development. There is some evidence that processing-based tasks (Cleary et al., 2001; Cleary et al., 2000; Conway et al., 2007), can be used to measure implicit statistical learning in children. However, these tasks often contain a spatial component requiring motor responses, which may involve additional mechanisms than are typically found in implicit statistical learning (Conway, 2005).

There are few studies that have used processing-based measures to assess the developmental trajectory of implicit statistical learning. Bertels et al. (2015) compared the performance of children and adults on both a processing-based task (using reaction times) and a reflection-based task (a forced-choice task). For each sequence in the forced-choice task, participants were also asked to rate their confidence in their choice, as above chance performance in the absence of increased confidence is typically taken as additional evidence that implicit, as opposed to explicit, learning has occurred (Chan, 1991; Dienes et al., 1995). In this experiment, children and adults both showed learning across both the processing- and reflection-based tasks (Bertels et al., 2015). However, although group level performance was above chance in the reflection-based measure, approximately half of both the children and adults did not perform above chance, despite evidence of learning in these participants in the processing-based task. This indicates that processing-based measures may be a more accurate measure of implicit statistical learning irrespective of age. Although both children and adults showed evidence of implicit statistical learning, there were some differences in the nature of the knowledge acquired. While adults performed above chance in the reflection-based task even when their confidence was low (indicating implicit knowledge), children showed no such relationship, which may indicate that reflection-based tasks may not accurately represent the implicit statistical learning abilities of children.

Auditory serial recall tasks that were previously used as a processing-based measure of learning in adults (Isbilen et al., 2017) have since been shown to measure learning in children aged 5 to 7 years. Kidd et al. (2020) showed that similarly to adults, children showed improved verbal recall of predictable nonword sequences over unpredictable sequences. Performance in this task was weakly correlated with performance in a reflection-based measure of learning, which the authors suggest reflects a difference in the precision of how processing- and reflection-based tasks measure knowledge acquired during the experiment. However, although these tasks demonstrate that serial recall can effectively measure implicit statistical learning in children, such tasks have yet to be used to assess learning over a broader range of ages, in order to assess the developmental trajectory of implicit statistical learning across development. Furthermore, it is not clear whether the efficacy of serial recall tasks in children extend beyond auditory phonological stimuli, which would be required to create a task that is suitable for measuring learning in children with language difficulties.

In this experiment, we aimed to use the same Visual Recall task we have previously used to measure learning in adults (Experiment 2.5) to measure learning in children aged 8 to 15 years. As the tasks used to measure learning in adults and children were identical, we then aimed to compare the performance of children to the previously recruited sample of adults (Experiment 2.5). We predicted that we would see evidence of learning across both the processing-based Visual recall task and subsequent reflection-based tasks. We also predicted that we would see no difference in implicit statistical learning in the Visual Recall task based on age, but if reflection-based tasks do not provide an accurate reflection of learning in children, then adults would show improved performance in the reflection-based tasks. Based on the findings of previous experiments, we predicted that there would be positive correlations between performance in the Visual Recall task and the reflection-based measures, and that performance across the reflection-based measures would be positively correlated.

Methods

Participants

89 native English speaking children aged between 8 and 15 years (44 female, 45 male; mean age = 12.08, see Appendix 3.1. for more details) were recruited to complete the online version of the Visual Recall task using the Chunking grammar. Prior to the pandemic, we aimed to recruit children in schools within 2 age groups: children aged between 8 and 11 years (primary

school), and children aged between 12 and 15 years (secondary school). We wanted to recruit children between these ages as we were interested in investigating whether implicit statistical learning abilities improved across development similarly to language skills (Rowe, 1992). We aimed to recruit 40 children per group, to provide a similar sample size to previous experiments (Experiments 2.3, 2.5 and 2.6). Due to the pandemic, we altered the recruitment strategy to instead recruit as many children as possible between the ages of 8 and 15 years. The age demographics of the children can be found in Appendix 3.2. We recruited children aged 8 years or older to ensure they would be able to understand the task instructions in an online environment, without the researcher present. The children were not pre-screened based on their ability to speak any additional languages. The study was advertised to parents on social media and through school newsletters. Parents who were interested in their child completing the study completed a form on Qualtrics which involved providing their name and email address, along with their child's name and date of birth. Parents were then emailed a link to the study which directed their child to complete on a desktop or laptop computer in a quiet area, free from distractions. An additional 14 children completed the experiment but were excluded from the analysis for failing attention checks.

Stimuli

The artificial grammar and stimuli were the same as in Experiments 2.4 and 2.5.

Procedure

At the start of the experiment, the child was asked if they were happy to complete the computer tasks; if not, then the experiment was aborted, although no children declined to take part. The tasks were identical to those from Experiment 2.5: The Visual Recall, Grammaticality Judgement, Sequence Generation and Sequence Completion tasks had previously been coded in PsychoPy (version 2021.2.3) and were completed online through Pavlovia. Similar to Experiment 2.5, the children were offered two opportunities to take a short break during the recall task, between Recall Blocks 2 and 3 and Recall Blocks 5 and 6.

Data Analysis

As this experiment was completed online, to ensure the children were completing the task properly we included the same quality checks prior to analysis as in Experiment 2.5. Furthermore, the data from the Visual Recall, Grammaticality Judgement, Sequence Generation and Sequence Completion task were analysed in the same way as Experiment 2.5. Although as in previous experiments, we calculated both absolute and proportion correct scores in the Visual Recall task, because absolute correct scores reflect the mean number of trials where the whole sequence was correctly recalled, poorer working memory in children may mean that this method of scoring is less appropriate than using proportion correct scores when comparing the performance of children and adults. Therefore, although we conduct the analysis of the Visual Recall task using both methods of scoring, it is likely that proportion correct scores in the Visual Recall task provide the fairest comparison of implicit statistical learning ability between children and adults. Due to errors in saving the data files online, the Sequence Completion data from 11 children was incomplete, and therefore not included in the Sequence Completion task analysis. We first analysed performance across tasks from all 89 children, and then compared their performance to the performance of adults. However, we found that a large number of children (30 out of 89) responded to all 48 trials with the same response in the Grammaticality Judgement task (i.e., exhibited a very strong grammaticality bias). Therefore, we removed the data from these children and re-ran the analysis.

Results

Learning occurred in the Visual Recall and subsequent reflection-based measures: when using absolute correct scores, recall accuracy was significantly higher in Recall Block 6 compared to the Testing Block (t_{88} = 4.60, p < .001), and significantly higher in the Recovery Block than the Testing Block (t_{88} = 3.74, p < .001). However, we found no improvement in recall accuracy between Recall Block 1 and Recall Block 6 (t_{88} = 1.06, p = .290). This pattern was reflected in the proportion correct scores: recall accuracy was significantly higher in Recall Block 6 compared to the Testing Block (t_{88} = 4.822, p < .001), and higher in the Recovery Block than the Testing Block (t_{88} = 3.98, p < .001), although there was no improvement in recall accuracy between Recall Block 1 and Recall Block 6 (t_{88} = .304, p = .762). As there was no improvement in recall accuracy across the Learning Blocks, these results differ from our previous experiments (see Experiments 2.4, 2.5 and 2.6). The children's lack of improvement across the Visual Recall task may be due to the bimodality of the data in this task (see Figure 3.1. A. and B.). Unlike adult's performance in previous experiments, children's performance is highly consistent across the Visual Recall task: children who show high levels of recall accuracy do so from the beginning of the task, whereas those who do not do not show any improvement across the task. However, given that at the end of the task recall accuracy is significantly higher for predictable sequences over unpredictable sequences, there is still strong evidence that learning has taken place during the Visual Recall task, and therefore that visual serial recall can effectively measure implicit statistical learning in children.

Table 3.1. Descriptive Statistics for the Visual Recall task in Experiment 3.1.

	Absolute Correct		Proportion Correct	
	Mean	SEM	Mean	SEM
Baseline	0.12	0.03	0.30	0.03
Recall 1	0.20	0.03	0.35	0.03
Recall 2	0.21	0.04	0.35	0.03
Recall 3	0.21	0.04	0.35	0.04
Recall 4	0.20	0.04	0.34	0.04
Recall 5	0.20	0.04	0.33	0.03
Recall 6	0.23	0.04	0.35	0.04
Testing	0.09	0.02	0.26	0.03
Recovery	0.19	0.03	0.33	0.03

The children performed significantly above chance (50%) in the Grammaticality Judgement task (M = 0.55, SEM = 0.01; t_{88} = 3.97, p < .001; Figure 3.1. C). However, this above chance performance was strongly driven by good performance in classifying the grammatical sequences only (M = 0.90, SEM = 0.02; t_{88} = 24.08, p < .001; Figure 3.1. D.). Performance across the three ungrammatical conditions was significantly below chance (within: M = 0.23, SEM = 0.03; t_{88} = - 8.44, p < .001; between: M = 0.12, SEM = 0.02; t_{88} = - 16.79, p < .001; both: M = 0.25, SEM = 0.03; t_{88} = - 7.51, p < .001). These findings indicate a bias for classifying sequences as grammatical. Performance was correlated across the reflection-based measures: Performance in the Grammaticality Judgement was positively correlated with performance in both the Sequence Generation (r = .626, p < .001) and Sequence Completion tasks (r = .559, p < .001), which suggests that children that perform well in the Grammaticality Judgement task have some explicit knowledge of the regularities, which is needed to create their own sequences. Performance on the Sequence Generation and Sequence Completion tasks were also positively correlated (r = .428, p < .001). Taken together, these findings provide further evidence that learning has occurred during the experiment and suggests that some explicit knowledge of the structure has been acquired.

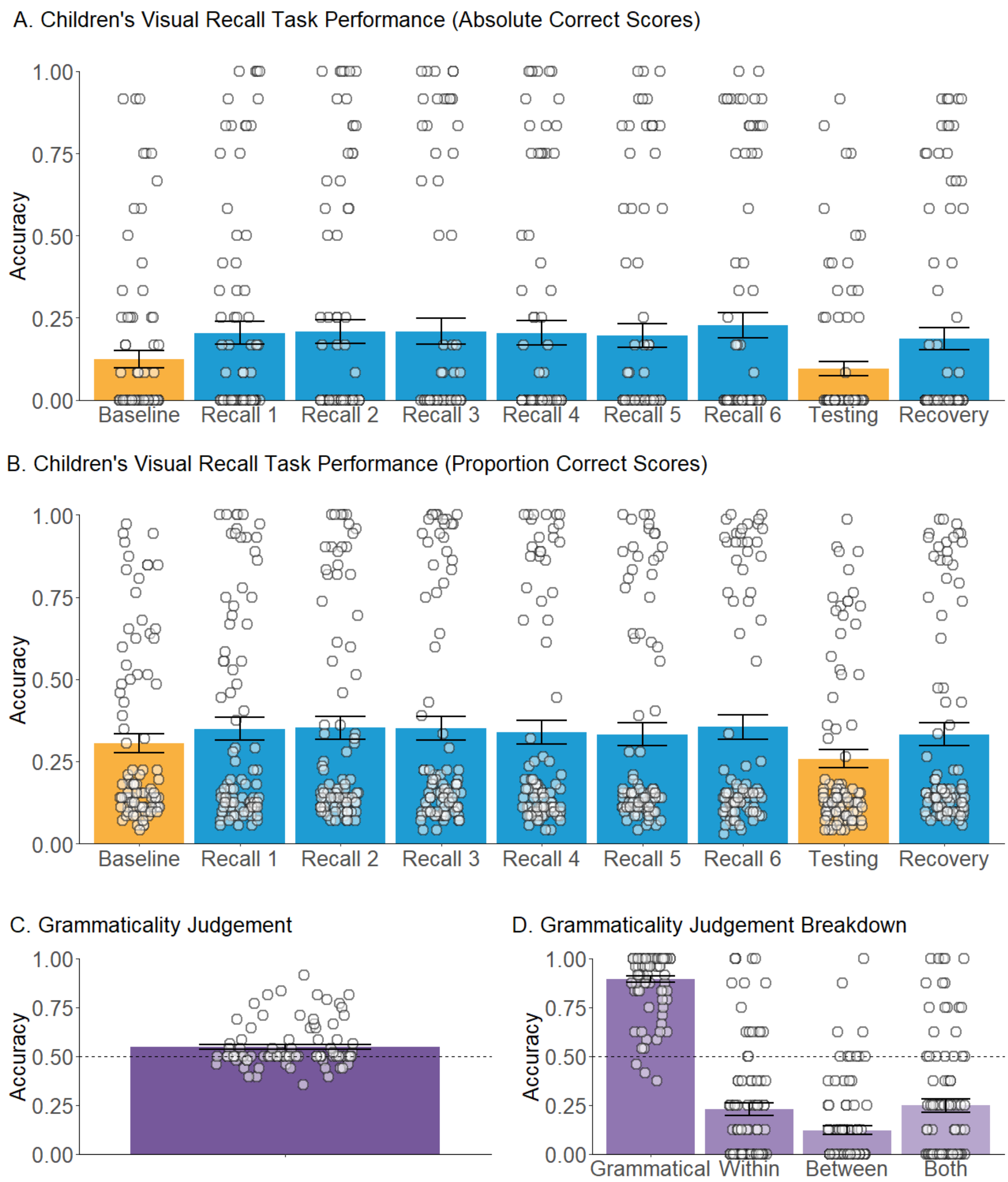


Figure 3.1. Children's Visual Recall and Grammaticality Judgement task performance. A) Mean recall (+/- SEM) accuracy in the Visual Recall task, based on absolute correct scores. Individual performance is shown as white circles. Orange bars indicate blocks of random sequences, whereas blue bars indicate blocks of grammatical sequences. While recall does not increase across grammatical Recall Blocks, there is a significant decrease in recall accuracy between the final Recall Block and the Testing Block. In the final Recovery Block, performance is significantly higher than in the Testing Block. This pattern of results provides clear evidence that implicit statistical learning has taken place. B) The same pattern of performance was observed during the Visual Recall task when using proportion correct scores. C) Children showed above chance performance in the Grammaticality Judgement task (indicated by the dashed line). D) Breakdown of performance across the different types of sequences in the Grammaticality Judgement task. Participants performance above chance (indicated by the dashed line) at correctly classifying the grammatical sequences only and performed below chance at classifying the ungrammatical sequences.

As in previous experiments, we calculated composite measures of learning across the Visual Recall task and subsequent reflection-based tasks. There was a positive correlation between performance in the Visual Recall task and performance across the reflection-based tasks (r = .523, p < .001), which suggests that, similarly to Experiments 2.4 and 2.5 with adult populations, in this experiment the Visual Recall and subsequent reflection-based measures may tap into the same mechanisms.

To assess the developmental trajectory of implicit statistical learning, we first determined whether there were age-related variations in implicit statistical learning across children. There was no correlation between the age of the children and composite performance across the Visual Recall task when using absolute correct scores (r = .063, p = .556; Figure 3.2. A.) or proportion correct scores (r = .065, p = .545; Figure 3.2. B.), or between age and composite performance across the reflection-based tasks (r = -.035, p = .744; Figure 3.2. C.). This indicates that implicit statistical learning did not differ based on age in the sample of children.

To assess whether this extends into adulthood, we also compared children and adult's performance across the Visual Recall task (using absolute and proportion correct scores) with the online data from an adult population (Experiment 2.5). We may expect to see some differences between children and adults in both the time-course of learning and in the amount of learning that has occurred at the end of the Visual Recall task. Therefore, to examine age-related differences in the time-course of learning, we first compared children and adult's performance across Recall Blocks 1 to 6 of Visual Recall task using a mixed 2x6 ANOVA, with age group (child or adult) as a between subjects factor, and run (Recall Blocks 1 to 6) as a within-subjects factor. When using absolute correct scores we found a significant main effect of age group ($F_{1,123}$ = 97.05, p = .002), with adults showing improved recall accuracy across the Recall Blocks compared to children, which is unsurprising given that working memory is thought to be better in adults compared to children (Gathercole et al., 2004). There was also a main effect of run ($F_{5,615}$ = 13.01, p < .001), with post-hoc (Bonferroni corrected) tests indicating significant difference in recall accuracy between Recall Block 6 and the previous Recall Blocks (Recall Block 1: p = .003; Recall Block 2: p = .044; Recall Block 3: p = .069; Recall Block 4: p = .008; Recall Block 5: p = .006), but no differences between other blocks. There was an interaction between run and age group ($F_{5,615}$ = 10.40, p < .001), indicating that there is a greater improvement in performance for adults than for children over the course of the learning period. This may indicate that there are differences in implicit statistical learning between children and adults.

These results were also reflected when using proportion correct scores: there was a main effect of age group ($F_{1,123}$ = 42.80, p < .001) and run ($F_{5,615}$ = 6.26, p < .001), with post-hoc (Bonferroni corrected) tests showed significant differences in recall accuracy between Recall Block 6 and Recall Block 5 (p = .043) only. There was also an interaction between age group and run ($F_{5,615}$ = 6.98, p < .001), further suggesting that there may be age-related differences in the trajectory of learning.

To examine any differences in learning at the end of the Visual Recall task, we then compared performance across the final three blocks using a mixed 2x3 ANOVA, with age group (child or adult) as a between subjects factor, and run (Recall Block 6, and Testing and Recovery Blocks) as a within-subjects factor. Using absolute correct scores, there was a main effect of age group ($F_{1,123}$ = 27.03, p < .001), indicating that similarly to the previous blocks in the Visual Recall task, adults showed improved recall accuracy across the final three blocks compared to children. There was also a main effect of run ($F_{2,246}$ = 36.84, p < .001); post-hoc (Bonferroni corrected) tests indicated that performance in the Testing Block was significantly poorer than in Recall Block 6 (p < .001) and the Recovery Block (p < .001), however there was no difference in performance between Recall Block 6 and the Recovery Block (p = .066). There was also an interaction between age group and run ($F_{2,246}$ = 4.44, p = .013), indicating that there was a more pronounced learning effect in adults compared to children and suggesting that adults implicit statistical learning was better than children's in the Visual Recall task. When using proportion correct scores, there was a similar main effect of age group ($F_{1,123}$ = 62.92, p < .001) and run ($F_{2,246}$ = 29.95, p < .001), with post-hoc (Bonferroni corrected) tests again showing significant differences in recall accuracy between the Testing Block and Recall Block 6 (p < .001) and the Recovery Block (p < .001), but no differences between Recall Block 6 and the Recovery Block (p = .364). There was no interaction between age group and run ($F_{2,246}$ = 1.19, p = .306). This may suggest that when using proportion correct scores, there may not be any difference between children and adults in the final amount of implicit statistical learning that has taken place.

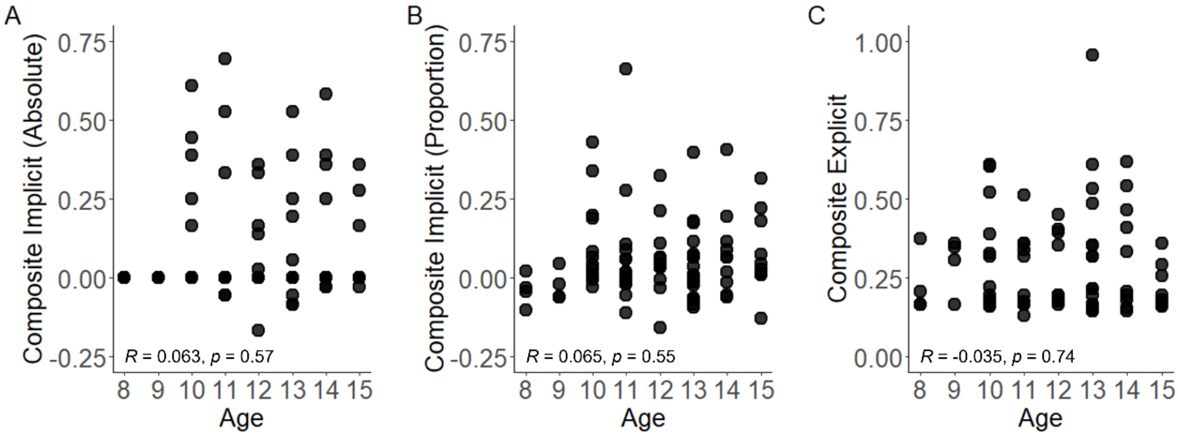


Figure 3.2. Correlations between age and performance in the Visual Recall and reflection-based tasks. A) There was no correlation between age and performance in the Visual Recall task based on absolute correct scores. B) There was no correlation between age and performance in the Visual Recall task based on proportion correct scores. C) There was no correlation between age and performance across the reflection-based tasks.

As previously highlighted, there was evidence that children exhibited a bias for categorising sequences as grammatical in the Grammaticality Judgement task; in fact, 30 out of 89 children categorised all 48 Grammaticality Judgement trials as grammatical. As children were informed at the start of the Grammaticality Judgement task that they would see sequences that both follow and break the pattern, we can conclude that these children were either unmotivated to complete the task accurately or did not understand the task instructions. If the data across all tasks are reanalysed without these children who showed pronounced bias, there is still strong evidence that children show learning across the tasks (see Appendix 3.2.). To compare the performance of children who do not show a pronounced bias with adults, we repeated both the mixed ANOVAs assessing differences between non-biased children and adults in the time-course of learning and the final amount of implicit statistical learning in the Visual Recall task.

When assessing differences in the time-course between non-biased children and adults, when using absolute scores there was no main effect of age group ($F_{1,93} = 1.99$, $p = .161$) indicating that there was no difference in recall accuracy across the Recall Blocks between the non-biased children and adults when using this method of scoring. There was also a main effect of run ($F_{5,465} = 10.91$, $p < .001$), with post-hoc (Bonferroni corrected) tests indicating significant difference in recall accuracy between Recall Block 6 and the previous Recall Blocks ($p < .05$), but no differences between other blocks. There was an interaction between age group and run ($F_{5,465} = 6.02$, $p < .001$), indicating that adults' recall accuracy improves to a greater extent than the non-biased children across the learning period. When using proportion correct scores, there was a main effect of age group ($F_{1,93} = 18.72$, $p < .001$), indicating that when we consider proportion correct scoring adults showed improved recall accuracy compared to the non-biased children. There was a main effect of run ($F_{5,465} = 6.45$, $p < .001$), and post-hoc (Bonferroni corrected) tests showed significant differences in recall accuracy between Recall Block 6 and Recall Blocks 1 and 5 ($p < .05$), however there were no significant differences between the other Recall Blocks. There was also an interaction between age group and run ($F_{5,465} = 3.94$, $p = .012$), further indicating that adults showed greater improvement across the learning period than the non-biased children, and that there may be differences in implicit statistical learning between adults and children in the learning period.

When comparing performance of non-biased children and adults across the final three blocks of the Visual Recall task, using absolute scoring there was a main effect of age group ($F_{1,93} = 10.35$, $p = .002$), with adults again showing improved recall accuracy over the children who did not show a pronounced bias. There was also a main effect of run ($F_{2,186} = 34.04$, $p < .001$); post-hoc (Bonferroni corrected) tests indicated that performance in the Testing Block was significantly poorer than in Recall Block 6 ($p < .001$) and the Recovery Block ($p < .001$), however there was no difference in performance between Recall Block 6 and the Recovery Block ($p = .086$). There was no significant interaction between age group and run ($F_{2,186} = 1.15$, $p = .320$), which suggests that there are no differences between non-biased children and adults in the amount of implicit statistical learning that has occurred during the Visual Recall task. These findings were also reflected in the proportion correct scores: there was a main effect of age group ($F_{1,93} = 30.59$, $p < .001$) and run ($F_{2,186} = 28.48$, $p < .001$), with post-hoc (Bonferroni corrected) tests again showing significant differences in recall accuracy between the Testing Block and Recall Block 6 and the Recovery Block ($p < .05$), but no differences between Recall Block 6 and the Recovery Block ($p > .05$). There was no significant interaction between age group and run ($F_{2,186} = .271$, $p = .763$). This further suggests that by the end of the Visual Recall task there was no difference in the amount of implicit statistical learning that had occurred between non-biased children and adults.

