Learning to read

Effects of memory consolidation on orthographic and lexical learning

Connor Quinn

Supervisor: **Dr. Matthew H. Davis**
Dr. John Williams

Department of Theoretical and Applied Linguistics
Cognition and Brain Sciences Unit
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Wolfson College July 2018

I would like to dedicate this thesis to my mother.
You have always been a role model and a champion for me.

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements. This dissertation contains fewer than 80,000 words including references, appendices, footnotes, tables and equations and has fewer than 150 figures.

Elements of this thesis have been published and presented in the following locations:

Publications

Quinn, C., Taylor, J. S. H., & Davis, M. H. (2017). Learning and retrieving holistic and componential visual-verbal associations in reading and object naming. *Neuropsychologia*, 98, 68-84. https://doi.org/10.1016/j.neuropsychologia.2016.09.025

Presentations

Quinn, C., Taylor, J. S. H., & Davis, M. H. (2016). *Componential and holistic processing of novel visual-verbal associations*. Poster presented at the Society for Neurobiology of Language meeting, London, August 2016.

Quinn, C., Taylor, J. S. H., & Davis, M. H. (2016). *How do training and consolidation affect reading and recognising an artificial orthography?* Poster presented at UK Orthography Group meeting, Oxford, July 2016.

Quinn, C., Taylor, J. S. H., & Davis, M. H. (2015). *Holistic and componential processing in learning to read.* Invited talk at the Donders Discussions, November 2015.

Quinn, C., Taylor, J. S. H., & Davis, M. H. (2015). *Methodological challenges in functional imaging of reading.* Invited talk presented at the Looking at Language Acquisition meeting, Cambridge, May 2015.

Quinn, C., Taylor, J. S. H., & Davis, M. H. (2015). *Generalisation and consolidation of novel written words and objects in an fMRI study.* Poster presented the Interdisciplinary Advances in Statistical Learning meeting, San Sebastian, June 2015.

Quinn, C., Taylor, J. S. H., & Davis, M. H. (2015). *Generalisation and consolidation in learning to read words and name objects: an fMRI study.* Talk presented at the Experimental Psychology Society meeting, Leeds, April 2015.

Quinn, C., Taylor, J. S. H., & Davis, M. H. (2015). *Exploring the effect of sleep on the learning of novel written words and objects.* Talk presented at the Sleep and Memory meeting, Royal Holloway, January 2015.

Quinn, C., Taylor, J. S. H., & Davis, M. H. (2014). *Learning to read words and name novel objects.* Poster presented at the Society for Neurobiology of Language meeting, Amsterdam, August 2014.

Connor Quinn
July 2018

Acknowledgements

I would like to first thank the ESRC for the opportunity to complete this PhD project. It has been a privilege to spend this time exploring such an exciting body of research.

I have been very fortunate in my academic mentors. I would thank Matt Davis for insightful guidance on the research process, for encouragement and support throughout, and for expressing such enthusiasm for nifty experimental designs and new data. I am grateful also to Jo Taylor who was not only supremely knowledgeable, but generous with her time. Your kindness has been appreciated more than you can know. Finally, I thank my second supervisor John Williams for being a cheerleader at every stage and for the endlessly enjoyable reading group.

Thank you to my friends from DTAL Giulia Bovolenta, Carla Pastorino-Campos, Dimitrios Alikaniotis, Elspeth Wilson, and Andrew Caines. You are a wonderful group of people and made every week a joy. A particular mention goes to Dimitrios Alikaniotis who contributed more than any other to all aspects of this PhD. As well as being a friend, you have been a mentor who has always pushed me in the right direction. I would also like to thank the staff members of the MRC Cognition and Brain Sciences Unit and the Department of Theoretical and Applied Linguistics for their friendship and practical support. It has been a pleasure to work with such a group.

It is important to thank the 131 participants who each completed two long days of learning. You have joined a select group who can communicate successfully in this writing system. Perhaps a new linguistic group will be born from these experiments. I would also like to mention the students of Dublin Adult Learning Centre who prompted the beginnings of this journey. Your bravery as lifelong learners has sustained me when times were hard. I am grateful to others who have led me to this point: Michelle Ellefson, who has been both a mentor and a friend, and Fred Cummins for giving me insight into to true meaning of education.

I thank my wonderful family. Through everything we have stuck together. Without your faith and encouragement this would not have been possible.