Taken together, these findings suggest that any differences in implicit statistical learning between children and adults in this experiment may have been driven by a number of children that showed a pronounced bias for classifying sequences as "grammatical" only in the Grammaticality Judgement task, and therefore either did not understand the instructions or were not paying attention throughout the tasks. After removing these biased children from the analysis, we no long see any differences between children and adults in implicit statistical learning towards the end of the Visual Recall task. However, there remains some differences between children and adults in the time-course of learning, with adults showing improvement across the Recall Blocks, but non-biased children showing no such improvement. This may highlight that there are still some differences in the nature of implicit statistical learning across the task, which is only revealed when using processing-based measures.

After reanalysing the reflection-based tasks without children with pronounced bias, we no longer found a difference between children and adults in the Sequence Completion task ($t_{93} = 1.60$, $p = .112$). There was still no difference in performance in the Grammaticality Judgement task based on age ($t_{93} = 0.56$, $p = .579$), and adults still performed significantly better than children in the Sequence Generation task ($t_{93} = 2.74$, $p = .007$). Differences in performance between children and adults in the Sequence Generation task may indicate that children are less able to explicitly access the knowledge they have acquired during the experiment and use this to create their own sequences. We did not find any differences in the performance of children and adults in the Grammaticality Judgement task, which may suggest that this task is not underestimating implicit statistical learning abilities in children. However, as this task measures implicit statistical learning post-exposure, it may not capture any differences between children and adults in the time-course of learning. Indeed, while we may see similarities in the performance of children and adults towards the end of the Visual Recall task, the time-course of learning across the Recall Blocks is different between children and adults, even after excluding children who show a pronounced bias in the Grammaticality Judgement task. This suggests that there may be differences between children and adults in the time-course of learning across the task that is not captured by traditional reflection-based tasks. Furthermore, given that a considerable proportion of our sample of children showed a profound bias in the Grammaticality Judgement task, this suggests that these tasks are not the most suitable measure of implicit statistical learning in children.

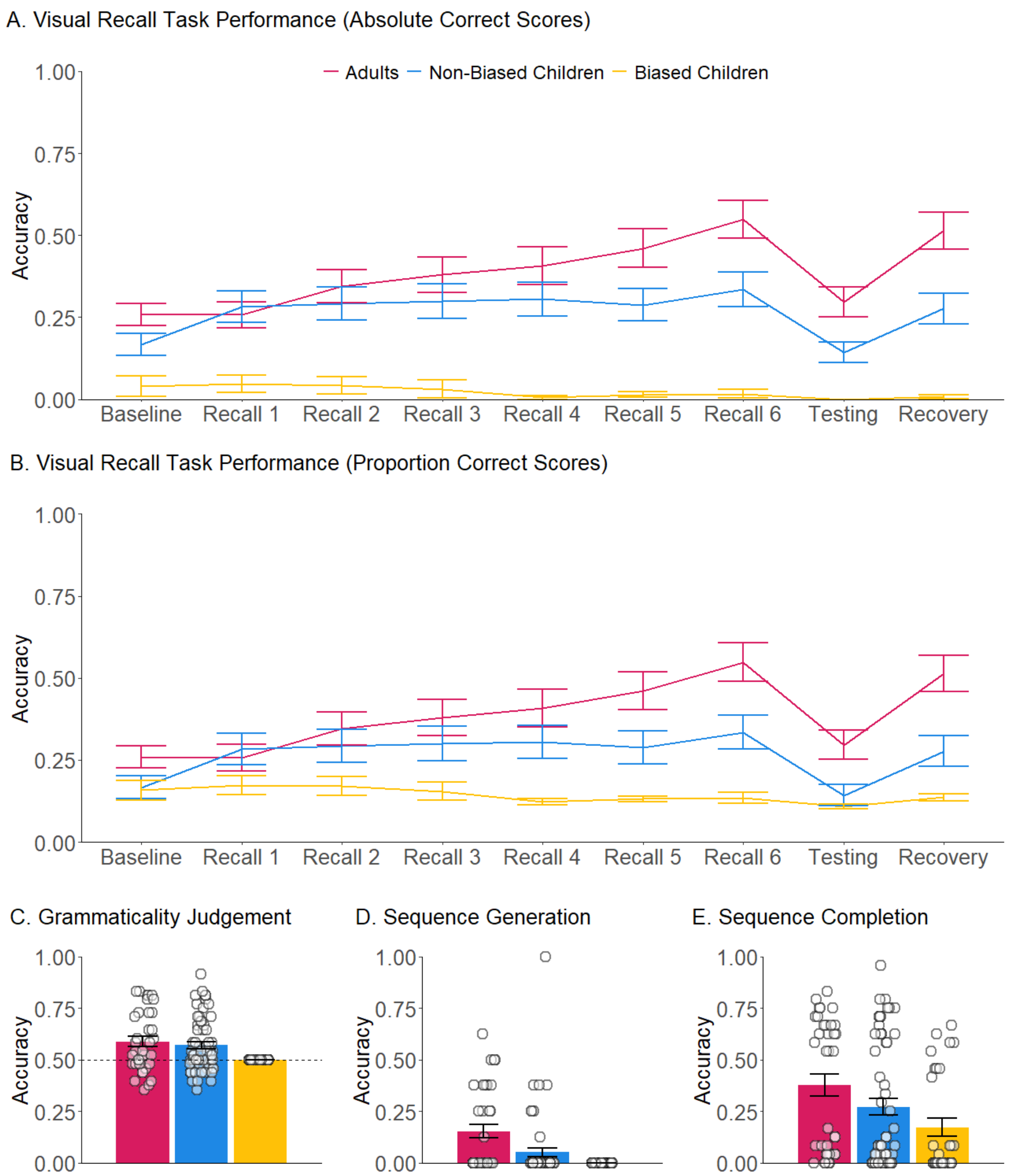


Figure 3.3. Children and adult's Visual Recall Task and reflection-based task performance. A) Mean recall (+/- SEM) accuracy of children who show bias (yellow), children who do not show bias (blue) and adults (pink, from Experiment 2.5, Chapter 2) in the Visual Recall task, based on absolute correct scores. B) Mean recall (+/- SEM) accuracy of children who show bias (yellow), children who do not show pronounced bias (blue) and adults (pink, from Experiment 2.5, Chapter 2) in the Visual Recall task, based on proportion correct scores. C) Performance in the Grammaticality Judgement task for children who show bias (yellow), children who do not show bias (blue) and adults (pink). All groups showed above chance performance (indicated by the dashed line). Individual performance is shown as white circles. D) Adults (pink) show improved performance in the Sequence Generation task compared to biased children (yellow) and children who do not show bias (blue). Individual performance is shown as white circles. E) Adults (pink) show improved performance in the Sequence Generation task compared to biased children (yellow), but not compared to children who do not show bias (blue). Individual performance is shown as white circles.

Discussion

We aimed to use the same Visual Recall task we have previously used to measure learning in adults (Experiment 2.5) to measure learning in children. In children aged between 8 and 15 years, we found evidence of learning across the Visual Recall task and subsequent reflection-based measures, which indicates that the Visual Recall task is a suitable processing-based measure of implicit statistical learning in children. Performance in the Visual Recall tasks was positively correlated with performance across the reflection-based measures, which suggests that like adults, children may be gaining some explicit knowledge of the rules and using this to aid explicit decision-making in the reflection-based tasks.

We found no difference in performance across the experiment based on children's age, which mirrors previous studies suggesting that implicit statistical learning abilities are stable across childhood (Raviv & Arnon, 2018; Saffran et al., 1997). Regarding the developmental trajectory of implicit statistical learning beyond childhood, there was some evidence of a difference in implicit statistical learning ability between children and adults in the Visual Recall, Sequence Generation and Sequence Completion tasks, but not the Grammaticality Judgement task. However, after removing children who showed a pronounced bias in the Grammaticality Judgement task, there was little evidence of a difference in implicit statistical learning between children and adults in the Visual Recall task, although there was still some indication that the time-course of learning across the task differed based on age group. The findings suggest that any differences in implicit statistical learning between children and adults in this experiment may have been due to differences in attention and or motivation, or a lack of understanding of the reflection-based task instructions. This reflects the findings of a number of previous studies that suggest that implicit statistical learning is age-invariant (g., Thiessen et al., 2013).

While the Grammaticality Judgement task did not appear to underestimate children's implicit statistical learning abilities in this experiment, performance in this task did not capture the potential differences in the time-course of learning indicated by the processing-based Visual Recall task. Despite clear evidence of learning towards the end of the Visual Recall task, the children showed no improvement in recall accuracy over the grammatical Recall Blocks, which differed from the performance of adults across a number of our previous experiments (e.g., Experiments 2.4, 2.5, and 2.6). This lack of improvement at a group level may be due to the bimodal distribution in performance across the Visual Recall task: children who are engaging with the task show high levels of recall accuracy from the first Recall Block and continue to perform well across the task, meaning they do not show an improvement in performance across

Recall Blocks, because recall accuracy is already high. Children who show poor recall accuracy do not improve across the Recall Blocks, which may be due to a lack of attention and/or motivation to complete the task. In fact, poor attention has been suggested to influence implicit statistical learning in previous studies (Emberson et al., 2011; Pacton & Perruchet, 2008; Toro et al., 2005), and may explain the different time-course of learning between children and adults in the Visual Recall task.

In this experiment, we found consistent differences between children and adult's performance in the Sequence Generation task. This task requires participants to create their own sequences, therefore it is thought to require more explicit awareness of the structure underlying the sequences. Age-related differences in performance in this task may suggest that there may be differences between children and adults regarding the extent to which they can access any implicitly learned knowledge. These findings reflect those of Bertels et al. (2015), although there is a clear need for more research investigating any differences between children and adults in the nature of the knowledge acquired during these tasks. To accomplish this, future research should combine processing- and reflection- based measures of learning to provide a comprehensive overview of the nature of the knowledge acquired through implicit statistical learning.

These findings suggest that the effectiveness of recall paradigms as a measure of learning in children extend beyond the auditory domain, and critically, beyond the use of nonword stimuli, which is essential for the development of a processing-based measure that is suitable for children and adults with language difficulties. Furthermore, these findings suggest that the Visual Recall task is an effective measure of learning in online samples; however, the extent to which children's performance may differ between online and in-person testing is yet to be seen. It is possible that the potential lack of motivation, attention or understanding that is present across our sample is a result of completing the experiment online, as opposed to in a more controlled environment.

The results mirror those of a previous experiment using auditory serial recall to measure implicit statistical learning in children. Kidd et al. (2020) found that children showed improved recall of predictable over random sequences of nonwords in their auditory serial recall task, and above chance performance in a subsequent reflection-based measure. Similarly to the findings of our experiment in children, and our previous experiments with adults, performance across these two tasks was positively correlated. As both our sample of children and adults show correlations between performance across tasks, this may suggest that similar mechanisms underpin performance on these tasks irrespective of age (Bertels et al., 2015). These findings provide further support for the idea that whilst reflection-based measures may be affected by additional explicit processes, they are still to some extent measuring underlying implicit statistical learning mechanisms.

While parents were instructed to have their child complete the experiment in a quiet space, given that this study took place online, it is not possible to ensure that the children were free of distractions whilst completing the tasks. Additionally, the differences in implicit statistical learning at a group level may reflect differences in motivation between the children and adults. To recruit the sample of children, we advertised the study to parents, and whilst they may have been motivated for their child to take part, their child may have lacked this. By contrast, the adult sample was recruited from Prolific, where participants actively seek out studies to complete, and are therefore likely to be more motivated than our child sample. Due to the impact of the COVID-19 pandemic, data collection was only possible through online tasks, and therefore it would be useful to replicate this study in-person, where distractions can be minimised, and attention ensured. While we did not see any difference in adult's performance based on in-person or online completion, there is some suggestion that a poor performance across some children may be the result of a lack of attention, motivation, or the presence of distractions, rather than a lack of implicit statistical learning. This would provide additional information as to whether online tasks are a suitable substitute for in-person studies in children, in the same way that our previous experiments have indicated that they are for adults.

This experiment suggests that visual serial recall can measure implicit statistical learning in children, and that implicit statistical learning abilities remain stable across childhood. Although there were similarities in performance of children and adults across the Visual Recall task, there was some evidence that time-course of learning may differ between children and adults, which was not captured by traditional reflection-based tasks.

Chapter 4: Implicit Statistical Learning Deficits in Dyslexia

Abstract

Dyslexia is associated with a deficit in phonological processing. However, as dyslexia is also commonly associated with difficulties that are unrelated to language, alternative theories, such as implicit statistical learning deficits, have been proposed. In these experiments, we aimed to investigate the nature of implicit statistical learning deficits in dyslexia. In Experiments 4.1 and 4.2 we assessed implicit statistical learning of individuals with dyslexia (Experiment 4.1, in-person) and individuals with reading difficulties (Experiment 4.2, online) using the Visual Recall task, and compared performance to adults who had previously completed the same task in-person (Experiment 2.4) and online (Experiment 2.5). Across both experiments, we demonstrated that individuals with dyslexia or reading difficulties showed evidence of implicit statistical learning across both the Visual recall and subsequent reflection-based tasks. We found no evidence of impaired implicit statistical learning in individuals with dyslexia or reading difficulties; however it is possible that any deficits in dyslexia are associated with the processing of phonological stimuli only, and therefore would not result in impaired performance in Experiments 4.1 and 4.2. To directly examine whether any deficits in dyslexia are associated with processing phonological stimuli, In Experiments 4.3 and 4.4 we compared performance of individuals with and without dyslexia (Experiment 4.3, in-person) and individuals with and without reading difficulties (Experiment 4.4, online) in auditory artificial grammar learning tasks using both nonword and tone stimuli. Across both in-person and online experiments, we see evidence of learning in the nonword and tone artificial grammar learning tasks in all groups. Although in the in-person experiment (Experiment 4.3) there was some indication that individuals with dyslexia showed poorer performance across both nonword and tone tasks relative to controls, the COVID-19 pandemic resulted in recruitment being concluded prematurely, and therefore this experiment was underpowered due to a small sample size. In the online experiment (Experiment 4.4), conducted with a larger sample, we found no evidence that individuals with reading difficulties were impaired in either nonword or tone artificial grammar learning tasks. Across these 4 experiments, we found little evidence of a deficit in implicit statistical learning in individuals with dyslexia or reading difficulties. These findings reflect the conflicting nature of the literature in this field, which is likely a result of the considerable heterogeneity associated with dyslexia.

General Introduction

Dyslexia is characterised by difficulties in reading and spelling (Lyon et al., 2003), which rely on the efficient access and manipulation of phonemes. Therefore, the most prominent theory of dyslexia suggests that deficits in reading and spelling are due to difficulties in phonological processing, which is characterised by poorly specified phonological representations (Griffiths & Snowling, 2002; Snowling, 1998). Indeed, previous research has shown that individuals with dyslexia perform more poorly on tasks requiring the manipulation of phonemes (Denckla & Rudel, 1976; Farquharson et al., 2014; Jones et al., 2009; Melby-Lervåg et al., 2012; Rispens & Been, 2007; Vellutino et al., 1996). Furthermore, poor phonological processing is thought to be associated with other issues that may contribute to reading difficulties, such as the storage and retrieval of printed words (Gathercole & Baddeley, 1990; Vellutino et al., 1994). Difficulties in storing and retrieving printed words could interfere with successful mapping of grapheme-phoneme relationships, which in turn has a negative impact on the quality of word representations and the fluency of word identification (Vellutino et al., 2004). Taken together, there is strong evidence that dyslexia is associated with difficulties in phonological processing.

However, dyslexia is also associated with differences across other non-language related facets of cognition, including motor skills (Fawcett & Nicolson, 1995; Fawcett et al., 1996), auditory processing (Farmer & Klein, 1995; Tallal, 1984), and visual processing (Stein, 2001, 2019). Phonological theories of dyslexia are unlikely to account for these non-language related differences, therefore alternative theories have proposed that a more domain-general deficit, for example in in implicit statistical learning, may explain both the difficulties with language and differences in non-language related abilities often associated with dyslexia. However, previous research investigating implicit statistical learning deficits in dyslexia have yielded mixed results, with some studies providing evidence of a deficit (Katan et al., 2017; Pavlidou et al., 2009; Stoodley et al., 2008; Stoodley & Stein, 2006), and other showing no differences in implicit statistical learning between individuals with and without dyslexia (Inácio et al., 2018; Menghini et al., 2010; Nigro et al., 2016; Waber et al., 2003). As such, it is currently unclear whether any differences in dyslexia are restricted to the processing of phonological stimuli or can instead be attributed to a domain-general deficit in implicit statistical learning.

In a series of 4 experiments, I aimed to investigate the nature of implicit statistical learning deficits in dyslexia and assess whether any differences in dyslexia are restricted to processing of phonological stimuli. In Experiment 4.1, I used the Visual Recall and subsequent reflection-based tasks from previous experiments (Experiments 2.4, 2.5 and 3.1) to investigate the proposed deficits in implicit statistical learning in dyslexia in an in-person experiment. To assess whether any differences in dyslexia are due to a specific deficit in processing phonological stimuli, in Experiment 4.3 I aimed to compare performance in auditory nonword and tone tasks using traditional artificial grammar learning paradigms between individuals with and without dyslexia in an in-person experiment. However, due to the restrictions placed on testing in-person during the COVID-19 pandemic, I was forced to conclude recruitment for these studies prematurely. As such, this resulted in small sample sizes of individuals who we confirmed had dyslexia in Experiments 4.1 and 4.3. The data from these experiments have been included for completeness, however any findings should be interpreted with respect to the small sample size.

To address the issues with small sample sizes in Experiments 4.1 and 4.3, I adapted both experiments to run online in Experiments 4.2 and 4.4 respectively. In both online experiments, we recruited participants online through Prolific, which allows for the pre-screening of participants to only recruit individuals with reading difficulties. However, Prolific does not provide a method of specifically recruiting individuals with dyslexia. Although we asked participants during the experiment if they had dyslexia, there was no way to confirm this as participation was anonymous and standardised tasks cannot be administered online through Prolific. Therefore, in Experiments 4.2 and 4.4, we grouped individuals with reading difficulties and individuals that specifically indicated that they had dyslexia under broader term of “individuals with reading difficulties”. Although there is much debate regarding whether there are differences between individuals with dyslexia and ‘garden-variety’ poor readers (Elliott & Grigorenko, 2014; Gibbs & Elliott, 2020; Kirby, 2020), a number of studies have highlighted differences in implicit statistical learning between individuals with dyslexia and individuals with reading difficulties more generally (Badian, 1994; Stanovich, 1988). Therefore, it is important to note that individuals with dyslexia and individuals with reading difficulties may represent two distinct populations, and any deficits associated with dyslexia may not extend to those with reading difficulties more generally.

Experiments 4.1 & 4.2: Measuring Implicit Statistical Learning in Dyslexia Using Serial Visual Recall

Introduction

Deficits in implicit statistical learning have been proposed as a domain-general explanation for both the language and non-language related deficits in dyslexia (Folia et al., 2008; Gombert, 2003; Menghini et al., 2006; Ullman & Pierpont, 2005), although the findings from the literature are mixed (see Schmalz et al., 2017, for a review). However, the artificial grammar learning paradigms that are typically used to assess implicit statistical learning in dyslexia may not be the most appropriate method for testing these populations for several reasons. In order to explore potential differences in implicit statistical learning in dyslexia, it is important that measures are not confounded by other cognitive processes, such as decision-making, which are also thought to be impaired in dyslexia (Manning et al., 2022; Stefanac et al., 2021). As previously discussed, reflection-based tasks that are typically used to measure learning in artificial grammar learning paradigms require conscious decision-making, and therefore performance on these tasks may be more of a reflection of these explicit processes than they are of implicit statistical learning. This means that poorer performance of individuals with dyslexia relative to controls in reflection-based tasks may not indicate differences in implicit statistical learning, but differences in other, more explicit cognitive processes. Indeed, there is some indication that children with language difficulties show comparable implicit statistical learning to children without language difficulties in a processing-based measure of learning, but show impairments based on reflection-based measures of learning (Lukács et al., 2021).

Processing-based measures provide additional advantages over reflection-based measures when investigating implicit statistical learning in dyslexia. In addition to measuring learning without requiring more explicit decision-making, processing-based measures allow the time-course of learning to be assessed over the course of the experiment, as opposed to simply assessing whether learning has occurred after the fact (Siegelman et al., 2018). Assessing the time-course of learning is particularly important for investigating implicit statistical learning differences in dyslexia, as this may reveal differences in how learning occurs compared to controls (Lukács et al., 2021). For example, some studies have shown that while implicit statistical learning was intact in participants with dyslexia, these individuals provided slower responses (Kelly et al., 2002). Although processing-based measures such as serial reaction time tasks have previously been used to investigate differences in implicit statistical learning in dyslexia (with mixed findings; see Inácio et al, 2018; Menghini et al., 2010; Nigro et al., 2016;

Waber et al., 2003), these tasks may rely on different cognitive and neural mechanisms, as they rely on learning relationships between spatial locations rather than between sensory stimuli (Conway & Christiansen, 2005). Furthermore, these tasks require fast motor responses, which provides additional issues when investigating implicit statistical learning in dyslexia, in which motor (Fawcett & Nicolson, 1995; Fawcett et al., 1996) and processing-speed (Stoodley & Stein, 2006) deficits have been reported. More recently, processing-based measures of serial recall have been developed (Isbilen et al., 2017, 2020). However, all previous versions of these tasks require learning dependencies between phonological stimuli, which is not appropriate for exploring implicit statistical learning in dyslexia, due to the deficits in phonological processing that are associated with dyslexia (Demb et al., 1999; Farquharson et al., 2014; Rispens & Been, 2007; Vellutino et al., 1996). The serial Visual Recall task we have developed (Chapters 2 and 3) does not require processing of phonological stimuli, and therefore would be appropriate to measure implicit statistical learning in individuals with dyslexia.

In Experiment 4.1, we conducted an in-person study measure implicit statistical learning of adults with dyslexia using the serial visual recall task from Experiments 2.5 and 3.1. We compared their performance to the individuals without dyslexia who completed the same tasks in Experiment 2.5. As previously discussed, due to the COVID-19 pandemic, we were not able to recruit as many dyslexic participants as we had hoped. Therefore, we adapted the experiment to run online in Experiment 4.2. This sample was recruited through Prolific, and whilst we could specifically screen for individuals with reading difficulties, we could not confirm which of these individuals had dyslexia specifically. Therefore, although in Experiment 4.1 we specifically compared individuals with and without dyslexia, in Experiment 4.2 we compared individuals with reading difficulties to a previously recruited online sample of individuals without reading difficulties (see Experiment 2.5). For clarity and conciseness, in this chapter the results of these two experiments will be presented in the same section. Across both in-person and online experiments, we predicted that we would see evidence of learning across both the Visual Recall task and subsequent reflection-based measures, and that performance across the Visual Recall and reflection-based tasks would be correlated. If dyslexia or reading difficulties are associated with deficits in implicit statistical learning, then we would predict that these individuals would show poorer performance across the tasks than individuals without dyslexia or reading difficulties. If there is no implicit statistical learning deficit, then we would expect no differences in performance across groups.

Methods

Participants

In Experiment 4.1, 16 adults with dyslexia (9 female, 6 male, 1 other; mean age = 25.38) were recruited using the Newcastle University Neuroscience Participant Pool and the School of Psychology Student Participant Pool. Although I originally aimed to recruit 40 participants for this study, I was unable to complete recruitment for this experiment due to the COVID-19 pandemic. Although this is a very small sample size, the data are reported for completeness and transparency. All participants were native English speakers, and had normal or corrected-to-normal vision and hearing. Participants were not excluded based on their ability to speak any additional languages. Ethics was approved by the Faculty of Medical Sciences Ethics Committee at Newcastle University.

In Experiment 4.2, 38 adults (15 female, 21 male, 2 other; mean age = 25.88) were recruited from Prolific to complete the online version of the Chunking recall task. This sample size was selected based on previous online experiments (Experiments 2.3, 2.5, 2.6 and 3.1). We pre-screened participants using Prolific to recruit native English speakers with reading difficulties, and also included a question within the experiment asking whether the participant had dyslexic specifically. Participants were not excluded based on their ability to speak any additional languages. Of the 38 participants, 19 reported that they had dyslexia. An additional 3 participants completed the experiment but were excluded from the analysis for failing attention checks.

Stimuli

The stimuli were the same as in Experiments 2.4, 2.5 and 3.1.

Procedure

Experiment 4.1 took place in a testing lab within the Institute of Neuroscience at Newcastle University and was coded using MATLAB and Psychtoolbox. Participants were seated approximately 60cm in front of a computer monitor (24-inch Dell U2412M, screen resolution 1920*1200 pixels). The procedure was identical to Experiment 4: the same standardised cognitive and language tasks were administered (Appendix 2.1.), and participants completed the Visual Recall, Grammaticality Judgement, Sequence Generation and Sequence Completion tasks in the same order as Experiment 2.4.

In Experiment 4.2, the procedure was identical to Experiment 2.5.

Data Analysis

The data from the in-person (Experiments 4.1) and online (Experiment 4.2) experiments were analysed in the same way as Experiments 2.4 and 2.5 respectively. In Experiment 4.1, we compared scores on the standardised cognitive and language tasks across control and dyslexic participants in order to determine whether the participants with dyslexia performed more poorly on the language tasks. As such, we predicted that the participants with dyslexia would perform more poorly in the TOWRE Words and Nonwords, Recalling Sentences, RAN Digits and Objects tasks, but not in the WASI Blocks task: a measure of nonverbal IQ. Dyslexia is also commonly associated with deficits in working memory, and therefore we may also expect that the participants with dyslexia show poorer performance in the Backwards Digit Recall task as well. Controls showed better performance in the TOWRE Nonword (t_{43} = 3.00, p = .004) and Recalling Sentences (t_{43} = 2.63, p = .012) tasks only. We found no differences across the other standardised tasks (TOWRE Word: t_{43} = 1.53, p = .134; Backward Digit Recall: t_{43} = 1.71, p = .094; WASI Block Design: t_{43} = .084, p = .934; RAN Digits: t_{43} = -.725, p = .472; RAN Objects: t_{43} = -1.59, p = .118). This may suggest that there were no differences in language ability between the two groups, due to our dyslexic sample also being drawn from university students who generally show good language ability. Alternatively, the lack of differences may also be a consequence of the small size of our dyslexic sample.

In Experiment 4.2, which was conducted online, it was not possible to collect the same standardised cognitive and language task data. As analyses showed no differences in performance across any of the tasks between participants who reported reading difficulties and those who reported that they had dyslexia, we combined these groups to form a single group of individuals with reading difficulties. As previously discussed, it is worth noting that individuals with dyslexia may show distinct differences to individuals who report having more general reading difficulties, and therefore comparisons between the in-person and online experiment should be considered with this in mind.

Results

We found strong evidence of learning across all tasks in both the in-person and online versions of the experiment. In the in-person Visual Recall task, we saw the predicted pattern of performance based on both absolute and proportion correct scores: recall accuracy was significantly higher in Recall Block 6 compared to Recall Block 1 (absolute correct: t_{15}= 2.78, p = .014; proportion correct: t_{15}= 5.17, p < .001, Figure 4.1); higher in Recall Block 6 compared to the Testing Block (absolute correct: t_{15}= 2.72, p = .016; proportion correct: t_{15}= 3.54, p =

.003); and higher in the Recovery Block than the Testing Block (absolute correct: t_{15}= 3.07, p = .008; proportion correct: t_{15}= 3.78, p = .002).

The online experiment showed the same pattern of learning: Recall accuracy was higher in Recall Block 6 compared to Recall Block 1 (absolute correct: t_{37} = 4.76, p = < .001; proportion correct: t_{37} = 5.02, p < .001); higher in Recall Block 6 compared to the Testing Block (absolute correct: t_{37} = 3.32, p = .002; proportion correct: t_{37} = 3.49, p = .001) and higher in the Recovery Block compared to the Testing Block (absolute correct: t_{37} = -3.07, p = .004; proportion correct: t_{37} = 2.75, p = .009). These results provide strong evidence that implicit statistical learning has occurred during the Visual Recall task.

Table 4.1. Descriptive Statistics for the Visual Recall Task in Experiments 4.1 and 4.2.

	Experiment 4.1 (In-Person)				Experiment 4.2 (Online)			
	Absolute Correct		Proportion Correct		Absolute Correct		Proportion Correct	
	Mean	SEM	Mean	SEM	Mean	SEM	Mean	SEM
Baseline	0.10	0.02	0.51	0.03	0.21	0.03	0.55	0.02
Recall 1	0.08	0.03	0.51	0.03	0.22	0.05	0.57	0.03
Recall 2	0.16	0.04	0.55	0.04	0.30	0.05	0.64	0.04
Recall 3	0.22	0.05	0.62	0.04	0.38	0.06	0.67	0.04
Recall 4	0.19	0.07	0.62	0.05	0.42	0.06	0.70	0.04
Recall 5	0.25	0.06	0.65	0.04	0.40	0.06	0.68	0.04
Recall 6	0.26	0.07	0.66	0.05	0.46	0.06	0.73	0.04
Testing	0.10	0.03	0.56	0.03	0.30	0.05	0.63	0.04
Recovery	0.30	0.08	0.69	0.04	0.45	0.06	0.70	0.04

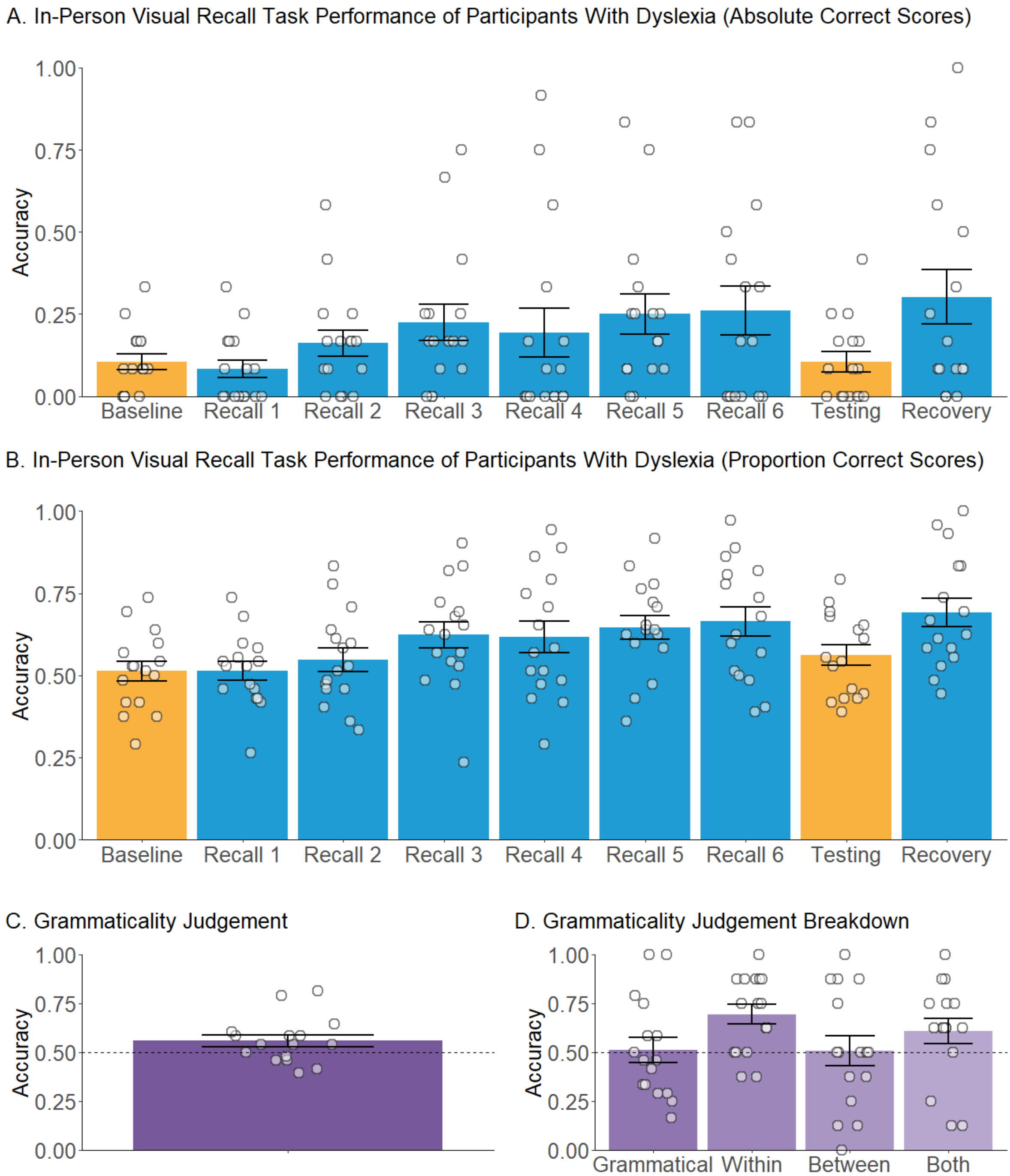


Figure 4.1. Visual Recall and Grammaticality Judgement task performance for individuals with dyslexia recruited in-person (Experiment 4.1). A) Mean recall (+/- SEM) accuracy in the Visual Recall task, based on absolute correct scores. Orange bars indicate blocks of random sequences, whereas blue bars indicate blocks of grammatical sequences. Recall increases across grammatical Recall Blocks, before significantly decreasing in the Testing Block. In the final Recovery Block, performance is significantly higher than in the Testing Block. This pattern of results provides clear evidence that implicit statistical learning has taken place. B) The same pattern was observed when using proportion correct scores. C) Performance on the Grammaticality Judgement task was above chance (indicated by the dashed line). Individual performance is shown as white circles. D) Breakdown of performance across the different types of sequences in the Grammaticality Judgement task. Participants performed above chance (indicated by the dashed line) at correctly classifying the within sequences only.

In the in-person experiment (Experiment 4.1), while participants were close to performing above chance in the Grammaticality Judgement task (*M* = 0.56, SEM = 0.03; t_{15}= 1.99, p = .065; Figure 4.1), although it is perhaps unsurprising that this did not reach significance given the small sample size. When examining the breakdown of performance across within, between and both sequence conditions, participants performed above chance only in classifying the sequences containing violations within chunks (*M* = 0.70, SEM = 0.05; t_{15} = 3.93, p = .001). Performance across the other conditions did not differ from chance (grammatical: *M* = 0.51, SEM = 0.06; t_{15} = .20, p = .842; between chunks: *M* = 0.51, SEM = 0.08; t_{15} = .10, p = .919; both between and within chunks: *M* = 0.61, SEM = 0.06; t_{15} = 1.73, p = .105).

In the online version of the experiment, with a larger sample size, performance in the Grammaticality Judgement task was above chance (*M* = 0.62, SEM = 0.02; t_{37} = 5.34, $p < .001$; Figure 4.2), suggesting that learning had occurred in this experiment. Participants performed above chance in classifying the grammatical (*M* = 0.70, SEM = 0.05; t_{37} = 4.36, $p < .001$), within (*M* = 0.64, SEM = 0.04; t_{37} = 3.32, p = .002) and both (*M* = 0.64, SEM = 0.05; t_{37} = 3.10, p = .004) sequences, but below chance in classifying the between sequences (*M* = 0.35, SEM = 0.05; t_{37} = -2.82, p = .008) in the online version of the task. This pattern of performance suggests that participants were learning the within chunk relationships, but not the between chunk relationships.

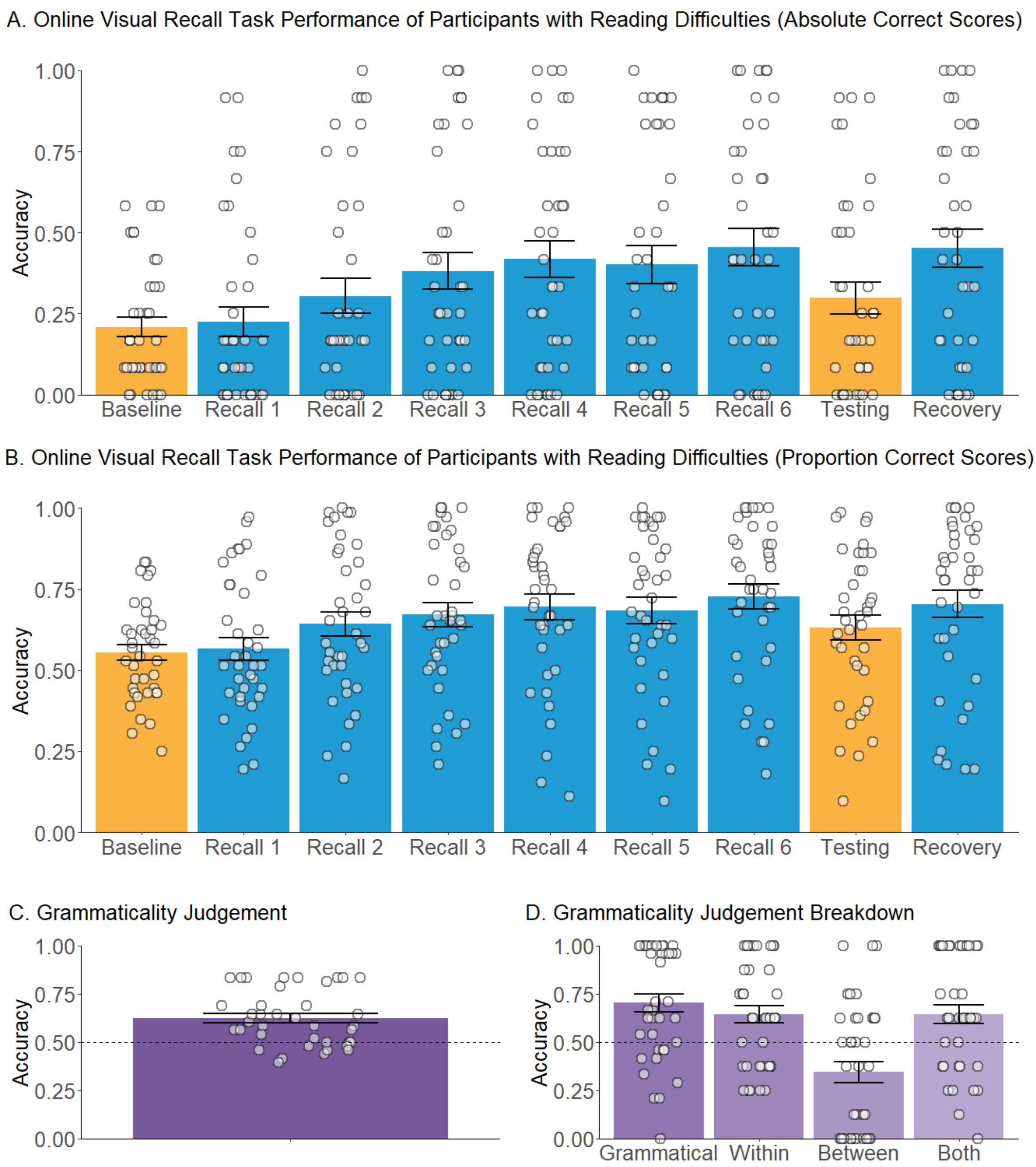


Figure 4.2. Visual Recall and Grammaticality Judgement task performance for individuals with reading difficulties recruited online (Experiment 4.2). A) Mean recall (+/- SEM) accuracy in the Visual Recall task, based on absolute correct scores. Orange bars indicate blocks of random sequences, whereas blue bars indicate blocks of grammatical sequences. Recall increases across grammatical Recall Blocks, before significantly decreasing in the Testing Block. In the final Recovery Block, performance is significantly higher than in the Testing Block. This pattern of results provides clear evidence that implicit statistical learning has taken place. B) The same pattern was observed when using proportion correct scores. C) Performance on the Grammaticality Judgement task was above chance (indicated by the dashed line). Individual performance is shown as white circles. D) Breakdown of performance across the different types of sequences in the Grammaticality Judgement task. Participants performed above chance (indicated by the dashed line) at correctly classifying the grammatical sequences, as well as the within and both violation sequences, but below chance in the between violation sequences.

As in previous experiments, we predicted that participants who performed well in the Grammaticality Judgement task would also perform well in the Sequence Generation and Sequence Completion tasks. For completeness, I report these correlational tests for both in-person and online experiments, although it is clear that the sample size for the in-person experiment is too small to draw any reliable conclusions from these results. In the in-person experiment, we did not find a correlation between performance in the Grammaticality Judgement task and performance in either the Sequence Generation (r = .334, p = .206) or Sequence Completion task (r = .355, p = .177), although performance in the Sequence Generation and Sequence Completion tasks was positively correlated (r = .520, p = .039).

In the online version, performance in the Grammaticality Judgement task was positively correlated with performance in the Sequence Generation (r = .462, p = .004) and Sequence Completion (r = .861, p < .001) tasks as predicted, which suggests that participants who performed well in the Grammaticality Judgement task had acquired some explicit knowledge of the structure which was needed to create their own sequences. Performance in the Sequence Generation and Sequence Completion tasks was also positively correlated (r = .636, p < .001).

As in our previous experiments (Experiments 2.2 – 2.6 and Experiment 3.1) using serial visual recall paradigms, to investigate whether the processing- and reflection-based measures were assessing similar processes, we calculated composite measures of both performance in the Visual Recall task and performance across the reflection-based tasks (see Experiment 2.4). These measures were positively correlated in both the in-person (r = .611, p = .012), and online (r = .545, p < .001) versions of the experiment, which reflects the findings across our previous experiments and suggests that the processing- and reflection-based measures may be assessing similar processes in these experiments.

In the in-person experiment, we also conducted a number of standardised cognitive and language tasks, to assess whether language ability was correlated with implicit statistical learning, as measured by either the Visual Recall or subsequent reflection-based tasks. Composite performance in the Visual Recall task was correlated with performance in the WASI Block Design task only (Appendix 4.1). Composite performance across the reflection-based tasks was correlated with performance in the Backwards Digit Recall and WASI Block Design tasks (Appendix 4.1); however, given the small sample size, we cannot draw strong conclusions based on these findings.

To assess whether dyslexia is associated with differences in implicit statistical learning across the in-person and online versions of the Visual Recall task, we conducted a 9x2x2 mixed

ANOVA, with run (9 runs) as a within-subjects factor, and group (control or reading difficulties) and task (in-person or online) as between-subjects factors (Figures 4.3 and 4.4). There was no main effect of task ($F_{1,116} = 2.89$, $p = .092$) or interaction between any of the variables, indicating that performance did not differ based on whether the experiment was completed in-person or online. There was a main effect of group ($F_{1,116} = 5.69$, $p = .019$), with individuals with reading difficulties showing poorer memory overall compared to individuals without reading difficulties. However, memory deficits are often reported in dyslexia and poorer recall accuracy across the Visual Recall task is not indicative of a deficit in implicit statistical learning. Instead, differences in implicit statistical learning between groups may be reflected in an interaction between run and group, but most importantly, through a difference in composite performance across the Visual Recall task. There was a main effect of run ($F_{3.9,458} = 32.17$, $p < .001$). Bonferroni corrected post hoc tests indicated that there were significant differences in recall accuracy between the Baseline Block and all other blocks ($p < .05$) except Recall Block 1 and the Testing Block ($p > .05$). There were also differences between Recall Block 1 and all other blocks ($p < .05$) except the Testing Block ($p > .05$). There were significant differences in recall accuracy between Recall Block 2 and Recall Blocks 5 and 6, and the Recovery Block ($p < .05$) but not with other blocks ($p > .05$). There were further significant differences between recall accuracy in Recall Block 3 and Recall Block 6 and the Recovery Block ($p < .05$), and similarly between Recall Block 4 and Recall Block 6 and the Recovery Block ($p < .05$). There was a significant difference between Recall Block 5 and the Testing Block, between Recall Block 6 and the Testing Block, and between the Recovery Block and the Testing Block ($p < .05$). There were no other significant differences in recall accuracy between blocks ($p > .05$ in all cases).

There was no interaction between run and group ($F_{1,116} = 2.71$, $p = .102$). This suggests that individuals with reading difficulties did not show any differences in implicit statistical learning in these experiments, including in the time-course of learning. Indeed, when comparing composite performance across the Visual Recall task between individuals with and without reading difficulties, there was no difference in implicit statistical learning either in-person ($t_{44} = .973$, $p = .336$) or online ($t_{72} = 1.14$, $p = .259$). Furthermore, there was no difference in composite performance across the reflection-based measures between individuals with and without reading difficulties in the in-person ($t_{44} = .853$, $p = .398$) or online ($t_{73} = -0.19$, $p = .852$) experiments. These findings suggest that reading difficulties do not impact implicit statistical learning across the processing- and reflection-based tasks.