Finally, I am grateful to my beloved Charlotte for reminding me just how much joy life can hold.

Abstract

Learning to Read
Effects of memory consolidation on orthographic and lexical learning

Connor Quinn

In recent years the role of offline consolidation in supporting word learning has attracted great interest and has provided valuable insight into how novel spoken and written words are learned. Relatively little attention has focused on whether offline consolidation supports the learning and generalisation of novel orthographic knowledge. Meanwhile, laboratory-based approaches have proven valuable in overcoming the methodological challenges of studying reading acquisition, i.e. learning letter-sound knowledge. This thesis combines laboratory-based orthographic learning with an overnight consolidation framework to track the effects of sleep on learning novel letters and novel written words in six experiments. Experiment 1 validated the artificial orthography paradigm by using fMRI to show the novel orthography activated similar neural regions to pseudowords written in familiar orthography. Comparing recently learned words and objects additionally highlighted the componential and holistic processes that distinguish reading from object naming. Experiments 2, 3, and 4 investigated whether overnight consolidation had contrasting effects on learning novel letters and learning novel written words. All three studies showed overnight improvements in the ability to use and generalise knowledge of letters. Experiment 3 further assessed whether consolidation supported the formation of bigram representations. While the results did not show bigram consolidation, a recognition memory task indicated participants had consolidated the novel spoken words. Experiment 4 manipulated the internal statistical structure of the novel words finding, in contrast to Experiment 3, participants had consolidated the written forms of the novel words. Experiments 5 and 6 asked whether consolidated and unconsolidated spoken words would support orthographic learning. These studies failed to observe previous findings of spoken word consolidation and did not demonstrate clear effects of lexical knowledge on orthographic learning. The findings of the thesis demonstrate the importance of letter-level learning and consolidation during reading acquisition as well as highlighting the value of laboratory-based studies for understanding the interdependent trajectories of the skills involved in reading.

Table of contents

List of figures **15**

List of tables **17**

1 Introduction **1**

1.1 Teaching letters or teaching words 2

2 Literature Review **7**

2.1 Computational models of reading 7

2.1.1 The Interactive Activation model 8

2.1.2 The Dual Route Cascade model 9

2.1.3 The Triangle model. 11

2.1.4 Themes and questions 14

2.2 Neural basis of reading 17

2.2.1 Neural basis of skilled reading 18

2.2.2 Neural perspectives on learning to read 22

2.3 Artificial reading studies 25

2.3.1 Behavioural studies using artificial orthographies 26

2.3.2 Brain studies using artificial orthographies. 28

2.4 Memory Consolidation 30

2.4.1 Complementary Learning Systems 31

2.4.2 Spoken word learning in the CLS 34

2.4.3 Orthographic consolidation 36

2.5 Summary and thesis overview 41

3 Neural basis of holistic and componential visual-verbal mappings **45**

3.1 Introduction 46

3.1.1 Neural bases of holistic and componential processes in reading . . . 47

3.2 Experiment 1 49

3.3 Methods 52
3.3.1 Participants 52
3.3.2 Materials 52
3.3.3 Procedure 54
3.3.4 Imaging analysis 59
3.4 Results 62
3.4.1 Artificial words and objects during training 62
3.4.2 Results during scanning 62
3.4.3 Whole-brain imaging analyses 64
3.4.4 ROI analyses based on real words and objects in the localiser 71
3.4.5 Hippocampal ROI analysis 72
3.5 Discussion 73
3.5.1 Behavioural results highlight componential and holistic learning . . 73
3.5.2 VOT contributions to reading and naming 74
3.5.3 Parietal and frontal contributions to reading and naming 77
3.5.4 Emergence of holistic representations for consolidated items 78
3.5.5 Validation of laboratory-based learning paradigms and implications for education 79

4 Lexical and sublexical orthographic consolidation 81
4.1 Background 82
4.1.1 Outstanding issues from Experiment 1 82
4.2 Experiment 2 85
4.3 Methods 86
4.3.1 Participants 86
4.3.2 Materials 86
4.3.3 Procedure 87
4.4 Results 92
4.4.1 Training task 93
4.4.2 Reading task 93
4.4.3 Matching task 95
4.5 Discussion 100
4.5.1 Sublexical overnight consolidation 100
4.5.2 Multiple measures of whole-word learning 101
4.6 Experiment 3 105
4.7 Methods 107
4.7.1 Participants 107