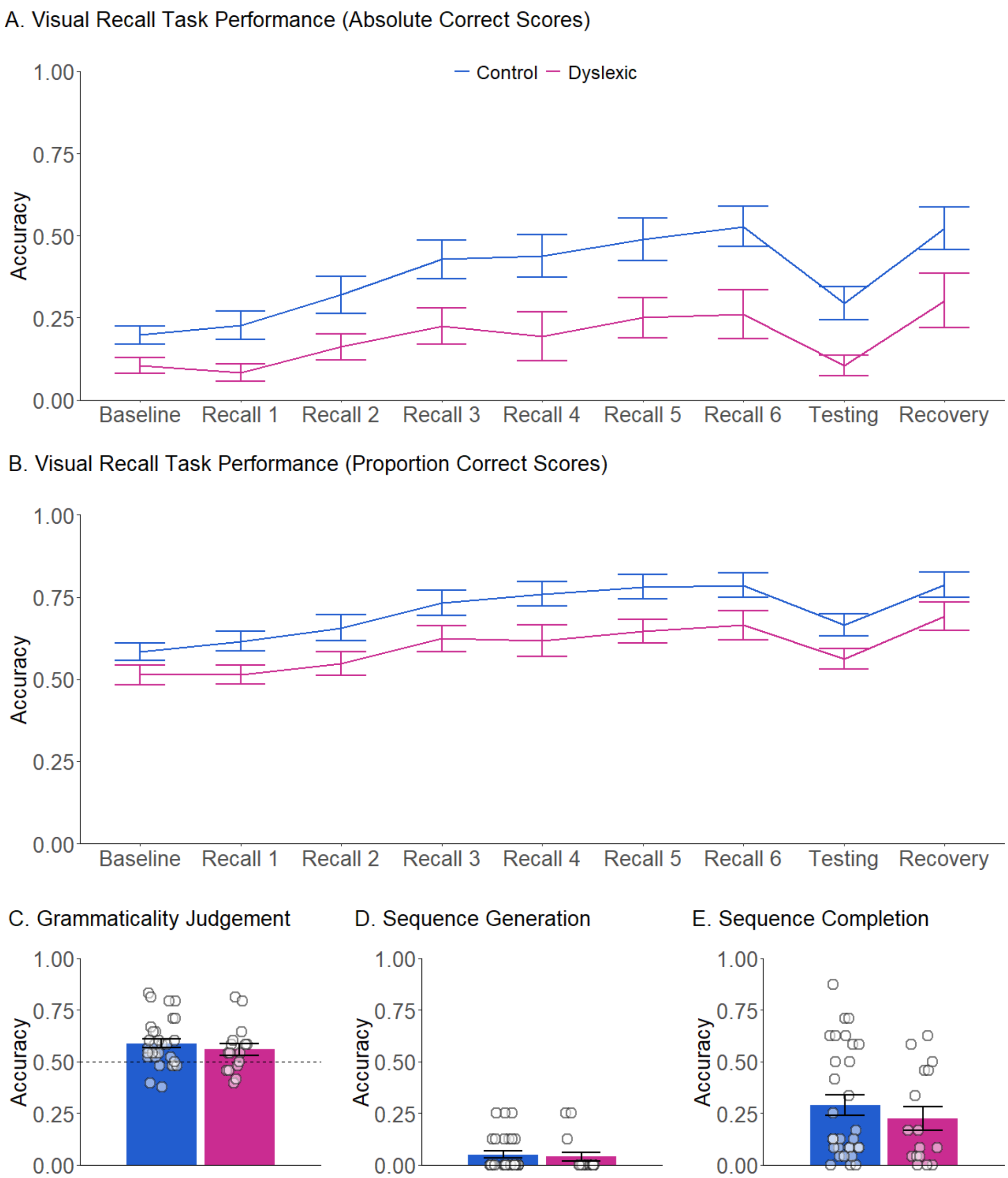


Figure 4.3. Performance of individuals with and without dyslexia across the in-person Visual Recall and reflection-based tasks (Experiment 4.1). A) Mean recall (+/- SEM) accuracy of individuals without dyslexia (blue, from Experiment 2.4) and individuals with dyslexia (purple) in the Visual Recall task, based on absolute correct scores. B) Mean recall (+/- SEM) accuracy of accuracy of individuals without dyslexia (blue) and individuals with dyslexia (purple) in the Visual Recall task, based on proportion correct scores. Based on both methods of scoring, there is no difference in implicit statistical learning between individuals with and without dyslexia. C) Performance in the Grammaticality Judgement task for individuals without dyslexia (blue) and individuals with dyslexia (purple). Both groups showed above chance performance (indicated by the dashed line). Individual performance is shown as white circles. D) There was no difference between performance of individuals without dyslexia (blue) and individuals with dyslexia (purple) in the Sequence Generation task. Individual performance is shown as white circles. E) There was no difference between performance of individuals without dyslexia (blue) and individuals with dyslexia (purple) in the Sequence Completion task. Individual performance is shown as white circles.

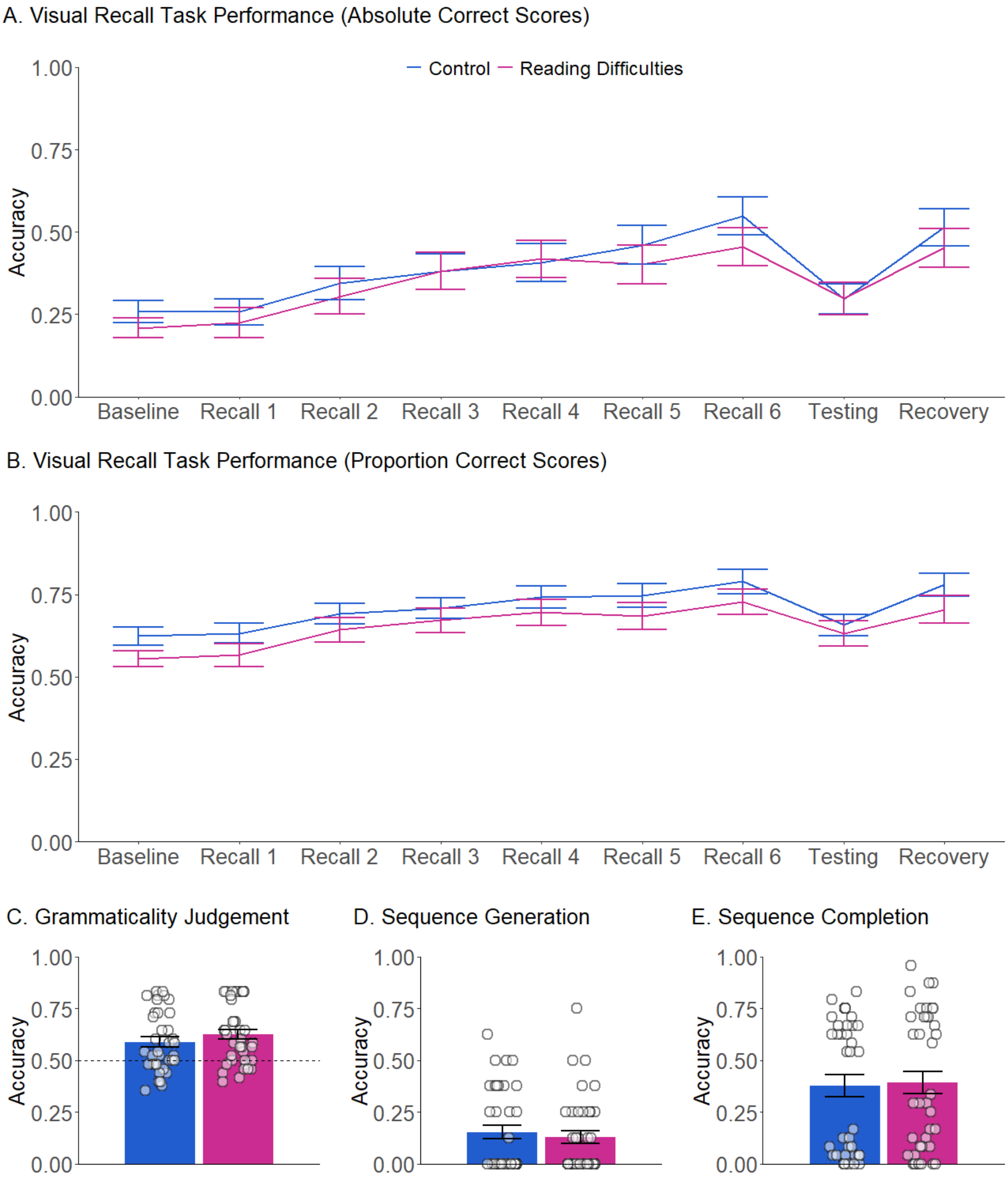


Figure 4.4. Performance of individuals with and without reading difficulties across the online Visual Recall and reflection-based tasks (Experiment 4.2). A) Mean recall (+/- SEM) accuracy of individuals without reading difficulties (blue, from Experiment 2.5) and individuals with reading difficulties (purple) in the Visual Recall task, based on absolute correct scores. B) Mean recall (+/- SEM) accuracy of accuracy of individuals without reading difficulties (blue) and individuals with reading difficulties (purple) in the Visual Recall task, based on proportion correct scores. Based on both methods of scoring, there is no difference in implicit statistical learning between individuals with and without reading difficulties. C) Performance in the Grammaticality Judgement task for individuals without reading difficulties (blue) and individuals with reading difficulties (purple). Both groups showed above chance performance (indicated by the dashed line). Individual performance is shown as white circles. D) There was no difference between performance of individuals without reading difficulties (blue) and individuals with reading difficulties (purple) in the Sequence Generation task. E) There was no difference between performance of individuals without reading difficulties (blue) and individuals with reading difficulties (purple) in the Sequence Completion task.

Discussion

In Experiments 4.1 and 4.2 we found that individuals with reading difficulties show evidence of learning in the processing-based Visual Recall task. In Experiment 4.1, conducted in-person, performance in the Grammaticality Judgement task just failed to reach significance, although this is likely a reflection of the very small sample size (16 participants) that resulted from stopping in-person testing during the COVID-19 lockdown. This idea is supported by the findings from Experiment 4.2, conducted online with a larger sample (38 participants), which showed good evidence of learning across the reflection-based tasks. The findings from these experiments suggests that the Visual Serial Recall task is suitable for measuring implicit statistical learning in individuals with reading difficulties. However, we found no difference in performance across any of the tasks between individuals with and without reading difficulties, including no differences in the time-course of learning across the task, which may provide evidence against an impairment in implicit statistical learning in these populations.

The findings from these experiments do not provide support for differences in implicit statistical learning in individuals with reading difficulties such as dyslexia. Whilst previous literature has found some evidence of an impairment in these populations (Kahta & Schiff, 2016, 2019; Katan et al., 2017; Pavlidou et al., 2009; Stoodley & Stein, 2006), there is also a considerable number of studies that do not report any differences between individuals with and without dyslexia (Inácio et al., 2018; Menghini et al., 2010; Nigro et al., 2016; Samara & Caravolas, 2017; Waber et al., 2003). Many of the studies that report differences in implicit statistical learning in dyslexia have measured learning using traditional artificial grammar learning paradigms involving passive exposure followed by a reflection-based task. It is possible that poor performance in previous studies may be a consequence of differences in other aspects of cognition, such as difficulties with attention, which are commonly found in individuals with dyslexia (Facoetti & Molteni, 2001; Facoetti et al., 2000; Marzocchi et al., 2009). If such differences affect dyslexic individuals' attention in the passive exposure phase, poorer performance in the reflection-based task may be a result of these attentional difficulties as opposed to a deficit in implicit statistical learning. This would explain the findings of Experiments 4.1 and 4.2, where we find no evidence of implicit statistical learning deficits in individuals with dyslexia. Furthermore, in these experiments there is no passive exposure. Instead, participants are exposed to grammatical sequences whilst completing the Visual Recall task, which may result in improved attention during learning. If poorer performance of individuals with dyslexia in traditional artificial grammar learning tasks can be attributed to a lack of attention during passive exposure, then processing-based tasks containing active exposure may be particularly beneficial for individuals with dyslexia, as this increased attention may facilitate performance and account for the lack of differences in performance across the reflection-based tasks between individuals with and without dyslexia in these experiments.

However, it is possible that we did not find any differences in implicit statistical learning between individuals with and without dyslexia in Experiment 4.1 because there were no differences in language ability between the two groups, which may be due to both groups consisting of university students, who generally have good language ability. It could be argued that if the groups show no differences in language ability, then we may not expect to see any differences in implicit statistical learning between groups. Furthermore, due to the inability to recruit participants for in-person testing over the course of the pandemic, we were not able to recruit as many individuals with dyslexia as originally planned, and therefore the sample size in Experiment 4.1 is very small. Although we have reported this for completeness, to address this, in Experiment 4.2 we ran the same tasks online, however this resulted in other issues. First, although Prolific did allow pre-screening to recruit individuals with reading difficulties only, it did not provide an option to specifically recruit individuals with dyslexia. As previously discussed, although in Experiment 4.2 we grouped participants who reported reading difficulties and those who reported that they had dyslexia, these groups may represent distinct populations, and any deficits associated with dyslexia may not extend to those with reading difficulties more generally. As the experiment was completed online, we could not administer the standardised cognitive and language tasks. Therefore, it is difficult to determine whether there was a difference in language ability between individuals with and without reading difficulties, which raises the same issues as in Experiment 4.1.

Despite these issues, we provide evidence that the Visual Recall task can be used to measure implicit statistical learning in individuals with reading difficulties. However, we do not find evidence of a deficit in implicit statistical learning in individuals with dyslexia or reading difficulties. However, dyslexia may be characterised by a deficit in processing phonological information only, and the lack of group differences could be due to the tasks relying on the processing of non-phonological stimuli. Therefore, in Experiments 4.3 and 4.4, we aimed to investigate whether any deficits in dyslexia are restricted to the processing of phonological information.

Experiments 4.3 & 4.4: Is Dyslexia Associated with Language-Specific or Domain-General Deficits?

Introduction

Experiments 4.1 and 4.2 suggest that there is no difference in implicit statistical learning between individuals with and without reading difficulties, including dyslexia. These experiments were designed to avoid the need to process phonological stimuli and ensure any group differences could be attributed to differences in implicit statistical learning. However, it is possible that any difficulties associated with dyslexia are domain-specific, that is, restricted to the processing of phonological information (Griffiths & Snowling, 2002; Snowling, 1998; Vellutino et al., 2004), as opposed to a more general domain-general difference in implicit statistical learning (Folia et al., 2008; Gombert, 2003).

Despite contrasting evidence, few studies have sought to directly address whether dyslexia is associated with a domain-specific deficit in processing phonological information or a domain-general deficit in implicit statistical learning. Gabay et al. (2015) compared the performance of individuals with and without dyslexia in two artificial grammar learning paradigms: one using nonword stimuli (phonological) and one using tone stimuli (non-phonological). A deficit in phonological processing specifically would impair dyslexic individuals' processing of nonword stimuli only (Denckla & Rudel, 1976; Farquharson et al., 2014; Jones et al., 2009; Melby-Lervåg et al., 2012; Rispens & Been, 2007; Vellutino et al., 1996)., whereas a domain-general statistical learning deficit would result in an impairment in both nonword and tone stimuli (Katan et al., 2017; Pavlidou et al., 2009; Stoodley et al., 2008; Stoodley & Stein, 2006). Both control and dyslexic participants performed above chance on both the linguistic and non-linguistic versions of the task, however participants in the control group performed significantly better on both tasks compared to the individuals with dyslexia, providing evidence supporting a domain-general deficit in implicit statistical learning.

Although Experiments 4.1 and 4.2 aimed to assess implicit statistical learning deficits in dyslexia using processing-based measures, in Experiments 4.3 and 4.4, conducted in-person and online respectively, we aimed to directly compare domain-specific and domain-general theories of dyslexia. To do this, we used similar auditory nonword and tone artificial grammar learning paradigms as Gabay et al. (2015) to measure learning of the Chunking grammar we had previously developed (Figure 2.8, or Experiment 2.4 for more details).

In Experiments 4.3 and 4.4, conducted in-person and online respectively, we aimed to extend these findings using the Chunking Grammar that we had previously developed (see Chapter 2, Experiments 2.4 and 2.5), by comparing the performance of adults with and without dyslexia or reading difficulties using both nonword and tone tasks. As in Gabay et al. (2015), we used a traditional artificial grammar learning paradigm. This consisted of passive exposure to a continuous stream of either nonword syllables or sine-wave tones generated by the Chunking grammar. This stimulus stream therefore contained both within and between chunk transitions which aimed to mimic speech in natural language. Following this, participants completed a traditional grammaticality judgement task. If dyslexia is caused by a specific deficit in phonological processing, then we would expect dyslexic individuals to be impaired on the nonword artificial grammar learning (AGL) task relative to controls, but not the Tone AGL task. If there is a more domain-general deficit in implicit statistical learning, then we would predict that individuals with dyslexia would be impaired relative to controls on both the Nonword and Tone AGL tasks.

Methods

Participants

In Experiment 4.3, conducted in-person, 26 participants (12 female, 14 male, mean age = 21.62), were recruited using the Newcastle University Neuroscience Participant Pool and the School of Psychology Student Participant Pool. 16 participants did not report any language difficulties (controls), and 10 reported that they had dyslexia. As in Experiment 4.1, although we originally aimed to recruit 40 participants for this study, data collection for this experiment was concluded prematurely due to COVID-19. Although this is a very small sample size, the data are reported for completeness and transparency. All participants were native English speakers and had normal or corrected-to-normal vision and hearing. Participants were not excluded based on their ability to speak any additional languages.

In Experiment 4.4, conducted online, we recruited 99 participants (36 female, 62 male, 1 other; mean age = 23.40) through Prolific. This sample size was selected to result in similar numbers of participants per group as in previous online experiments (Experiments 2.3, 2.5, 2.6, 3.1 and 4.2). As in Experiment 4.2, we were able to selectively recruit individuals who were native English speakers and who reported that they had reading difficulties through Prolific. We also included a question within the experiment asking whether the participant had dyslexic specifically. 45 participants did not report any language difficulties (controls) and 54 participants reported that they had reading difficulties. Of these 54 participants, 26 reported that they had dyslexia. As we could not pre-screen participants who had dyslexia specifically, we recruited more participants with reading difficulties compared to those without, to try and increase the number of participants with dyslexia specifically in this experiment. Participants were not excluded based on their ability to speak any additional languages. An additional 7 participants were recruited but subsequently excluded from the analysis for failing attention checks. Ethics was approved by the Faculty of Medical Sciences Ethics Committee at Newcastle University for both experiments.

Stimuli

In both Experiments 4.3 and 4.4, the stimuli were generated based on the Chunking Grammar (Figure 2.8, see Experiment 2.4 for more details). For the nonword task, 8 mono-syllabic nonwords corresponding to each of the elements in the grammar were generated using MBROLA (Dutoit et al., 1996; see Table 4.1.). For the tone task, 8 pure tones based on Conway and Christiansen (2006) were generated using MATLAB. The frequencies of these tones were deliberately chosen to avoid musical notes and intervals between musical notes. In both the nonword and tone tasks, the duration of each of the elements was 250ms.

Table 4.2. Experiments 4.3 and 4.4 nonword and tone stimuli. 8 mono-syllabic nonwords were generated using MBROLA for the nonword task, and 8 pure tones were generated using MATLAB for the tone task. The frequencies of the tones were deliberately chosen to avoid musical notes or the intervals between musical notes.

Element	Nonword	Tone
A	Tor	180 Hz
B	Li	389 Hz
C	Mo	210 Hz
D	Ku	333 Hz
E	Di	454 Hz
F	Ga	286 Hz
G	Nu	531 Hz
H	Fay	245 Hz

Both tasks were split into exposure and testing phases. There were two types of exposure phase: initial exposure phases, which were a longer period of exposure which occurred at the start of each AGL task, and short exposure phases, which were spread throughout the AGL tasks to refamiliarise participants with the grammatical transitions. In both types of exposure phase, the nonword or tone elements were combined in MBROLA in accordance with the rules of the grammar to form a continuous stream of stimuli, containing within and between chunk transitions, that aimed to mimic natural speech. In the longer initial exposure phases, 720 elements were presented, resulting in 3 minutes of exposure. In the short exposure phases, 60 elements were presented, resulting in 30 seconds of exposure. The testing phase consisted of a Grammaticality Judgement task, which consisted of the same 48 sequences that we had previously used in the Visual Recall task using the Chunking Grammar (see Chapter 2, Table 2.1).

Procedure

In Experiment 4.3, the lab set-up was identical to previous in-person experiments. First, the TOWRE Words and Nonwords, Backwards Digit Recall and the Recalling Sentences Task were administered (see Appendix 2.1). Following this, participants completed both the Nonword and Tone AGL tasks, with the order of completion of these tasks counterbalanced across participants. The nonword and tone tasks were identical, apart from the stimuli used. Participants completed two runs of each task. In each run, participants were first exposed to a 3-minute-long continuous stream of stimuli which were generated in accordance with the grammar. Participants were told to pay attention to the stimulus stream but were not informed about the existence of rules within the stimulus stream. Following the initial exposure phase, participants were told that the sequences they had just heard followed a pattern, and that they would be presented with new sequences and asked if they fit the pattern they had previously heard. After each of the sequences was presented, participants were told to press the “c” key if they thought the sequence fit the pattern, or the “m” key if the sequence did not fit the pattern. The Grammaticality Judgement task was split into 6 blocks, each containing 8 sequences, half of which were grammatical. The ungrammatical sequences were split into three types of violation: within, between and both within and between chunk violations (see Table 2.1., Chapter 2). Each block only contained two of the three types of violations, and the types of violation that occurred in each block were balanced so that each violation type occurred with the other violation types equally. Each block was separated by a short exposure phase lasting 30 seconds, which was designed to re-familiarise the participants with the grammatical sequences. Finally, the WASI Block Design and Rapid Automatized Naming Digits and Objects tasks were administered.

In Experiment 4.4, we were not able to administer the standardised cognitive and language tasks, as this experiment took place online. Participants completed the same computer tasks as

Experiment 4.3 on their own desktop computer or laptop. To ensure participants were paying attention in the exposure phases, we included attention checks. Participants were instructed to press the 'space' key if they heard a 'click' sound during the exposure phase. In the initial exposure phase, 16 nonword or tone stimuli were randomly selected and replaced with a click sound. The same process was repeated for the short exposure phases, however only 3 nonword or tone stimuli were replaced by a click sound.

Data analysis

In Experiment 4.3, we compared scores on the standardised cognitive and language tasks across control and dyslexic participants which showed that controls scored significantly better than the dyslexic participants in the TOWRE Words ($t_{24} = 3.04$, $p = .006$) and Nonwords ($t_{24} = 6.11$, $p < .001$) and Backwards Digit Recall ($t_{24} = 2.49$, $p = .020$), and the difference between performance on the Rapid Automatized Naming Digits ($t_{24} = 1.86$, $p = .075$) and Objects ($t_{24} = 1.94$, $p = .065$) tasks was close to significant. Although this suggests that there was a difference in language ability between participants with and without dyslexia in this experiment, this is still based on a small sample size. As in Experiment 4.2, Experiment 4.4 was conducted online and therefore it was not possible to collect the same standardised cognitive and language task data.

The data were analysed in the same way for both Experiments 4.3 and 4.4. The Nonword and Tone AGL tasks were counterbalanced across participants to account for order effects. In the Grammaticality Judgement task, a trial was scored as correct if the participants successfully classified it as grammatical or ungrammatical, and performance on this task was compared to chance levels (50%). In the online experiment, if participants missed over 50% of clicks in any exposure phase, then they were removed from the analysis. Similarly to Experiments 4.1 and 4.2, the analyses showed no difference in performance across the tasks between participants who reported reading difficulties and those who reported that they had dyslexia. As in Experiment 4.2, we combined these groups into a single group with reading difficulties for analysis.

Results

In the in-person experiment (Experiment 4.3), controls performed above chance in both runs of the Nonword task (Run 1: $t_{15} = 9.75$, $p < .001$; Run 2: $t_{15} = 5.044$, $p < .001$) and the Tone task (Run 1: $t_{15} = 6.642$, $p < .001$; Run 2: $t_{15} = 4.85$, $p < .001$). The dyslexic group performed above chance in both runs of the Nonword task (Run 1: $t_9 = 4.961$, $p < .001$; Run 2: $t_9 = 3.807$,

$p < .0042$), and above chance in run 1 of the Tone task ($t_9 = 2.898$, $p = .018$), but not run 2 ($t_9 = 2.03$, $p = .055$), although performance in this run was approaching significance.

The findings were somewhat similar in the online version of the task comparing the performance of individuals with and without reading difficulties: individuals without reading difficulties performed above chance in both runs of the Nonword task (Run 1: $t_{44} = 12.84$, $p < .001$; Run 2: $t_{44} = 10.46$, $p < .001$), but above chance in run 2 of the Tone task only (Run 1: $t_{44} = 0.64$, $p = .526$; Run 2: $t_{44} = 3.84$, $p < .001$). Individuals with reading difficulties performed above chance in both runs of the Nonword (Run 1: $t_{53} = 2.87$, $p = .006$; Run 2: $t_{53} = 3.81$, $p < .001$), and Tone task (Run 1: $t_{53} = 2.87$, $p = .006$; Run 2: $t_{53} = 3.81$, $p < .001$). Taken together, these results suggest that learning occurred in both Nonword and Tone versions of the task in individuals with and without reading difficulties.

We aimed to compare performance of individuals with and without reading difficulties. In Experiment 4.4, we conducted a mixed 2x2x2 ANOVA with stimuli (nonword and tone), run (run 1 and 2) as within-subject factors, and group (control or dyslexic) as a between-subject factor. In the in-person Experiment, we found some evidence of a significant main effect of group ($F_{1,24} = 4.52$, $p = .044$), which indicates that there was a difference in performance on the nonword and tone tasks between the control and dyslexic participants. We found a main effect of stimuli ($F_{1,24} = 22.03$, $p < .001$), with participants performing better at the nonword task than the tone task. We found some evidence for a main effect of run ($F_{1,24} = 4.48$, $p = .045$), and a significant interaction between stimuli and run ($F_{1,24} = 5.96$, $p = .022$), which reflected a decrease in performance across runs for the nonword task but not the tone task. There were no other significant interactions. The findings from Experiment 4.3 may provide some suggestion of a domain-general deficit in implicit statistical learning in dyslexia; however, given the small sample size, it would be sensible to treat these conclusions with caution.

In Experiment 4.4, we conducted the same 2x2x2 ANOVA, except the between-subjects factor was changed to assess differences between controls and individuals with reading difficulties, rather than dyslexia specifically. We found a main effect of stimuli ($F_{1,97} = 237.961$, $p < .001$), with participants performing better in the Nonword than the Tone task, but no main effect of group ($F_{1,97} = .073$, $p = .787$), suggesting no difference in performance across runs or between individuals with and without reading difficulties. Whilst we found no main effect of run ($F_{1,97} = .718$, $p = .399$), there was a significant interaction between run and stimuli ($F_{1,97} = 18.01$, $p < .001$), which reflected a decrease in performance across runs in the Nonword task, but an increase in performance across runs in the tone task. There were no other significant interactions. The results from Experiment 4.4 suggest that there is no difference in performance between individuals with and without reading difficulties across the Nonword and Tone task and provide evidence against an implicit statistical learning deficit in individuals with reading difficulties.

To assess whether there was a difference in performance between the in-person and online experiments, we added an additional between-subjects factor of task (in-person or online) to the previous ANOVA. We found no main effect of task ($F_{1,121} = 3.10$, $p = .081$), however there was a significant interaction between task and stimuli ($F_{1,121} = 7.17$, $p = .008$), which reflected a decrease in performance between in-person and online experiments in the Tone AGL task but not the Nonword AGL task, which may suggest that performance in the Tone task is less suitable for online testing, perhaps due to factors outsider the experimenters control, such as differences in audio equipment across participants. Furthermore, we found a significant interaction between task and group, ($F_{1,121} = 3.96$, $p = .049$), although this only just reached significance. This interaction reflected poorer performance of controls in the online version of the task compared to the in-person task, whereas individuals with reading difficulties performance did not differ.

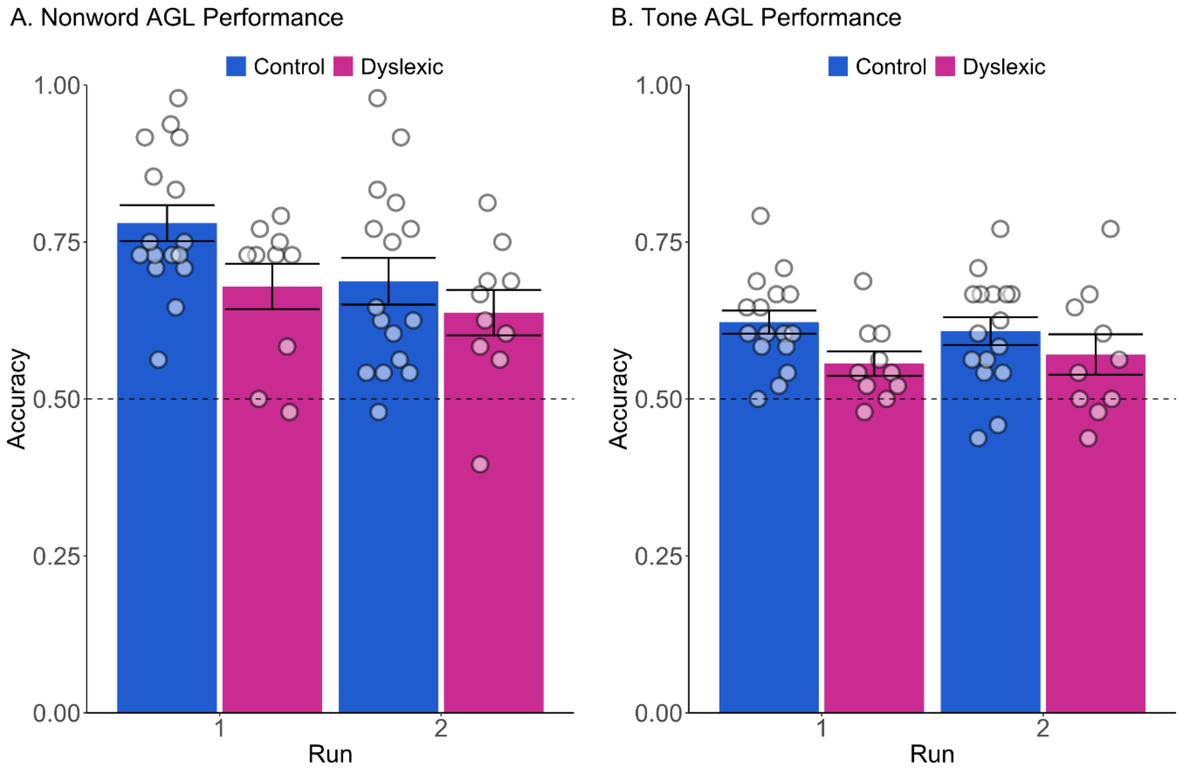


Figure 4.5. In-person Nonword and Tone auditory AGL performance. A) Nonword AGL performance for control (blue) and dyslexic (purple) participants. Individual performance is shown in circles. Both controls and individuals with dyslexia performed above chance (indicated by the dashed line) in both runs of the Nonword AGL task. B) Tone AGL performance for control (blue) and dyslexic (purple) participants. Individual performance is shown in circles. Both controls and individuals with dyslexia both performed above chance (indicated by the dashed line) in run 1 of the Tone AGL task, but only control participants performed above chance in run 2 of the tone task. There was some suggestion that control participants performed better than dyslexic participants across both Nonword and Tone AGL tasks.

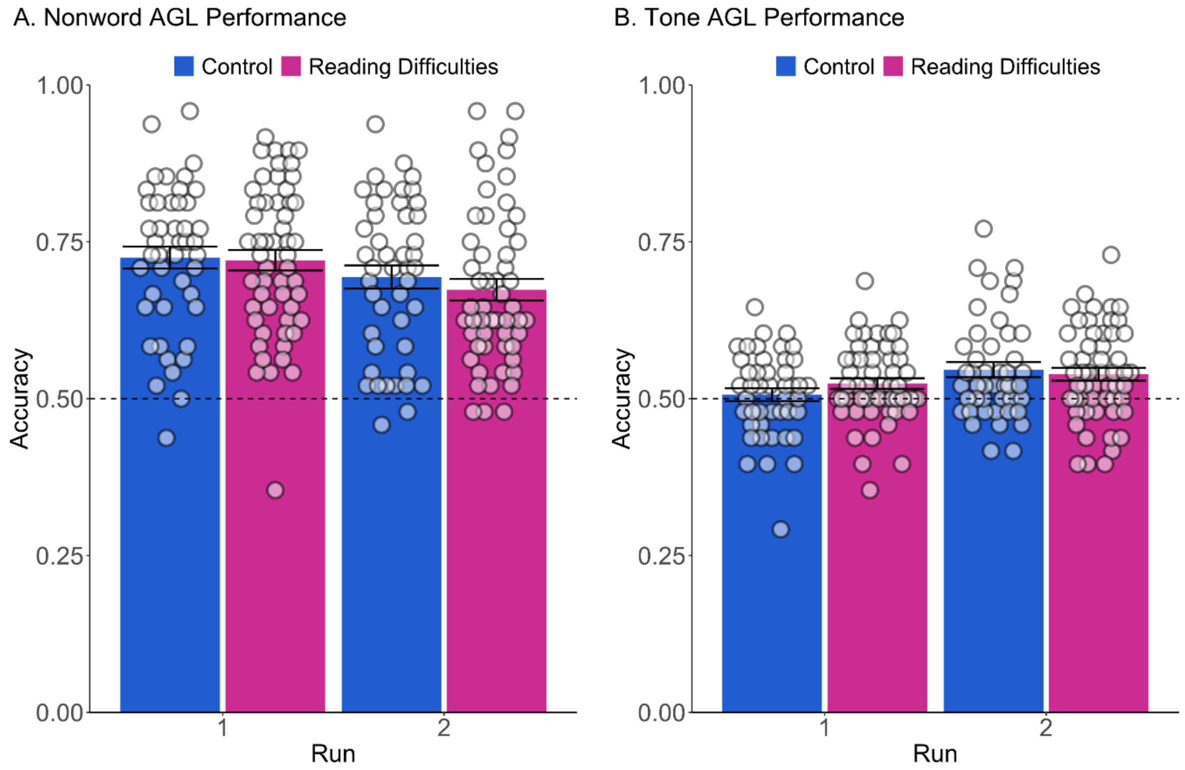


Figure 4.6. Online Nonword and Tone auditory AGL performance. A) Nonword AGL performance for control participants (blue) and participants with reading difficulties (purple). Individual performance is shown in circles. Both controls and individuals with reading difficulties performed above chance (indicated by the dashed line) in both runs of the Nonword AGL task. B) Tone AGL performance for control participants (blue) and participants with reading difficulties (purple). Individual performance is shown in circles. Both controls and individuals with reading difficulties performed above chance in run 2 of the Tone AGL task, however only individuals with reading difficulties performed above chance in run 1. Across both Nonword and Tone AGL tasks, there was no evidence of a difference in performance between individuals with and without reading difficulties.

To investigate whether an individual's performance is related to language ability, in Experiment 4.3 we correlated performance on the Nonword and Tone tasks with performance on the battery of standardised cognitive and language tasks (Figure 4.8). The language tasks were chosen because they require the rapid and/or precise manipulation of phonemes, an impairment in which is characteristic of dyslexia (Appendix 2.1). Performance in the Nonword task was correlated with performance on the TOWRE Word and Nonword tasks, as well as the Recalling Sentences task and Rapid Automatized Naming Objects task, with individuals who performed better in the Nonword AGL task performing better in the language tasks. Performance on the tone task was correlated with some of the language tasks: the TOWRE Word and Recalling Sentences tasks. These correlations may suggest that the Nonword and Tone AGL tasks are tapping into mechanisms that are similar to those that underpin performance in the standardised language tasks. Performance on the language tasks was more highly correlated with performance on the Nonword AGL task than the Tone AGL task; however, given that the Nonword AGL task used phonological stimuli, this is unsurprising.

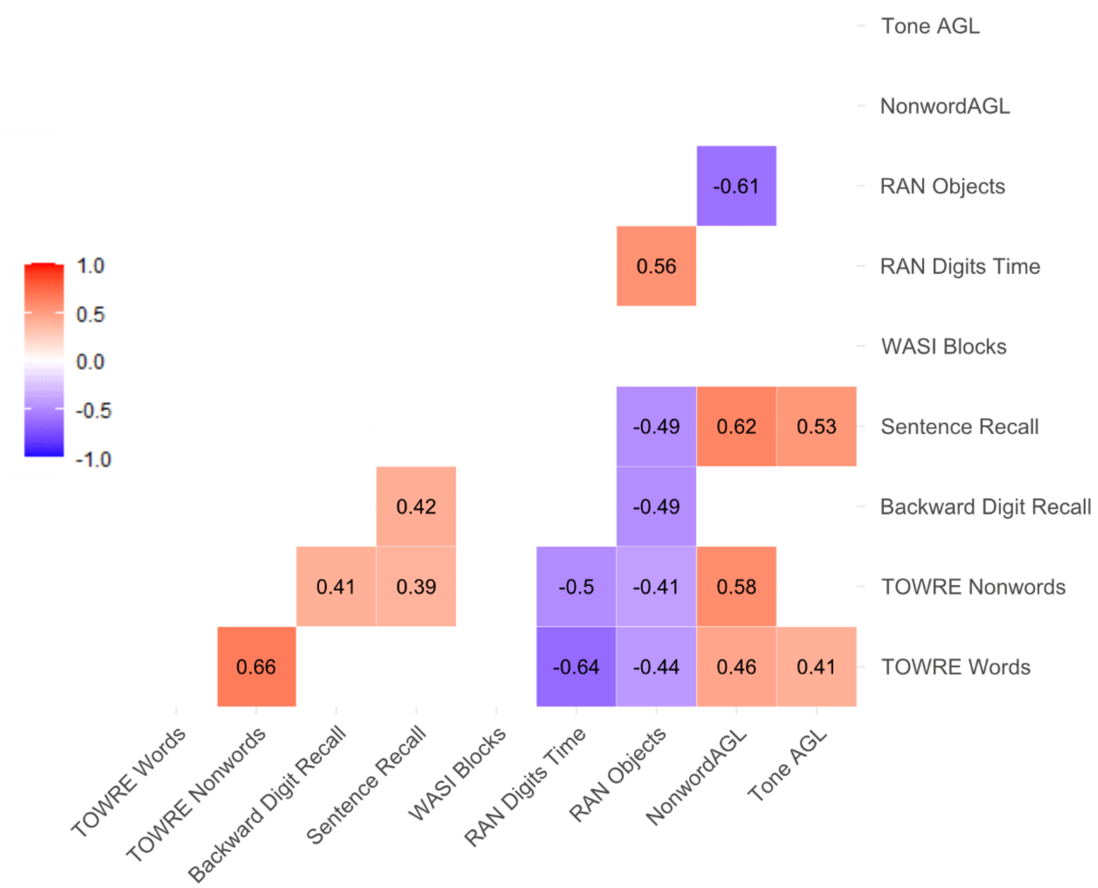


Figure 4.7. Correlations between Nonword and Tone AGL task performance and standardised cognitive and language tasks across participants with and without dyslexia. Only significant *r* values are shown ($p < 0.05$). The *r* values are colour-coded according to the correlation statistics, with negative correlations being shown in blue and positive correlations shown in red. Performance in the Nonword task was correlated with performance on the TOWRE Word and Nonword, Recalling Sentences and Rapid Automatized Naming Objects tasks, with individuals who performed better in the Nonword AGL task performing better in the language tasks. Performance on the Tone AGL task was correlated with the TOWRE Word and Recalling Sentences tasks. Performance in the Nonword and Tone AGL tasks were not correlated.

Discussion

In Experiments 4.3 and 4.4, we aimed to investigate whether dyslexia is associated with domain-specific differences in processing phonological information specifically, or a domain-general deficit in implicit statistical learning. In Experiment 4.3, we used a paradigm similar to Gabay et al. (2015) to test individuals with and without dyslexia in-person. There may have been some indication of a difference in performance between individuals with and without dyslexia across both nonword and tone tasks. However, due to the restrictions placed on in-person recruitment during the COVID-19 pandemic, the sample size for this study was much smaller than we had originally planned and therefore the study was underpowered. As such, we adapted the experiment to run online and recruited a larger sample. In Experiment 4.4, we compared the performance of individuals with and without reading difficulties and found no differences in performance across the Nonword and Tone AGL tasks between these groups. Overall, there is little evidence of a deficit in implicit statistical learning in individuals with reading difficulties, including dyslexia.

The findings from the in-person experiment somewhat mirror those of previous experiments using similar designs, which also found evidence of a domain-general implicit statistical learning deficit (Gabay et al., 2015). However, it is important to note that due to the issues surrounding in-person testing during the pandemic, a small number of participants were recruited in Experiment 4.3, meaning that this study is underpowered. Experiment 4.4 was conducted online, with a larger sample of individuals with and without reading difficulties and did not show any differences between groups. However, this experiment faces similar issues to Experiment 4.2: first, it is possible that the predicted differences in performance are specific to dyslexia and would not be reflected in the performance of individuals with reading difficulties who are not dyslexic. Second, we cannot confirm that the group with reading difficulties have poorer language ability compared to controls, as we have in Experiment 4.3.

Across both in-person and online versions of the experiment, participants perform better in the Nonword AGL task compared to the Tone AGL task. This does not reflect the findings from Gabay et al. (2015), where no difference was found between performance in Nonword and Tone AGL tasks. This is likely because in their experiment, the tones were presented at a slower rate compared to the nonwords, as previous research has suggested that a slower presentation rate is required to achieve comparable performance to nonwords (Saffran et al., 1999). In the current experiments, the presentation rate was kept consistent across Nonword and Tone AGL tasks, and differences in performance of participants across tasks highlights that there may be potential constraints on implicit statistical learning depending on the perceptual features of the stimuli (Conway, 2020; Conway & Christiansen, 2006).

The findings from Experiments 4.3 and 4.4 reflect the current state of the literature in that mixed findings are common when investigating the underlying causes of dyslexia. Indeed, regarding implicit statistical learning deficits, although there are a number of studies that suggest an impairment in implicit statistical learning in dyslexia (Katan et al., 2017; Pavlidou et al., 2009; Stoodley et al., 2008; Stoodley & Stein, 2006), there are as many that do not show any difference in implicit statistical learning between individuals with and without dyslexia (Inácio et al., 2018; Menghini et al., 2010; Nigro et al., 2016; Samara & Caravolas, 2017; Waber et al., 2003). Several meta-analyses have concluded that while there is some evidence of an implicit statistical learning deficit in dyslexia, there was also evidence of a publication bias, and that unpublished data may well null the result (van Witteloostuijn et al., 2017). A systematic review of this literature stressed the difficulty in drawing conclusions regarding implicit statistical learning deficits in dyslexia due to the lack of high quality data (Schmalz et al., 2017). The contradictory findings associated with the dyslexia literature also extend beyond those examining implicit statistical learning deficits. Meta-analyses of studies investigating two separate deficits in dyslexia - auditory processing (Witton et al., 2020) and visuo-spatial deficits (Chamberlain et al., 2018; Tafti et al., 2014) – both highlight that there is considerable heterogeneity within dyslexia. This heterogeneity may be an important factor that accounts for the mixed findings associated with most of the proposed deficits in dyslexia and suggests that dyslexia is unlikely to be the consequence of a single cause. It is more likely that there are multiple factors that interact in more complex ways to result in the reading impairments that characterise dyslexia (Castles et al., 2010).

Conclusions

The findings from the in-person and online Visual Recall tasks (Experiments 4.1 and 4.2 respectively) suggest that there is no deficit in implicit statistical learning. Experiment 4.3 provided some evidence of impaired implicit statistical learning in dyslexia when measuring learning using traditional AGL paradigms but was underpowered due to issues with recruitment during the COVID-19 pandemic. We did not replicate these findings in an online version of the task in Experiment 4.4 in which we recruited individuals with and without reading difficulties. Taken together, these experiments do not provide convincing evidence for an impairment in implicit statistical learning in dyslexia.

Chapter 5: General Discussion

Implicit statistical learning is thought to play an important role in language acquisition (Romberg & Saffran, 2010). While implicit statistical learning in adults and older children is typically measured using tasks that require explicit reflection regarding the grammaticality of a sequence, it can also be measured using processing-based tasks which do not require conscious reflection, and instead measure processes that are facilitated by implicit statistical learning (Isbilen et al., 2022; Isbilen et al., 2017, 2020; Kidd et al., 2020; Lammertink et al., 2019). The aim of this thesis was to develop and test novel processing-based measures of implicit statistical learning and combine these with traditional reflection-based tasks to further investigate the processes that occur during implicit statistical learning in children and adults with and without dyslexia.

To assess the efficacy of processing-based measures, in Chapter 2, I developed and tested both reaction time and serial visual recall tasks as measures of implicit statistical learning. Previous research suggested that implicit statistical learning can be measured using processing-based SRT-AGL tasks (Misyak & Christiansen, 2012; Misyak et al., 2009, 2010; Vuong et al., 2011) and serial recall (Isbilen et al., 2022; Isbilen et al., 2017, 2020; Kidd et al., 2020). The findings from Experiment 2.1 suggest that the SRT-AGL task was not an effective measure of learning. However, the Visual Recall task did show evidence of learning across a number of experiments, but only when the task required using active exposure.

The lack of learning in the SRT-AGL task may be due to participants not completing a passive exposure phase, which is thought to be an important feature of implicit statistical learning (Aslin, 2017). Instead, the SRT-AGL task contained active exposure in the form of a simple matching task. However, the data from the experiments using Visual Recall tasks suggest that active exposure may facilitate learning over passive exposure, as this method more accurately reflects how natural language is acquired (Heyes, 2018). These contradictory findings may be explained by differences in the tasks that participants were engaging with during active exposure. In the Visual Recall task, participants completed a cognitively demanding memory task, whereas the SRT-AGL task involved a less engaging matching task. Krishnan et al. (2021) have provided some evidence that passive and active exposure can lead to similar levels of performance; however, this study also demonstrated that knowledge acquired from passive exposure does not transfer during active recall. This may explain why in our experiments involving both passive and active exposure (Experiments 2.2 and 2.3), performance appears to be most influenced by active exposure. Taken together, these experiments suggest that active exposure can facilitate learning over passive exposure, but only when the task itself is engaging. Previous research demonstrated that serial recall can be used as a processing-based measure of highly regular transitions, such as those found within words (Isbilen et al., 2022; Isbilen et al., 2017, 2020; Kidd et al., 2020). The experiments within Chapter 2 indicate that the Visual Recall task can measure learning of grammars containing both highly regular and more variable transitions. Whilst it is generally understood that increasing complexity is associated with poorer performance in artificial grammar learning tasks (Pothos, 2010; Schiff & Katan, 2014), these experiments suggest that that the complexity of the grammar may also affect the way in which these regularities are processed during learning. Participants are more likely to gain explicit awareness of simpler predictable relationships, and this explicit awareness can be used to aid performance in the reflection-based tasks. Conversely, complex grammars containing more variable transitions are less likely to be learned explicitly, and therefore do not result in explicit knowledge that can be relied on in reflection-based tasks. Indeed, previous research has shown performance in artificial grammar learning tasks is facilitated by explicitly informing participants of the rules prior to exposure (Reber et al., 1980). Furthermore, Batterink, Reber and Paller (2015) showed that participants are more likely to rely on explicit knowledge if it is available. Given that the method by which learning occurs has important implications regarding learning, future research should focus on systematically varying complexity across experiments to further assess how such factors influence the cognitive processes underlying implicit statistical learning.

The findings from these experiments suggest that the Visual Recall task is a valuable processing-based measure of implicit statistical learning. However, the Visual Recall tasks from Chapter 2 focused on demonstrating that implicit statistical learning had occurred and provided less detail on what knowledge had been acquired. Instead, the subsequent reflection-based tasks were used to gain further insight into which regularities participants had become sensitive to during the Visual Recall task and suggested that participants learned the salient within chunk transitions but not the more variable between chunk transitions. Although future research should expand on these experiments to investigate whether serial visual recall can reveal more about what has been learned, these findings highlight the importance of utilising both processing- and reflection-based measures when investigating learning. Previous research has suggested that processing- and reflection-based measures tap into different processes (Isbilen et al., 2017; Lammertink et al., 2019), and therefore both types of tasks can provide useful information regarding the learning that occurs through implicit statistical learning.