4.7.2 Materials 109
4.7.3 Procedure 110
4.8 Results 112
4.8.1 Reading task 112
4.8.2 Matching task 113
4.8.3 Recognition memory 115
4.9 Discussion 119
4.9.1 Replication of sublexical consolidation 119
4.9.2 Consolidation of spoken word representations 120
4.10 Experiment 4 121
4.11 Methods 123
4.11.1 Participants 123
4.11.2 Materials 123
4.11.3 Procedure 126
4.12 Results 127
4.12.1 Reading task 127
4.12.2 Matching task 130
4.12.3 Recognition memory 133
4.13 Discussion 135
4.13.1 Replication and extension of sublexical consolidation 136
4.13.2 Consolidation of written word representations 137
4.14 Overall discussion 138
4.15 Conclusion 143

5 The impact of spoken word consolidation on learning a novel orthography. 145
5.1 Background 146
5.2 Experiment 5 149
5.3 Methods 150
5.3.1 Participants 150
5.3.2 Materials 150
5.3.3 Procedure 150
5.4 Results 154
5.4.1 Phoneme monitoring 154
5.4.2 Spoken word repetition 154
5.4.3 Orthographic training 154
5.4.4 Matching task 155
5.5 Discussion 157

5.5.1 Failure to observe consolidation effects in positive control. 158
5.5.2 Effects of lexical knowledge on learning to read. 159
5.5.3 No differences on the matching task. 160
5.6 Experiment 6 160
5.7 Methods 163
5.7.1 Participants 163
5.7.2 Materials 163
5.7.3 Procedure 163
5.8 Results 166
5.8.1 Phoneme monitoring 166
5.8.2 Object learning 166
5.8.3 Spoken word repetition 167
5.8.4 Orthographic training 168
5.8.5 Matching task 168
5.8.6 Cross-experiment comparison. 169
5.9 Discussion 172
5.9.1 Failure to observe spoken word consolidation 172
5.10 Conclusion 174

6 Thesis summary and conclusions 177
6.1 Thesis summary 178
6.2 Limitations of this thesis 180
6.3 Future directions 182
6.4 Conclusions 183

References 187

List of figures

2.1 DRC model of word recognition 10
2.2 Triangle model of word recognition 12
2.3 Sequence encoder architecture 15
2.4 Meta-analysis of brain function in reading 19
2.5 Model of system-level consolidation 33
2.6 Oscillatory activation for memory reactivation during sleep 34

3.1 Experiment 1: Timeline and procedures 56
3.2 Experiment 1: Response times on localiser task 61
3.3 Experiment 1: Behavioural results during training and scanning 65
3.4 Experiment 1: Localiser activation 66
3.5 Experiment 1: Whole brain analysis excluding incorrect responses 67
3.6 Experiment 1: Hippocampal ROI analysis 73

4.1 Experiment 2: Lexical items and characters 87
4.2 Experiment 2: Timeline and procedures 88
4.3 Experiment 2: Examples of probe items for matching task 91
4.4 Experiment 2: Improvements in reading performance during training. . . . 92
4.5 Experiment 2: Performance on reading task 94
4.6 Experiment 2: Repetition priming on reading task 95
4.7 Experiment 2: Comparison of visual familiarity in probes 97
4.8 Experiment 2: Effects of training and consolidation on the matching task . . 98
4.9 Experiment 2: Response times to detect mismatching probes 99
4.10 Experiment 3: Timeline and procedures 108
4.11 Experiment 3: Lexical items and characters 109
4.12 Experiment 3: Sample stimulus assignment 110
4.13 Experiment 3: Reading latency according to bigram familiarity 112
4.14 Experiment 3: Performance on the matching task 114

4.15 Experiment 3: Response times to detect mismatching probes 116
4.16 Experiment 3: Recognition memory for written words from days 1 and 2 . . 117
4.17 Experiment 3: Recognition memory for spoken words from days 1 and 2 . . 118
4.18 Experiment 4: Timeline and procedures 124
4.19 Experiment 4: Example of frequent bigrams in Experiment 4 125
4.20 Experiment 4: Accuracy and response time on the reading task 128
4.21 Experiment 4: Reading latency for untrained words according to bigram training status 129
4.22 Experiment 4: Accuracy and response time on the matching task 130
4.23 Experiment 4: Response times on matching task according to bigram training status 131
4.24 Experiment 4: Response times to detect mismatching probes according to error position. 132
4.25 Experiment 4: Recognition memory for written words from days 1 and 2 . . 133
4.26 Experiment 4: Recognition memory for spoken words from days 1 and 2 . . 135