Previous literature investigating the developmental trajectory of implicit statistical learning has yielded mixed results: some studies have suggested that implicit statistical learning is age-invariant (Raviv & Arnon, 2018; Saffran et al., 1997; Thiessen et al., 2013), while other studies have shown that implicit statistical learning abilities improve alongside other cognitive abilities (Arciuli & Simpson, 2011). It has also been suggested that implicit statistical learning is better in children compared to adults (Jost et al., 2015). To investigate implicit statistical learning in children, in Chapter 3, I recruited children aged 8 to 15 years to complete the Visual Recall and subsequent reflection-based tasks. There was evidence of implicit statistical learning across the tasks, and no difference in implicit statistical learning based on age within the sample of children. These findings suggest that implicit statistical learning abilities remain stable over this period of childhood. When comparing the performance of these children with adults, there was some indication that implicit statistical learning was poorer in children compared to adults in the Visual Recall task, but there was no difference in performance in the Grammaticality Judgement task. However, differences in implicit statistical learning may have been driven by differences in attention or motivation. After removing children who showed a pronounced grammaticality bias (and therefore either did not understand or engage with the task), there was little evidence of a difference in implicit statistical learning in the Visual Recall or Grammaticality Judgement tasks. These findings provide some suggestion that implicit statistical learning is an age-invariant mechanism (Kirkham et al., 2002; Raviv & Arnon, 2018; Saffran et al., 1997). If implicit statistical learning is age-invariant from childhood through to adulthood, then this may have important implications for the role of such mechanisms in second language learning, in which implicit statistical learning may play an important role (Frost et al., 2013; Godfroid & Kim, 2021; Onnis, 2012). Specifically, this may affect how second languages are taught, with an emphasis being placed on presenting the learner with input containing specific distributional properties to accelerate learning (Onnis, 2012; Smith, 1991).

The Visual Recall task revealed differences in the time-course of learning between children and adults that were not captured by the reflection-based tasks, in that there is no improvement in performance across the grammatical Recall Blocks in the Visual Recall task. Indeed, these differences remained even after removing the children who showed pronounced bias. Previous research has typically used SRT tasks to compare children and adult's performance on processing-based measures. These studies have generally shown that children and adults learn at a similar rate (Bertels et al., 2015; Du et al., 2017; Karatekin et al., 2007; Meulemans et al., 1998; Weiermann & Meier, 2012). It is possible that the children's lack of improvement across Recall Blocks can again be attributed to differences in attention and/or motivation that may arise from completing the experiment online. Children who were motivated and engaged with the task showed high levels of recall accuracy from the first Recall Block and continued to perform well across the task, meaning they do not show an improvement in performance across Recall Blocks. Children who showed poor recall accuracy did not improve across the Recall Blocks, which may be due to a lack of attention and/or motivation to complete the task. Although little research has been done comparing performance on in-person versus online implicit statistical learning tasks, there is some evidence suggesting that adults report more distractions when completing experiments in an online setting (Clifford & Jerit, 2014). As attention improves with age (Plude et al., 1994), it is sensible to suggest that children may be more affected by these additional distractions during online experiments compared to adults. Therefore, it would be useful to replicate this Experiment 3.1 in-person in order to determine whether any differences in the time-course of learning between children and adults are a consequence of running the study online.

The experiments with adults from Chapter 2 highlight that participants may be more likely to gain explicit awareness of more simple, predictable relationships compared to more complex relationships, which results in explicit knowledge that can be used to aid performance in the reflection-based tasks. However, it is currently unclear whether these findings extend to children. Previous research has indicated that similarly to adults, children can readily learn simple patterns (Cassar & Treiman, 1997; Fayol, 2002; Treiman, 1993), but more complex regularities take much longer to acquire (Caravolas et al., 2005). This may suggest that like adults, children may also rely on different forms of processing depending on the complexity of the grammar being learned. However, the findings from Experiment 3.1 and previous experiments (Bertels et al., 2015; Hickey et al., 2019; Meulemans et al., 1998) suggest that children are less likely to gain explicit awareness of any implicit learned information, which may indicate that children may be less likely to rely on explicit knowledge even when learning more simple relationships. Future research, preferably using longitudinal approaches, should examine whether there are any developmental differences in the mechanisms that support learning of regularities of varying complexity (Conway, 2020).

One of the key debates within dyslexia research relates to whether dyslexia is associated with a domain-specific deficit that is restricted to the processing of phonological stimuli (Griffiths & Snowling, 2002; Snowling, 1998; Vellutino et al., 2004), or whether a domain-general deficit in implicit statistical learning may underlie dyslexia (Folia et al., 2008; Gombert, 2003; Menghini et al., 2006; Ullman & Pierpont, 2005). In Chapter 4, I aimed to investigate the nature of implicit statistical learning deficits in dyslexia using the Visual Recall task and then subsequently assess whether any differences in dyslexia are restricted to processing of phonological stimuli. Across a number of in-person and online experiments, there was no difference in implicit statistical learning between individuals with and without dyslexia or reading difficulties. Although there were some limitations associated with conducting some of these experiments online (discussed below), these findings likely reflect the heterogeneity of dyslexia. In fact, the literature relating to the majority of the proposed deficits in dyslexia is contradictory. For example, there is evidence for and against differences in auditory processing (for a meta-analysis, see Witton et al., 2020), visuo-spatial processing (for a meta-analysis, see Chamberlain et al., 2018) and implicit statistical learning (for a meta-analysis, see van Witteloostuijn et al., 2017) in dyslexia. As such, the lack of a difference between individuals with and without dyslexia or reading difficulties in the experiments from Chapter 4 is not out of place within the wider literature and likely implies that any differences associated with dyslexia are caused by multiple interacting factors, rather than a single cause (Castles et al., 2010). Future research should move beyond characterising dyslexia in terms of a "core-deficit" and instead develop methods that capture the complexity that exists with and between the diagnostic boundaries of the disorder (Astle & Fletcher-Watson, 2020).

The conflicting findings within the literature investigating implicit statistical learning deficits in dyslexia specifically may also be a consequence of inadequate methods for measuring implicit statistical learning. Tasks that are developed to measure differences in implicit statistical learning between groups (e.g., between individuals with and without dyslexia) are characterised by poor reliability in adults (Bogaerts et al., 2018; Hedge et al., 2018; Siegelman, 2020; Siegelman & Frost, 2015), and this may be exacerbated in children (Arnon, 2019; West et al., 2018). The development of more reliable implicit statistical learning tasks has been recognised as an important avenue for research. Siegelman (2020) highlighted a number of factors that may improve reliability, most notably the development of processing-based tasks, which measure the time-course of learning across exposure by assessing changes in cognitive processes that are facilitated by implicit statistical learning. Therefore, the Visual Recall task may provide a promising method of reliably assessing implicit statistical learning and future research should extend these findings to directly assess the reliability of this task across populations.

Implicit statistical learning plays an important role in language acquisition (Romberg & Saffran, 2010). Although language learning is typically associated with the auditory domain, by adulthood, the language system is more multimodal, and many of the different facets of language (for example reading and listening) engage similar neural systems (Malik-Moraleda et al., 2022). Therefore, it is important to consider the role of implicit statistical learning across different sensory modalities in language processing. Recent multicomponential frameworks have highlighted that as well as domain general processes that occur similarly across modalities, there are also modality-specific constraints that influence the implicit statistical learning of regularities across domains (Conway, 2020; Frost et al., 2019; Frost et al., 2015). Therefore, examining implicit statistical learning, ideally using processing-based measures, across multiple domains is important in order to gain a comprehensive understanding of the nature of these mechanisms and their role in language acquisition and processing. Previous research has explored these processes in the auditory domain using speech stimuli (Isbilen et al., 2017). Here, I have adapted and developed these methods for the visual domain, and identified the conditions under which implicit statistical learning occurs. While directly comparing visual and auditory implicit statistical learning within the same experiment was beyond the scope of this project, such comparisons could represent a valuable future research direction. Moreover, this research has explored the role of implicit statistical learning in language acquisition, by taking a developmental approach to assess learning in children using the same tasks as adults. Future research should further investigate how implicit statistical learning contributes to language acquisition (including in those with language difficulties) using these tasks, for example, by examining the learning of more complex dependencies across development. Much recent discussion in the field of implicit statistical learning has focused on the importance of considering how implicit learning processes contribute to language learning and processing (Conway, 2020; Frost et al., 2019; Frost et al., 2015). The development of effective, processing-based measures of statistical learning represents an important step forward in resolving aspects of these debates, and clarifying the role of implicit statistical learning in language acquisition and processing.

These experiments also provide some interesting insight into the differences in conducting experiments online as opposed to in-person. Although this thesis was not planned to contain both in-person and online studies, the COVID-19 pandemic meant that experiments had to be adapted to run online in order to collect data. There were no differences in performance between in-person and online versions of the Visual Recall task when testing adults, which provides some suggestion that performance in visual experiments may not be affected by conducting the experiments online. However, when testing children, performance may have been affected by a lack of attention and/or motivation. However, from this data, it is unclear whether this was due to conducting the experiment online, or simply as facet of collecting data in children. There was some indication that participants performed more poorly in the online version of the auditory AGL task compared to the in-person experiment; however, this was specific to AGL task using tone stimuli. This may be due to differences in performance between tone and Nonword AGL tasks being exaggerated in the online experiment. For example, if the Tone AGL task is more difficult, it may become less engaging to complete. Therefore, when completing the experiment online, without an experimenter present, participants may be less motivated to maintain attention. As there were differences between in-person and online performance in auditory, but not visual tasks, this may suggest that performance in auditory tasks may not be comparable between in-person and online settings. Indeed, when comparing multiple online packages with equivalent in-person setups, the presentation of visual stimuli appears to be more precise compared to auditory stimuli, as currently none of the available online packages can provide precisely timed onsets for auditory stimuli (Bridges et al., 2020). Furthermore, previous research has highlighted the challenges associated with online auditory testing. It has been suggested that results may rely on factors that are outside of the experimenter's control, such as sound presentation level, or perceptual thresholds (Milne et al., 2021; Zhao et al., 2022), which may not affect the perception of nonwords. These findings suggest that visual tasks may provide more consistent results when conducting online experiments.

Although adapting these experiments to run online enabled me to continue collecting data in the midst of a global pandemic, there were some limitations associated with these experiments. Most notably, there were issues in the recruitment of individuals with dyslexia specifically for the online experiments. As discussed in Chapter 4, Prolific, the online platform I used for recruitment, only allowed for the recruitment of individuals with reading difficulties, rather than dyslexia specifically. Furthermore, there was no method of confirming that participants had reading difficulties, as the standardised cognitive and language tasks that would typically be administered to assess language ability cannot be conducted online. This presented a number of issues. First, and any deficits associated with dyslexia may not extend to those with reading difficulties more generally. As previously discussed (see Chapter 4, General Introduction), there is much debate regarding whether individuals with dyslexia and individuals with reading difficulties more generally are distinct populations (Badian, 1994; Elliott & Grigorenko, 2014; Gibbs & Elliott, 2020; Kirby, 2020; Stanovich, 1988). If implicit statistical learning deficits are specific to dyslexia, then it may not be sensible to expect poorer performance in participants with reading difficulties. Second, we were not able to administer the standardised cognitive and language tasks through Prolific, therefore it was not possible to determine whether participants with reading difficulties had poorer language ability than our control sample. If deficits in implicit statistical learning are associated with poorer language ability, then differences between individuals with and without reading difficulties across the tasks may not be expected if there is not difference in their language ability. Despite these limitations, online experiments provided a valuable tool for recruiting a large number of participants and enabled the collection of data during a time when in-person testing was not possible.

Conclusion

The aim of this research was to develop novel processing-based measures of implicit statistical learning and combine them with traditional reflection-based tasks to assess implicit statistical learning in children and adults with and without dyslexia. The findings from these experiments indicate that a multimodal approach, incorporating both processing- and reflection-based tasks, provides a valuable method of assessing implicit statistical learning across a range of populations. These methods have provided further detail on the processes that occur during implicit statistical learning, as well as the developmental trajectory of these mechanisms and the role they play in dyslexia.

References

Abla, D., & Okanoya, K. (2009). Visual statistical learning of shape sequences: An ERP study. *Neuroscience research, 64*(2), 185-190.

Ambridge, B., & Lieven, E. V. (2011). *Child language acquisition: Contrasting theoretical approaches*. Cambridge University Press.

Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child neuropsychology*, *8*(2), 71-82.

Arciuli, J. (2018). Reading as statistical learning. *Language, Speech, and Hearing Services in Schools*, *49*(3S), 634-643.

Arciuli, J., & Conway, C. M. (2018). The promise—and challenge—of statistical learning for elucidating atypical language development. *Current Directions in Psychological Science*, *27*(6), 492-500.

Arciuli, J., & Simpson, I. C. (2011). Statistical learning in typically developing children: The role of age and speed of stimulus presentation. *Developmental science*, *14*(3), 464-473.

Arciuli, J., & Simpson, I. C. (2012). Statistical learning is related to reading ability in children and adults. *Cognitive science*, *36*(2), 286-304.

Arciuli, J., & Torkildsen, J. v. K. (2012). Advancing our understanding of the link between statistical learning and language acquisition: The need for longitudinal data. *Frontiers in psychology*, *3*, 324.

Arnon, I. (2019). Statistical learning, implicit learning, and first language acquisition: A critical evaluation of two developmental predictions. *Topics in Cognitive Science*, *11*(3), 504-519.

Aslin, R. N. (2017). Statistical learning: a powerful mechanism that operates by mere exposure. *Wiley Interdisciplinary Reviews: Cognitive Science*, *8*(1-2), e1373.

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.

Astle, D. E., & Fletcher-Watson, S. (2020). Beyond the core-deficit hypothesis in developmental disorders. *Current Directions in Psychological Science*, *29*(5), 431-437.

Badian, N. A. (1994). Do dyslexic and other poor readers differ in reading-related cognitive skills? *Reading and writing*, *6*(1), 45-63.

Bahlmann, J., Schubotz, R. I., & Friederici, A. D. (2008). Hierarchical artificial grammar processing engages Broca's area. *Neuroimage*, *42*(2), 525-534.

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.

Batterink, L. J., Paller, K. A., & Reber, P. J. (2019). Understanding the neural bases of implicit and statistical learning. *Topics in Cognitive Science*, *11*(3), 482-503.

Batterink, L. J., Reber, P. J., Neville, H. J., & Paller, K. A. (2015). Implicit and explicit contributions to statistical learning. *Journal of Memory and Language*, *83*, 62-78.

Batterink, L. J., Reber, P. J., & Paller, K. A. (2015). Functional differences between statistical learning with and without explicit training. *Learning & memory*, *22*(11), 544-556.

Bertels, J., Boursain, E., Destrebecqz, A., & Gaillard, V. (2015). Visual statistical learning in children and young adults: how implicit? *Frontiers in psychology*, *5*, 1541.

Birdsong, D. (1999). *Second language acquisition and the critical period hypothesis*. Routledge.

Bogaerts, L., Siegelman, N., Ben-Porat, T., & Frost, R. (2018). Is the Hebb repetition task a reliable measure of individual differences in sequence learning? *Quarterly Journal of Experimental Psychology*, *71*(4), 892-905.

Bogen, J. E., & Bogen, G. (1976). Wernicke's region--Where is it. *Annals of the New York Academy of Sciences*(280), 834-843.

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: comparing a range of experiment generators, both lab-based and online. *PeerJ*, *8*, e9414.

Broca, P. (1861). Remarks on the seat of the faculty of articulated language, following an observation of aphemia (loss of speech). *Bulletin de la Société Anatomique*, *6*, 330-357.

Bulf, H., Johnson, S. P., & Valenza, E. (2011). Visual statistical learning in the newborn infant. *Cognition*, *121*(1), 127-132.

Caravolas, M., Kessler, B., Hulme, C., & Snowling, M. (2005). Effects of orthographic consistency, frequency, and letter knowledge on children's vowel spelling development. *Journal of Experimental Child Psychology*, *92*(4), 307-321.

Cassar, M., & Treiman, R. (1997). The beginnings of orthographic knowledge: Children's knowledge of double letters in words. *Journal of Educational Psychology*, *89*(4), 631.

Castles, A., MT McLean, G., & McArthur, G. (2010). Dyslexia (neuropsychological). *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(3), 426-432.

Catts, H. W., Adlof, S. M., Hogan, T. P., & Weismer, S. E. (2005). Are specific language impairment and dyslexia distinct disorders? *Journal of Speech, Language, and Hearing Research*, *48*, 1378–1396.

Chamberlain, R., Brunswick, N., Siev, J., & McManus, I. C. (2018). Meta-analytic findings reveal lower means but higher variances in visuospatial ability in dyslexia. *British Journal of Psychology*, *109*(4), 897-916.

Chan, C. (1991). *Implicit cognitive processes: theoretical issues and applications in computer systems design* University of Oxford].

Chan, S.-h., Ryan, L., & Bever, T. G. (2013). Role of the striatum in language: Syntactic and conceptual sequencing. *Brain and language*, *125*(3), 283-294.

Christiansen, M. H. (2019). Implicit statistical learning: A tale of two literatures. *Topics in Cognitive Science*, *11*(3), 468-481.

Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and brain sciences*, *39*.

Christiansen, M. H., Conway, C. M., & Onnis, L. (2012). Similar neural correlates for language and sequential learning: Evidence from event-related brain potentials. *Language and cognitive processes*, *27*(2), 231-256.

Christiansen, M. H., Kelly, M. L., Shillcock, R. C., & Greenfield, K. (2010). Impaired artificial grammar learning in agrammatism. *Cognition*, *116*(3), 382-393.

Cleary, M., Pisoni, D. B., & Geers, A. E. (2001). Some measures of verbal and spatial working memory in eight-and nine-year-old hearing-impaired children with cochlear implants. *Ear and hearing*, *22*(5), 395.

Cleary, M., Pisoni, D. B., & Kirk, K. I. (2000). Working memory spans as predictors of spoken word recognition and receptive vocabulary in children with cochlear implants. *The Volta Review*, *102*(4), 259.

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: News from the front. *Trends in cognitive sciences*, *2*(10), 406-416.

Clifford, S., & Jerit, J. (2014). Is there a cost to convenience? An experimental comparison of data quality in laboratory and online studies. *Journal of Experimental Political Science*, *1*(2), 120-131.

Conway, C. (2005). An odyssey through sight, sound, and touch: Toward a perceptual theory of implicit statistical learning.

Conway, C. M. (2020). How does the brain learn environmental structure? Ten core principles for understanding the neurocognitive mechanisms of statistical learning. *Neuroscience & Biobehavioral Reviews*, *112*, 279-299.

Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition*, *114*(3), 356-371.

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 24.
Conway, C. M., & Christiansen, M. H. (2006). Statistical learning within and between modalities: Pitting abstract against stimulus-specific representations. *Psychological science*, *17*(10), 905-912.
Conway, C. M., & Christiansen, M. H. (2009). Seeing and hearing in space and time: Effects of modality and presentation rate on implicit statistical learning. *European Journal of Cognitive Psychology*, *21*(4), 561-580.
Conway, C. M., Karpicke, J., & Pisoni, D. B. (2007). Contribution of implicit sequence learning to spoken language processing: Some preliminary findings with hearing adults. *Journal of deaf studies and deaf education*, *12*(3), 317-334.
Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science*, *19*(1), 51-57.
Curran, T. (1997). Effects of aging on implicit sequence learning: Accounting for sequence structure and explicit knowledge. *Psychological research*, *60*(1), 24-41.
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.
Demb, J. B., Poldrack, R. A., & Gabrieli, J. D. (1999). Functional neuroimaging of word processing in normal and dyslexic readers. *Converging Methods for Understanding Reading and Dyslexia. Cambridge*, 245–304.
Denckla, M. B., & Rudel, R. G. (1976). Rapid 'automatized'naming (RAN): Dyslexia differentiated from other learning disabilities. *Neuropsychologia*, *14*(4), 471-479.
Deocampo, J. A., Smith, G. N., Kronenberger, W. G., Pisoni, D. B., & Conway, C. M. (2018). The role of statistical learning in understanding and treating spoken language outcomes in deaf children with cochlear implants. *Language, Speech, and Hearing Services in Schools*, *49*(3S), 723-739.
Destrebecqz, A., & Cleeremans, A. (2001). Can sequence learning be implicit? New evidence with the process dissociation procedure. *Psychonomic bulletin & review*, *8*(2), 343-350.
Dienes, Z. (2012). Conscious versus unconscious learning of structure. *Statistical learning and language acquisition*, *1*, 337-364.
Dienes, Z., & Altmann, G. (1997). Transfer of implicit knowledge across domains: How implicit and how abstract. *How implicit is implicit learning*, *5*, 107-123.
Dienes, Z., Altmann, G., Kwan, L., & Goode, A. (1995). Unconscious knowledge of artificial grammars is applied strategically. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(5), 1322.
Dienes, Z., & Berry, D. (1997). Implicit learning: Below the subjective threshold. *Psychonomic bulletin & review*, *4*(1), 3-23.
Dienes, Z., Broadbent, D., & Berry, D. C. (1991). Implicit and explicit knowledge bases in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(5), 875.
Du, Y., Valentini, N. C., Kim, M. J., Whitall, J., & Clark, J. E. (2017). Children and adults both learn motor sequences quickly, but do so differently. *Frontiers in psychology*, *8*, 158.
Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & Van der Vrecken, O. (1996). The MBROLA project: Towards a set of high quality speech synthesizers free of use for non commercial purposes. Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP'96,
Elliott, J. G., & Grigorenko, E. L. (2014). *The dyslexia debate*. Cambridge University Press.
Emberson, L. L., Conway, C. M., & Christiansen, M. H. (2011). Timing is everything: Changes in presentation rate have opposite effects on auditory and visual implicit statistical learning. *Quarterly Journal of Experimental Psychology*, *64*(5), 1021-1040.

Erickson, L. C., Kaschak, M. P., Thiessen, E. D., & Berry, C. A. (2016). Individual differences in statistical learning: Conceptual and measurement issues. *Collabra*, *2*(1).

Evans, J. L., Saffran, J. R., & Robe-Torres, K. (2009). Statistical learning in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *52*, 321–335.

Facoetti, A., & Molteni, M. (2001). The gradient of visual attention in developmental dyslexia. *Neuropsychologia*, *39*(4), 352-357.

Facoetti, A., Paganoni, P., Turatto, M., Marzola, V., & Mascetti, G. G. (2000). Visual-spatial attention in developmental dyslexia. *Cortex*, *36*(1), 109-123.

Farmer, M. E., & Klein, R. M. (1995). The evidence for a temporal processing deficit linked to dyslexia: A review. *Psychonomic bulletin & review*, *2*(4), 460-493.

Farquharson, K., Centanni, T. M., Franzluebbers, C. E., & Hogan, T. P. (2014). Phonological and lexical influences on phonological awareness in children with specific language impairment and dyslexia. *Frontiers in psychology*, *5*, 838.

Fawcett, A. J., & Nicolson, R. I. (1995). The dyslexia early screening test. *The Irish Journal of Psychology*, *16*(3), 248-259.

Fawcett, A. J., Nicolson, R. I., & Dean, P. (1996). Impaired performance of children with dyslexia on a range of cerebellar tasks. *Annals of dyslexia*, *46*(1), 259-283.

Fayol, M. (2002). The acquisition of untaught orthographic regularities in French. *Precursors of functional literacy*, *11*, 121.

Fiser, J., & Aslin, R. N. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological science*, *12*(6), 499-504.

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

Folia, V., Uddén, J., Forkstam, C., Ingvar, M., Hagoort, P., & Petersson, K. M. (2008). Implicit learning and dyslexia. *Annals of the New York Academy of Sciences*, *1145*(1), 132-150.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in cognitive sciences*, *6*(2), 78-84.

Friederici, A. D. (2011). The brain basis of language processing: from structure to function. *Physiological reviews*, *91*(4), 1357-1392.

Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I., & Anwander, A. (2006). The brain differentiates human and non-human grammars: functional localization and structural connectivity. *Proceedings of the National Academy of Sciences*, *103*(7), 2458-2463.

Friederici, A. D., Friedrich, R., & Makuuchi, M. (2011). The neural basis of recursion and complex syntactic hierarchy. *Biolinguistics*, *5*(1-2), 087-104.

Friederici, A. D., Hahne, A., & Saddy, D. (2002). Distinct neurophysiological patterns reflecting aspects of syntactic complexity and syntactic repair. *Journal of psycholinguistic research*, *31*(1), 45-63.

Frost, R., Armstrong, B. C., & Christiansen, M. H. (2019). Statistical learning research: A critical review and possible new directions. *Psychological bulletin*, *145*(12), 1128.

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: the paradox of statistical learning. *Trends in cognitive sciences*, *19*(3), 117-125.

Frost, R., Siegelman, N., Narkiss, A., & Afek, L. (2013). What predicts successful literacy acquisition in a second language? *Psychological science*, *24*(7), 1243-1252.

Gabay, Y., Thiessen, E. D., & Holt, L. L. (2015). Impaired statistical learning in developmental dyslexia. *Journal of Speech, Language, and Hearing Research*, *58*(3), 934-945.

Gathercole, S. E., & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language*, *29*(3), 336-360.

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental psychology*, *40*(2), 177.

Gebhart, A. L., Newport, E. L., & Aslin, R. N. (2009). Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. *Psychonomic bulletin & review*, *16*(3), 486-490.

Gibbs, S. J., & Elliott, J. G. (2020). The dyslexia debate: life without the label. *Oxford Review of Education*, *46*(4), 487-500.

Gillis, J. U., Gul, A., Fox, A., Parikh, A., & Arbel, Y. (2022). Artificial grammar learning in children with developmental language disorder. *Journal of Speech, Language, and Hearing Research*, *65*(1), 292-302.

Godfroid, A., & Kim, K. M. (2021). The contributions of implicit-statistical learning aptitude to implicit second-language knowledge. *Studies in Second Language Acquisition*, *43*(3), 606-634.

Gombert, J.-E. (2003). Implicit and explicit learning to read: Implication as for subtypes of dyslexia. *Current psychology letters. Behaviour, brain & cognition*(10, Vol. 1, 2003).

Gómez, R., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy*, *7*(2), 183-206.

Gomez, R. L. (2002). Variability and detection of invariant structure. *Psychological science*, *13*(5), 431-436.

Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, *70*(2), 109-135.

Gómez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in cognitive sciences*, *4*(5), 178-186.

Griffiths, Y. M., & Snowling, M. J. (2002). Predictors of exception word and nonword reading in dyslexic children: The severity hypothesis. *Journal of Educational Psychology*, *94*(1), 34.

Hagoort, P. (2008). The fractionation of spoken language understanding by measuring electrical and magnetic brain signals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 1055-1069.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, *304*(5669), 438-441.

Hagoort, P., & van Berkum, J. (2007). Beyond the sentence given. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1481), 801-811.

Hahne, A., & Friederici, A. D. (2002). Differential task effects on semantic and syntactic processes as revealed by ERPs. *Cognitive Brain Research*, *13*(3), 339-356.

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, *50*(3), 1166-1186.

Heyes, C. (2018). Enquire within: Cultural evolution and cognitive science. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1743), 20170051.

Heyes, C., & Foster, C. (2002). Motor learning by observation: evidence from a serial reaction time task. *The Quarterly Journal of Experimental Psychology Section A*, *55*(2), 593-607.

Hickey, A., Hayiou-Thomas, E., & Mirkovic, J. (2019). Grammatical Generalisation in Statistical Learning: Is it implicit and invariant across development?. In *CogSci* (pp. 1908-1914).

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature reviews neuroscience*, *8*(5), 393-402.

Homae, F., Hashimoto, R., Nakajima, K., Miyashita, Y., & Sakai, K. L. (2002). From perception to sentence comprehension: the convergence of auditory and visual information of language in the left inferior frontal cortex. *Neuroimage*, *16*(4), 883-900.

Hsu, H. J., & Bishop, D. V. (2014). Sequence-specific procedural learning deficits in children with specific language impairment. *Developmental science*, *17*(3), 352-365.

Hunt, R. H., & Aslin, R. N. (2001). Statistical learning in a serial reaction time task: access to separable statistical cues by individual learners. *Journal of Experimental Psychology: General*, *130*(4), 658.

Inácio, F., Faísca, L., Forkstam, C., Araújo, S., Bramão, I., Reis, A., & Petersson, K. M. (2018). Implicit sequence learning is preserved in dyslexic children. *Annals of dyslexia*, *68*(1), 1-14.

Isbilen, E. S., McCauley, S. M., & Christiansen, M. H. (2022). Individual differences in artificial and natural language statistical learning. *Cognition*, *225*, 105123.

Isbilen, E. S., McCauley, S. M., Kidd, E., & Christiansen, M. H. (2017). Testing statistical learning implicitly: A novel chunk-based measure of statistical learning. the 39th Annual Conference of the Cognitive Science Society (CogSci 2017),

Isbilen, E. S., McCauley, S. M., Kidd, E., & Christiansen, M. H. (2020). Statistically induced chunking recall: A memory-based approach to statistical learning. *Cognitive science*, *44*(7), e12848.

Janacsek, K., Fiser, J., & Nemeth, D. (2012). The best time to acquire new skills: Age-related differences in implicit sequence learning across the human lifespan. *Developmental science*, *15*(4), 496-505.

Jones, M. W., Branigan, H. P., & Kelly, M. L. (2009). Dyslexic and nondyslexic reading fluency: Rapid automatized naming and the importance of continuous lists. *Psychonomic bulletin & review*, *16*(3), 567-572.

Jost, E., Conway, C. M., Purdy, J. D., Walk, A. M., & Hendricks, M. A. (2015). Exploring the neurodevelopment of visual statistical learning using event-related brain potentials. *Brain Research*, *1597*, 95-107.

Kahta, S., & Schiff, R. (2016). Implicit learning deficits among adults with developmental dyslexia. *Annals of dyslexia*, *66*(2), 235-250.

Kahta, S., & Schiff, R. (2019). Deficits in statistical leaning of auditory sequences among adults with dyslexia. *Dyslexia*, *25*(2), 142-157.

Karatekin, C., Marcus, D. J., & White, T. (2007). Oculomotor and manual indexes of incidental and intentional spatial sequence learning during middle childhood and adolescence. *Journal of Experimental Child Psychology*, *96*(2), 107-130.

Karuza, E. A., Newport, E. L., Aslin, R. N., Starling, S. J., Tivarus, M. E., & Bavelier, D. (2013). The neural correlates of statistical learning in a word segmentation task: An fMRI study. *Brain and language*, *127*(1), 46-54.

Katan, P., Kahta, S., Sasson, A., & Schiff, R. (2017). Performance of children with developmental dyslexia on high and low topological entropy artificial grammar learning task. *Annals of dyslexia*, *67*(2), 163-179.

Kelly, S. W., Griffiths, S., & Frith, U. (2002). Evidence for implicit sequence learning in dyslexia. *Dyslexia*, *8*(1), 43-52.

Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental psychology*, *48*(1), 171.

Kidd, E., & Arciuli, J. (2016). Individual differences in statistical learning predict children's comprehension of syntax. *Child development*, *87*(1), 184-193.

Kidd, E., Arciuli, J., Christiansen, M. H., Isbilen, E. S., Revius, K., & Smithson, M. (2020). Measuring children’s auditory statistical learning via serial recall. *Journal of Experimental Child Psychology*, *200*, 104964.

Kirby, P. (2020). Dyslexia debated, then and now: A historical perspective on the dyslexia debate. *Oxford Review of Education*, *46*(4), 472-486.

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.

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.

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

Knowlton, B. J., & Squire, L. R. (1994). The information acquired during artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(1), 79.

Krishnan, S., Carey, D., Dick, F., & Pearce, M. T. (2021). Effects of statistical learning in passive and active contexts on reproduction and recognition of auditory sequences. *Journal of Experimental Psychology: General*, *151*(3), 555.

Lammertink, I., Boersma, P., Wijnen, F., & Rispens, J. (2017). Statistical learning in specific language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, *60*(12), 3474-3486.

Lammertink, I., Boersma, P., Wijnen, F., & Rispens, J. (2020). Children with developmental language disorder have an auditory verbal statistical learning deficit: Evidence from an online measure. *Language Learning*, *70*(1), 137-178.

Lammertink, I., Van Witteloostuijn, M., Boersma, P., Wijnen, F., & Rispens, J. (2019). Auditory statistical learning in children: Novel insights from an online measure. *Applied Psycholinguistics*, *40*(2), 279-302.

Lukács, Á., & Kemény, F. (2015). Development of different forms of skill learning throughout the lifespan. *Cognitive science*, *39*(2), 383-404.

Lukács, Á., Lukics, K. S., & Dobó, D. (2021). Online Statistical Learning in Developmental Language Disorder. *Frontiers in Human Neuroscience*, 541.

Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2003). A definition of dyslexia. *Annals of dyslexia*, *53*(1), 1-14.

Malik-Moraleda, S., Ayyash, D., Gallée, J., Affourtit, J., Hoffmann, M., Mineroff, Z., Jouravlev, O., & Fedorenko, E. (2022). An investigation across 45 languages and 12 language families reveals a universal language network. *Nature neuroscience*, *25*(8), 1014-1019.

Manning, C., Hassall, C. D., Hunt, L. T., Norcia, A. M., Wagenmakers, E.-J., Snowling, M. J., Scerif, G., & Evans, N. J. (2022). Visual motion and decision-making in dyslexia: Reduced accumulation of sensory evidence and related neural dynamics. *Journal of Neuroscience*, *42*(1), 121-134.

Marzocchi, G. M., Ornaghi, S., & Barboglio, S. (2009). What are the causes of the attention deficits observed in children with dyslexia? *Child neuropsychology*, *15*(6), 567-581.

McCauley, S. M. & Christiansen, M. H. (2015b) Individual differences in chunking ability predict on-line sentence processing. In: Proceedings of the 37th Annual Conference of the Cognitive Science Society, Pasadena, CA, July 2015, pp. 1553–58, ed. D. C. Noelle & R. Dale. Cognitive Science Society.

Melby-Lervåg, M., Lervåg, A., Lyster, S.-A. H., Klem, M., Hagtvet, B., & Hulme, C. (2012). Nonword-repetition ability does not appear to be a causal influence on children's vocabulary development. *Psychological science*, *23*(10), 1092-1098.

Menghini, D., Finzi, A., Benassi, M., Bolzani, R., Facoetti, A., Giovagnoli, S., Ruffino, M., & Vicari, S. (2010). Different underlying neurocognitive deficits in developmental dyslexia: a comparative study. *Neuropsychologia*, *48*(4), 863-872.

Menghini, D., Hagberg, G. E., Caltagirone, C., Petrosini, L., & Vicari, S. (2006). Implicit learning deficits in dyslexic adults: An fMRI study. *Neuroimage*, *33*(4), 1218-1226.

Meulemans, T., Van der Linden, M., & Perruchet, P. (1998). Implicit sequence learning in children. *Journal of Experimental Child Psychology*, *69*(3), 199-221.

Meyer, P., Mecklinger, A., Grunwald, T., Fell, J., Elger, C. E., & Friederici, A. D. (2005). Language processing within the human medial temporal lobe. *Hippocampus*, *15*(4), 451-459.

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

Milne, A. E., Bianco, R., Poole, K. C., Zhao, S., Oxenham, A. J., Billig, A. J., & Chait, M. (2021). An online headphone screening test based on dichotic pitch. *Behavior Research Methods*, *53*(4), 1551-1562.

Milne, A. E., Petkov, C. I., & Wilson, B. (2018). Auditory and visual sequence learning in humans and monkeys using an artificial grammar learning paradigm. *Neuroscience*, *389*, 104-117.

Misyak, J. B., & Christiansen, M. H. (2012). Statistical learning and language: An individual differences study. *Language Learning*, *62*(1), 302-331.

Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2009). Statistical learning of nonadjacencies predicts on-line processing of long-distance dependencies in natural language. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 31, No. 31).

Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). On-line individual differences in statistical learning predict language processing. *Frontiers in psychology*, *1*, 31.

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive science*, *14*(1), 11-28.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive psychology*, *48*(2), 127-162.

Nigro, L., Jiménez-Fernández, G., Simpson, I. C., & Defior, S. (2016). Implicit learning of non-linguistic and linguistic regularities in children with dyslexia. *Annals of dyslexia*, *66*(2), 202-218.

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

Obeid, R., Brooks, P. J., Powers, K. L., Gillespie-Lynch, K., & Lum, J. A. (2016). Statistical learning in specific language impairment and autism spectrum disorder: A meta-analysis. *Frontiers in psychology*, *7*, 1245.

Ojima, S., & Okanoya, K. (2020). Children’s learning of a semantics-free artificial grammar with center embedding. *Biolinguistics*, *14*, 21-48.

Onnis, L. (2012). The potential contribution of statistical learning to second language acquisition. *Statistical learning and language acquisition*, 203-235.

Osterhout, L., Holcomb, P. J., & Swinney, D. A. (1994). Brain potentials elicited by garden-path sentences: evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 786.

Osugi, T., & Takeda, Y. (2013). The precision of visual memory for a complex contour shape measured by a freehand drawing task. *Vision research*, *79*, 17-26.

Pacton, S., & Perruchet, P. (2008). An attention-based associative account of adjacent and nonadjacent dependency learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(1), 80.

Pavlidou, E. V., Williams, J. M., & Kelly, L. M. (2009). Artificial grammar learning in primary school children with and without developmental dyslexia. *Annals of dyslexia*, *59*(1), 55-77.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical learning in a natural language by 8-month-old infants. *Child development*, *80*(3), 674-685.

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

Petersson, K.-M., Folia, V., & Hagoort, P. (2012). What artificial grammar learning reveals about the neurobiology of syntax. *Brain and language*, *120*(2), 83-95.

Plude, D. J., Enns, J. T., & Brodeur, D. (1994). The development of selective attention: A life-span overview. *Acta psychologica*, *86*(2-3), 227-272.
Pothos, E. (2010). An entropy model for artificial grammar learning. *Frontiers in psychology, 1*, 16.
Qi, Z., Sanchez Araujo, Y., Georgan, W. C., Gabrieli, J. D., & Arciuli, J. (2019). Hearing matters more than seeing: A cross-modality study of statistical learning and reading ability. *Scientific Studies of Reading*, *23*(1), 101-115.
Raviv, L., & Arnon, I. (2017). Differences between children and adults in the emergence of linguistic structure. the 30th Annual CUNY Conference on Human Sentence Processing,
Raviv, L., & Arnon, I. (2018). The developmental trajectory of children's auditory and visual statistical learning abilities: Modality-based differences in the effect of age. *Developmental science, 21*(4), e12593.
Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of verbal learning and verbal behavior*, *6*(6), 855-863.
Reber, A. S. (1976). Implicit learning of synthetic languages: The role of instructional set. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(1), 88.
Reber, A. S., Kassin, S. M., Lewis, S., & Cantor, G. (1980). On the relationship between implicit and explicit modes in the learning of a complex rule structure. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(5), 492.
Reber, P. J., & Squire, L. R. (1994). Parallel brain systems for learning with and without awareness. *Learning & memory*, *1*(4), 217-229.
Rispens, J., & Been, P. (2007). Subject–verb agreement and phonological processing in developmental dyslexia and specific language impairment (SLI): a closer look. *International Journal of Language & Communication Disorders, 42*(3), 293-305.
Robertson, E. M. (2007). The serial reaction time task: implicit motor skill learning? *Journal of Neuroscience*, *27*(38), 10073-10075.
Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(6), 906-914.
Roongpraiwan, R., Ruangdaraganon, N., Visudhiphan, P., & Santikul, K. (2002). Prevalence and clinical characteristics of dyslexia in primary school students. *Journal of the Medical Association of Thailand= Chotmaihet thangphaet, 85*, S1097-1103.
Rowe, V. (1992). Language Development in Adolescence and Beyond.
Saffran, J., Hauser, M., Seibel, R., Kapfhamer, J., Tsao, F., & Cushman, F. (2008). Grammatical pattern learning by human infants and cotton-top tamarin monkeys. *Cognition*, *107*(2), 479-500.
Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language*, *47*(1), 172-196.
Saffran, J. R. (2018). Statistical learning as a window into developmental disabilities. *Journal of Neurodevelopmental Disorders*, *10*(1), 1-5.
Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926-1928.
Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27-52.
Saffran, J. R., & Kirkham, N. Z. (2018). Infant statistical learning. *Annual review of psychology*, *69*, 181-203.
Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological science*, *8*(2), 101-105.
Sahin, N. T., Pinker, S., Cash, S. S., Schomer, D., & Halgren, E. (2009). Sequential processing of lexical, grammatical, and phonological information within Broca’s area. *Science*, *326*(5951), 445-449.

Samara, A., & Caravolas, M. (2017). Artificial grammar learning in dyslexic and nondyslexic adults: Implications for orthographic learning. *Scientific Studies of Reading*, *21*(1), 76-97.

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.

Schiff, R., & Katan, P. (2014). Does complexity matter? Meta-analysis of learner performance in artificial grammar tasks. *Frontiers in psychology*, *5*, 1084.

Schmalz, X., Altoè, G., & Mulatti, C. (2017). Statistical learning and dyslexia: A systematic review. *Annals of dyslexia*, *67*(2), 147-162.

Schmalz, X., Moll, K., Mulatti, C., & Schulte-Körne, G. (2019). Is statistical learning ability related to reading ability, and if so, why? *Scientific Studies of Reading*, *23*(1), 64-76.

Scott, R. B., & Dienes, Z. (2008). The conscious, the unconscious, and familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(5), 1264.

Seitz, A. R., Kim, R., van Wassenhove, V., & Shams, L. (2007). Simultaneous and independent acquisition of multisensory and unisensory associations. *Perception*, *36*(10), 1445-1453.

Siegelman, N. (2020). Statistical learning abilities and their relation to language. *Language and Linguistics Compass*, *14*(3), e12365.

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.

Siegelman, N., & Frost, R. (2015). Statistical learning as an individual ability: Theoretical perspectives and empirical evidence. *Journal of Memory and Language*, *81*, 105-120.

Smith, M. S. (1991). Speaking to many minds: On the relevance of different types of language information for the L2 learner. *Interlanguage studies bulletin (Utrecht)*, *7*(2), 118-132.

Snowling, M. (1998). Dyslexia as a phonological deficit: Evidence and implications. *Child Psychology and Psychiatry Review*, *3*(1), 4-11.

Spencer, M., Kaschak, M. P., Jones, J. L., & Lonigan, C. J. (2015). Statistical learning is related to early literacy-related skills. *Reading and writing*, *28*(4), 467-490.

Stanovich, K. E. (1988). Explaining the differences between the dyslexic and the garden-variety poor reader: The phonological-core variable-difference model. *Journal of learning disabilities*, *21*(10), 590-604.

Stefanac, N. R., Zhou, S.-H., Spencer-Smith, M. M., O'Connell, R., & Bellgrove, M. A. (2021). A neural index of inefficient evidence accumulation in dyslexia underlying slow perceptual decision making. *Cortex*, *142*, 122-137.

Stein, J. (2001). The magnocellular theory of developmental dyslexia. *Dyslexia*, *7*(1), 12-36.

Stein, J. (2019). The current status of the magnocellular theory of developmental dyslexia. *Neuropsychologia*, *130*, 66-77.

Stoodley, C. J., Ray, N. J., Jack, A., & Stein, J. F. (2008). Implicit learning in control, dyslexic, and garden-variety poor readers. *Annals of the New York Academy of Sciences*, *1145*(1), 173-183.

Stoodley, C. J., & Stein, J. F. (2006). A processing speed deficit in dyslexic adults? Evidence from a peg-moving task. *Neuroscience Letters*, *399*(3), 264-267.

Tabullo, Á., Sevilla, Y., Segura, E., Zanutto, S., & Wainselboim, A. (2013). An ERP study of structural anomalies in native and semantic free artificial grammar: evidence for shared processing mechanisms. *Brain Research*, *1527*, 149-160.

Tafti, M. A., Boyle, J. R., & Crawford, C. M. (2014). Meta-analysis of visual-spatial deficits in dyslexia. *International Journal of Brain and Cognitive Sciences*, *3*(1), 25-34.

Tallal, P. (1984). Temporal or phonetic processing deficit in dyslexia? That is the question. *Applied Psycholinguistics*, *5*(2), 167-169.

Thiessen, E. D., Kronstein, A. T., & Hufnagle, D. G. (2013). The extraction and integration framework: a two-process account of statistical learning. *Psychological bulletin*, *139*(4), 792.

Thomas, K. M., Hunt, R. H., Vizueta, N., Sommer, T., Durston, S., Yang, Y., & Worden, M. S. (2004). Evidence of developmental differences in implicit sequence learning: an fMRI study of children and adults. *Journal of cognitive neuroscience*, *16*(8), 1339-1351.

Tillmann, B., & McAdams, S. (2004). Implicit learning of musical timbre sequences: statistical regularities confronted with acoustical (dis) similarities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(5), 1131.

Toro, J. M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, *97*(2), B25-B34.

Treiman, R. (1993). *Beginning to spell: A study of first-grade children*. Oxford University Press on Demand.

Tunney, R. J., & Altmann, G. (2001). Two modes of transfer in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(3), 614.

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.

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of cognitive neuroscience*, *21*(10), 1934-1945.

Uddén, J., Araujo, S., Forkstam, C., Ingvar, M., Hagoort, P., & Petersson, K. M. (2009). A matter of time: Implicit acquisition of recursive sequence structures. In *31st Annual Conference of the Cognitive Science Society* (pp. 2444-2449).

Ullman, M. T., & Pierpont, E. I. (2005). Specific language impairment is not specific to language: The procedural deficit hypothesis. *Cortex*, *41*(3), 399-433.

Vaidya, C. J., Huger, M., Howard, D. V., & Howard Jr, J. H. (2007). Developmental differences in implicit learning of spatial context. *Neuropsychology*, *21*(4), 497.

van der Lely, H. K., Jones, M., & Marshall, C. R. (2011). Who did Buzz see someone? Grammaticality judgement of wh-questions in typically developing children and children with Grammatical-SLI. *Lingua*, *121*(3), 408-422.

van Witteloostuijn, M., Boersma, P., Wijnen, F., & Rispens, J. (2017). Visual artificial grammar learning in dyslexia: A meta-analysis. *Research in Developmental Disabilities*, *70*, 126-137.

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): What have we learned in the past four decades? *Journal of child psychology and psychiatry*, *45*(1), 2-40.

Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S. G., Pratt, A., Chen, R., & Denckla, M. B. (1996). Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology*, *88*(4), 601.

Vellutino, F.R., Scanlon, D.M., & Tanzman, M.S. (1994). Components of reading ability: Issues and problems in operationalizing word identification, phonological coding, and orthographic coding. In G.R. Lyon (Ed.), Frames of reference for the assessment of learning disabilities: New views on measurement issues (pp. 279–324). Baltimore, MD: Paul H. Brookes

Vigneau, M., Beaucousin, V., Hervé, P.-Y., Duffau, H., Crivello, F., Houde, O., Mazoyer, B., & Tzourio-Mazoyer, N. (2006). Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. *Neuroimage*, *30*(4), 1414-1432.

Vogel, E. K., Woodman, G. F., & Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of experimental psychology: human perception and performance*, *27*(1), 92.

von Koss Torkildsen, J., Arciuli, J., & Wie, O. B. (2019). Individual differences in statistical learning predict children's reading ability in a semi-transparent orthography. *Learning and individual differences*, *69*, 60-68.

Vuong, L., Meyer, A., & Christiansen, M. (2011). Simultaneous online tracking of adjacent and non-adjacent dependencies in statistical learning. Proceedings of the Annual Meeting of the Cognitive Science Society,

Waber, D. P., Marcus, D. J., Forbes, P. W., Bellinger, D. C., Weiler, M. D., Sorensen, L. G., & Curran, T. (2003). Motor sequence learning and reading ability: Is poor reading associated with sequencing deficits? *Journal of Experimental Child Psychology*, *84*(4), 338-354.

Weiermann, B., & Meier, B. (2012). Incidental sequence learning across the lifespan. *Cognition*, *123*(3), 380-391.

West, G., Vadillo, M. A., Shanks, D. R., & Hulme, C. (2018). The procedural learning deficit hypothesis of language learning disorders: We see some problems. *Developmental science*, *21*(2), e12552.

Wilkinson, L., & Shanks, D. R. (2004). Intentional control and implicit sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 354.

Willingham, D. B., & Goedert-Eschmann, K. (1999). The relation between implicit and explicit learning: Evidence for parallel development. *Psychological science*, *10*(6), 531-534.

Wilson, B., Smith, K., & Petkov, C. I. (2015). Mixed-complexity artificial grammar learning in humans and macaque monkeys: evaluating learning strategies. *European Journal of Neuroscience*, *41*(5), 568-578.

Winkler, M., Mueller, J. L., Friederici, A. D., & Männel, C. (2018). Infant cognition includes the potentially human-unique ability to encode embedding. *Science Advances*, *4*(11), eaar8334.

Witton, C., Swoboda, K., Shapiro, L. R., & Talcott, J. B. (2020). Auditory frequency discrimination in developmental dyslexia: A meta-analysis. *Dyslexia*, *26*(1), 36-51.

Zhao, S., Brown, C. A., Holt, L. L., & Dick, F. (2022). Robust and Efficient Online Auditory Psychophysics. *Trends in hearing*, *26*, 23312165221118792.

Zurif, E. B. (1980). Language Mechanisms: A Neuropsychological Perspective: The effects of focal brain damage on the processing of syntactic elements may provide an important clue to the manner in which language is organized in the brain. *American Scientist*, *68*(3), 305-311.

Appendix

Appendix 2.1. Standardised Cognitive and Language Tasks

Test of Word Reading Efficiency (TOWRE): Words and Nonwords

These tasks are a popular measure of an individual's ability to pronounce printed words and phonemically regular nonwords accurately and fluently (Torgesen, Wagner & Rashotte, 2012). The TOWRE consists of two tests: Sight Word Efficiency (SWE), which measures the ability to read familiar words aloud quickly, and Phonemic Decoding Efficiency (PDE), which measures proficiency in sounding out nonwords. The SWE task consists of 108 items in total, arranged into four columns. The PDE consists of 66 items arranged into three columns. Both lists were presented sequentially on an A4 sheet of paper, and the participant was asked to read as many of the items as possible in 45 seconds.

Backwards Digit Recall

Backwards digit recall (taken from the Wechsler Adult Intelligence Scale (WAIS); Wechsler, 2008) was used as a measure of verbal working memory. Although both forward and backward digit span tasks rely on short-term memory (Rosen & Engle, 1997), backwards digit span also recruits working memory systems (Alloway *et al.*, 2006). For this task, participants were instructed to listen to the experimenter say a sequence of numbers and repeat these numbers in reverse order. The task consisted of 6 blocks of numbers; each block contained 6 sequences of digits ranging from 2 to 7 digits in length. Participants scored 1 mark per correct response, and a discontinue rule was implemented: the participants had to score 4 marks per block to progress to the next block.

Recalling Sentences

The recalling sentences subtask was taken from the Clinical Evaluation of Language Fundamentals (CELF-5; Wiig, Secord & Semel, 2013), and requires the participant to repeat sentences of increasing length and grammatical complexity. There are 26 sentences in total, however as all of the participants in this study were older than 15 years, only sentences 16 through to 26 were administered. The maximum score for each sentence is 3 points, and for each mistake the participant makes one point is deducted until they reach 0 points.

WASI Block Design

We assessed nonverbal intelligence using the Block Design task from the Weschler Scale of Abbreviated Intelligence, Second Edition (WASI-II; Weschler, 2011). The task consists of 13 trials in which the participant is shown two-dimensional red and white geometric designs, and then asked to recreate each design using the top of red and white cubes. The task is standardised for ages 6 to 90 years. Each item has a specified time limit, and participants are given higher scores for completing the trials more quickly. Trials are scored as 0 if the design is not successfully recreated. The task is discontinued if the participants fail 2 consecutive trials.

Rapid Automatised Naming (RAN)

RAN was used as a measure of processing speed in two domains: digit and object processing. Therefore, two types of RAN task, taken from the York Adult Assessment Battery – Revised (YAA-R; Warmington, Stothard & Snowling, 2013) were used in this study. The RAN digits task consisted of 10 practice digits and 50 test digits, arranged in a 5x10 matrix on an A4 sheet of paper. The RAN objects consisted of a selection of four line drawings of objects (duck, shoe, car, frog) which were organised randomly on the page. 7 practice items and 50 test items were presented to the participant on an A4 sheet of paper, organised in a 5x10 matrix.

Appendix 2.2. Experiment 2.2 Correlations Between Standardised Cognitive and Language Tasks and performance in Composite Visual Recall and Reflection-Based tasks in Individuals Without Dyslexia

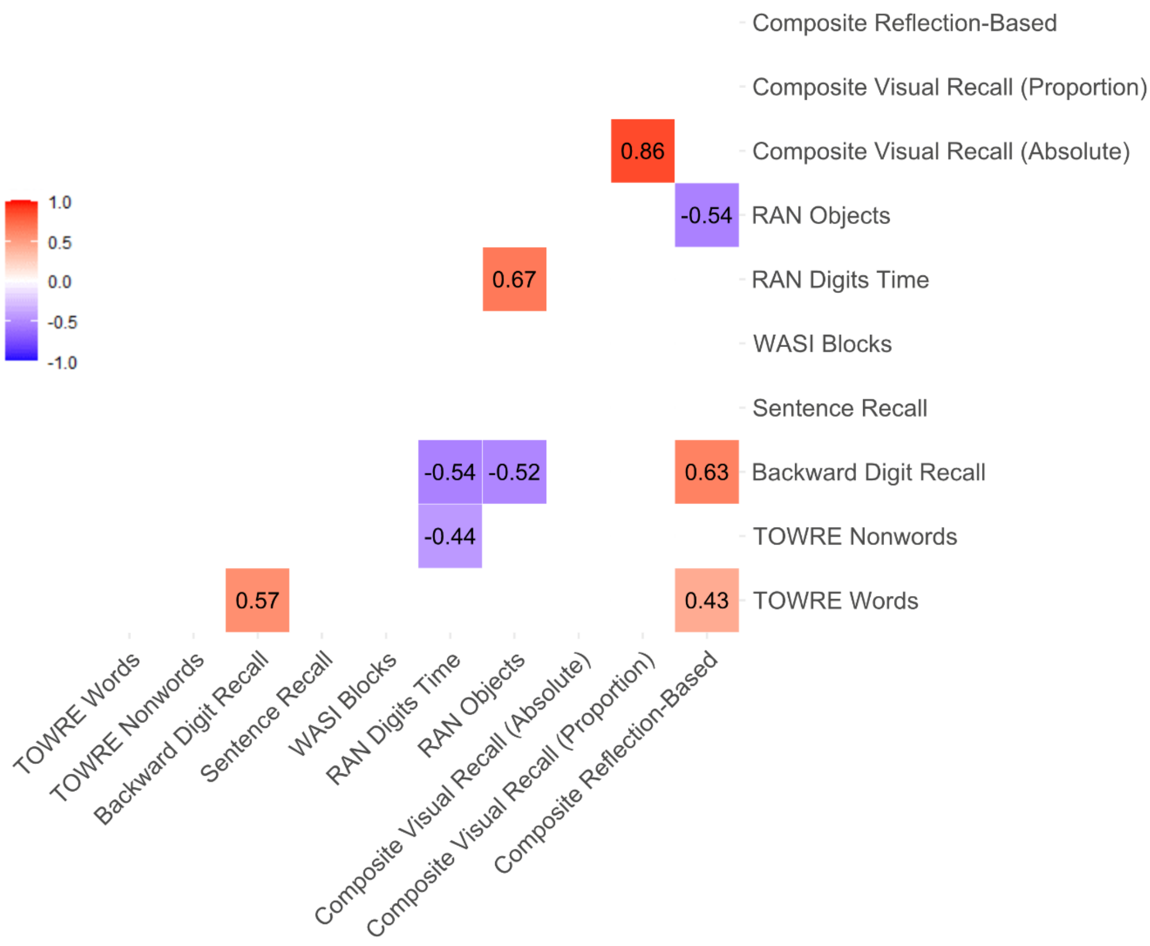


Figure 6.1. Correlations between standardised cognitive and language tasks and composite Visual Recall task performance and composite performance across reflection-based tasks in individuals without dyslexia in Experiment 2.2

Appendix 2.3. Experiment 2.4 Correlations Between Standardised Cognitive and Language Tasks and performance in Composite Visual Recall and Reflection-Based tasks in Individuals Without Dyslexia

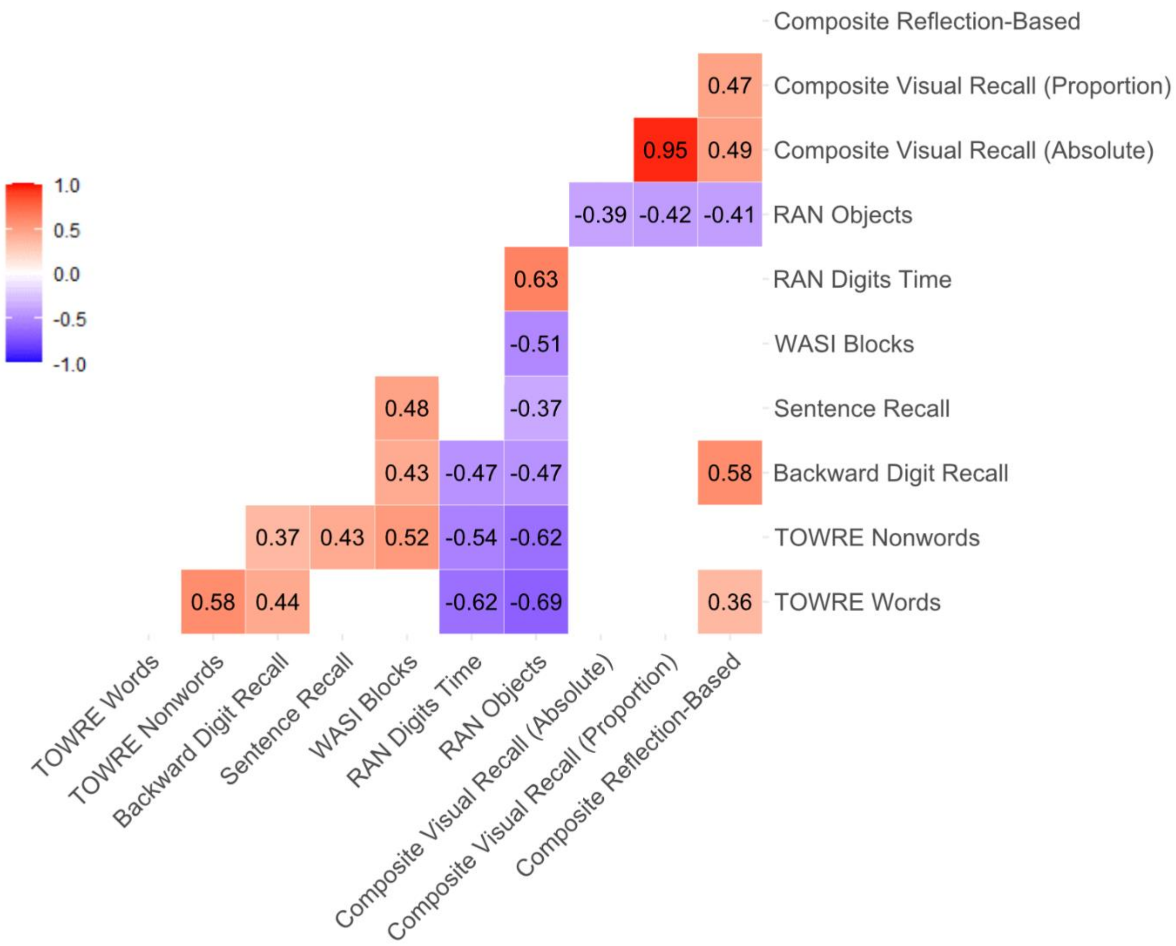


Figure 6.2. Correlations between standardised cognitive and language tasks and composite Visual Recall task performance and composite performance across reflection-based tasks in individuals without dyslexia in Experiment 2.4

Appendix 3.1. Number of children recruited by age in Experiment 3.1.

Table 6.1. Number of children recruited by age in Experiment 3.1.

Age	Number Recruited
8	4
9	4
10	12
11	15
12	12
13	21
14	10
15	11

Appendix 3.2. Performance Across Visual Recall and Reflection-Based Tasks After Excluding Children Who Show Pronounced Bias in Experiment 3.1

We reanalysed the data from Experiment 7 after excluding the 30 children (mean age = 11.67) who showed a very strong grammaticality bias in the Grammaticality Judgement task. Based on the 59 remaining children (mean age = 12.56 years), we still see evidence of learning in the Visual Recall task when using absolute scores: recall accuracy was still higher in Recall Block 6 compared to the Testing Block ($t_{58} = 4.68$, $p < .001$), and significantly higher in the Recovery Block than the Testing Block ($t_{58} = 3.75$, $p < .001$), although we still found no improvement in recall accuracy across Recall Blocks ($t_{58} = 1.58$, $p = .119$). A similar pattern was found when using proportion correct scores (Recall Block 6 > Testing Block: $t_{58} = 4.78$, $p < .001$; Recovery Block > Testing Block: $t_{58} = 3.68$, $p < .001$; no difference between Recall Block 2 and Recall Block 6: $t_{58} = 1.06$, $p = .294$). Performance in the Grammaticality Judgement task unsurprisingly remained above chance ($t_{58} = 4.16$, $p < .001$), and we still found positive correlations between performance in the Grammaticality Judgement task and performance in the Sequence Generation ($r = .609$, $p < .001$) and Sequence Completion ($r = .623$, $p < .001$), and between the Sequence Generation and Sequence Completion tasks ($r = .470$, $p < .001$).

Appendix 4.1. Experiment 4.1 Correlations Between Standardised Cognitive and Language Tasks and performance in Composite Visual Recall and Reflection-Based tasks in Individuals with Dyslexia

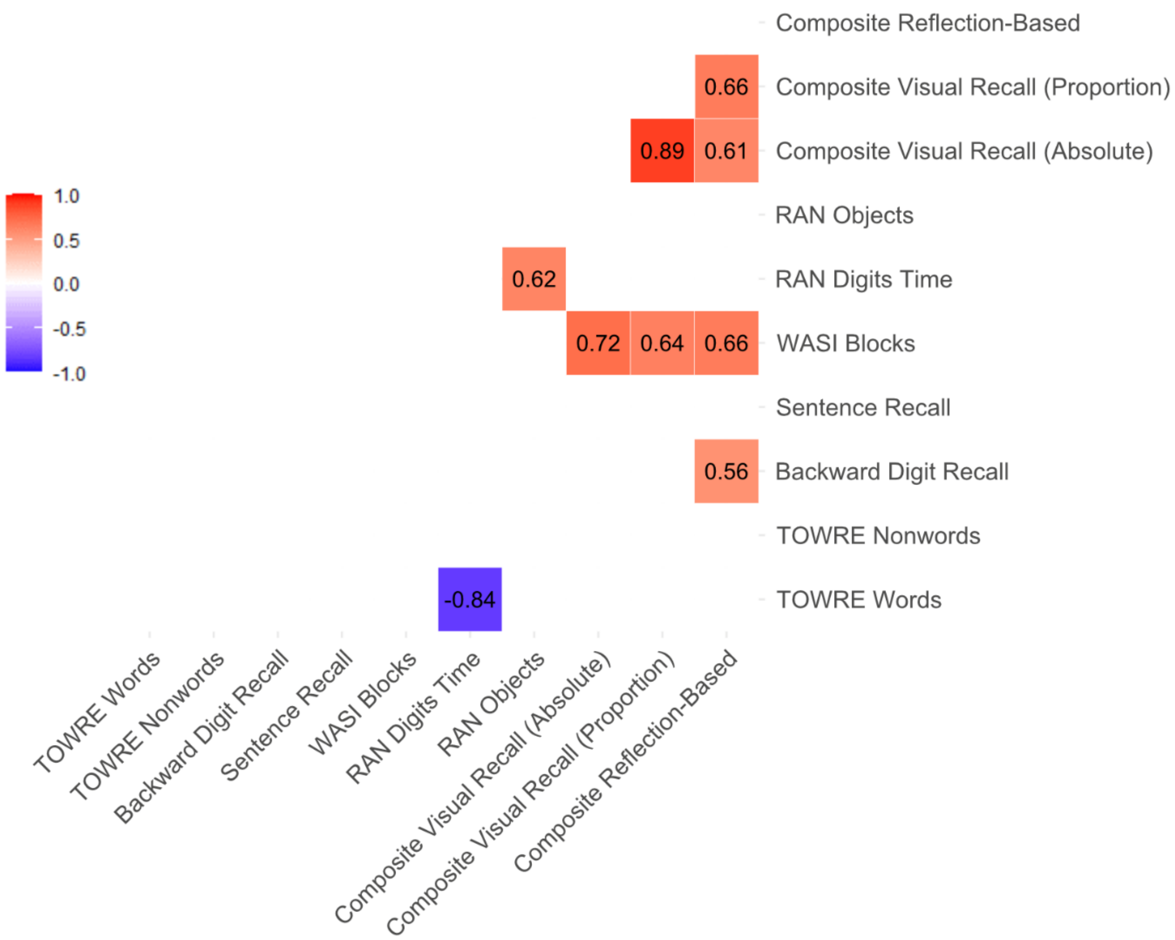


Figure 6.3. Correlations between standardised cognitive and language tasks and composite Visual Recall task performance and composite performance across reflection-based tasks in individuals with dyslexia in Experiment 4.1

Appendix 6.1. Descriptive Statistics for Standardised Cognitive and Language Tasks.

Table 6.2. Descriptive Statistics for Standardised Cognitive and Language Tasks in Experiment 2.1.

TOWRE Words	TOWRE Nonwords	Backward Digit Recall	Sentence Recall	WASI Blocks	RAN Digits	RAN Objects
92.93	57.00	20.00	66.25	47.09	16.42	27.65
10.01	7.23	5.79	8.26	10.89	3.19	3.63

Table 6.3. Descriptive Statistics for Standardised Cognitive and Language Tasks in Experiment 2.2.

TOWRE Words	TOWRE Nonwords	Backward Digit Recall	Sentence Recall	WASI Blocks	RAN Digits	RAN Objects
94.42	57.61	22.30	66.07	53.88	15.55	25.48
11.20	5.63	7.48	9.69	11.93	3.81	3.76

Table 6.4. Descriptive Statistics for Standardised Cognitive and Language Tasks in Experiment 2.4.

TOWRE Words	TOWRE Nonwords	Backward Digit Recall	Sentence Recall	WASI Blocks	RAN Digits	RAN Objects
90.46	54.60	19.90	68.03	50.36	17.80	29.62
13.32	7.19	6.02	5.84	11.40	5.21	9.18

Table 6.5. Descriptive Statistics for Standardised Cognitive and Language Tasks in Experiment 4.1.

TOWRE Words	TOWRE Nonwords	Backward Digit Recall	Sentence Recall	WASI Blocks	RAN Digits	RAN Objects
84.85	45.57	16.28	64.00	46.00	19.16	33.86
7.42	2.99	7.11	7.78	12.34	4.42	22.44

Table 6.6. Descriptive Statistics for Standardised Cognitive and Language Tasks for Individuals without Dyslexia in Experiment 4.3.

TOWRE Words	TOWRE Nonwords	Backward Digit Recall	Sentence Recall	WASI Blocks	RAN Digits	RAN Objects
92.25	57.81	25.37	67.68	52.25	16.13	28.66

11.40	4.80	6.72	5.12	14.28	4.27	3.84

Table 6.7. Descriptive Statistics for Standardised Cognitive and Language Tasks for Individuals with Dyslexia in Experiment 4.3.

TOWRE Words	TOWRE Nonwords	Backward Digit Recall	Sentence Recall	WASI Blocks	RAN Digits	RAN Objects
78.40	43.20	18.80	64.30	53.90	19.29	31.58
11.08	7.43	6.26	6.25	10.58	4.08	3.57

Appendix 6.2. Sequence Generation and Sequence Completion Descriptive Statistics

Table 6.8. Experiment 2.1 Sequence Generation and Sequence Completion Descriptive Statistics

Sequence Generation		Sequence Completion	
Mean	Standard Error	Mean	Standard Error
0.27	0.04	0.48	0.02

Table 6.9. Experiment 2.2 Sequence Generation and Sequence Completion Descriptive Statistics

Sequence Generation		Sequence Completion	
Mean	Standard Error	Mean	Standard Error
0.35	0.07	0.80	0.03

Table 6.10 Experiment 2.3 Sequence Generation and Sequence Completion Descriptive Statistics

Sequence Generation		Sequence Completion	
Mean	Standard Error	Mean	Standard Error
0.27	0.03	0.76	0.01

Table 6.11. Experiment 2.4 Sequence Generation and Sequence Completion Descriptive Statistics

Sequence Generation		Sequence Completion	
Mean	Standard Error	Mean	Standard Error
0.05	0.02	0.29	0.05

Table 6.12. Experiment 2.5 Sequence Generation and Sequence Completion Descriptive Statistics

Sequence Generation		Sequence Completion	
Mean	Standard Error	Mean	Standard Error
0.15	0.03	0.38	0.05

Table 6.13. Experiment 2.6 Sequence Generation and Sequence Completion Descriptive Statistics

Sequence Generation		Sequence Completion	
Mean	Standard Error	Mean	Standard Error
0.27	0.04	0.79	0.03

Table 6.14. Experiment 3.1 Sequence Generation and Sequence Completion Descriptive Statistics

Sequence Generation		Sequence Completion	
Mean	Standard Error	Mean	Standard Error
0.03	0.01	0.27	0.03

Table 6.15. Experiment 4.1 Sequence Generation and Sequence Completion Descriptive Statistics

Sequence Generation		Sequence Completion	
Mean	Standard Error	Mean	Standard Error
0.04	0.02	0.22	0.06

Table 6.16. Experiment 4.2 Sequence Generation and Sequence Completion Descriptive Statistics

Sequence Generation		Sequence Completion	
Mean	Standard Error	Mean	Standard Error
0.13	0.03	0.39	0.05