5.1 Experiment 5: Lexical items and characters 151
5.2 Experiment 5: Timeline and procedures 152
5.3 Experiment 5: Response time on non-word repetition task 155
5.4 Experiment 5: Improvements in reading performance during training 156
5.5 Experiment 5: Performance on matching task 157
5.6 Experiment 5: Response times to detect mismatching probes 158
5.7 Experiment 6: Timeline and procedures 164
5.8 Experiment 6: Response time on non-word repetition task 167
5.9 Experiment 6: Improvements in reading performance during training 168
5.10 Experiment 6: Performance on matching task 170
5.11 Experiment 6: Response times to detect mismatching probes 170

List of tables

3.1 Experiment 1: Artificial stimulus sets 53
3.2 Experiment 1: Word and pseudoword stimuli 55
3.3 Experiment 1: Behavioural results 63
3.4 Experiment 1: Brain regions involved in reading real words and objects in the localiser 68
3.5 Experiment 1: Brain regions involved in reading real words and pseudowords in the localiser 69
3.6 Experiment 1: Brain regions involved in reading artificial words and naming artificial objects. 70
3.7 Experiment 1: Region of interest analyses 71

4.1 Experiment 2: Performance on reading task 94
4.2 Experiment 3: Performance on reading task 113
4.3 Experiment 3: Performance on matching task 114
4.4 Experiment 4: Performance on reading task 128
4.5 Experiment 4: Performance on matching task 130

5.1 Experiment 5: Performance on spoken word repetition task 155
5.2 Experiment 5: Accuracy during training 156
5.3 Experiment 6: Performance on spoken word repetition task 167
5.4 Experiment 6: Accuracy during training 169

Chapter 1

Introduction

At its most simple reading is the ability to access spoken language by way of visual symbols. Fluent reading draws upon language and vision but differs from these core developmental skills in that reading is a culturally transmitted ability. Over time and with extensive instruction and practice, learners come to combine vision and language in order to fluently access language through print.

From the combination of vision and language two competing features of reading emerge: the ability to represent the component parts of written words using letters, and the ability to recognise holistic representations of written words. The great power of literacy lies in the componential nature of letter-sound mappings. Just 26 letters allow readers of English to identify a vast array of possible words. A skilled reader can even access the spoken forms of entirely unfamiliar words such as ‘spape’ without difficulty; our ability to generalise knowledge of letters and sounds allows us to read words beyond even our spoken lexicons. Reading differs in this way from being able to name objects in the world; upon encountering an unfamiliar object in the world we cannot determine what it is called. Nevertheless, it may be that fluent reading also relies on the ability of our visual system to quickly identify visual inputs. When we read a familiar word that we have encountered numerous times we do not laboriously decode that word letter-by-letter but instead recognise the whole word. Thus, when reading a word such as ‘pint’ we can only know how to say it correctly if we already know that the word is not homophonous with ‘mint’. These contrasting aspects of reading have provoked strong disagreements about the best approach to reading instruction.

This thesis revolves around the question of how these competing features of reading interact with the processes of memory formation. On the one hand learning the componential mappings from letters to sounds allows us to decode words, while on the other learning to recognise whole-word forms allows us to rapidly identify familiar or irregular words. Simply, to what extent does knowing parts or knowing wholes impact on the process of learning to read? By adapting recent approaches that make use of sleep-related changes in the representations of knowledge, we aim to shed light on how these reading processes play out in the earliest stages of learning to read.

Literacy and society. Reading is a core part of modern life and a skill that most children and adults are expected to master; so much so that successful literacy acquisition is a necessary precursor to engage meaningfully in modern society. The impact of literacy plays out at all levels of society and propagates over generations, improving not only individual attainment but also employment, community, health and overall wellbeing (Gakidou et al., 2010; Snow, 2016; Viner et al., 2012). As the economic structure of society shifts evermore towards skilled labour, the increasing pace of technological change favours those with higher educational attainment. As a result, the prospects of those with low levels of literacy become increasingly limited (Acemoglu, 2002; Levy and Murnane, 2004; Snow, 2016). Moreover, poor readers are at risk of social exclusion; illiteracy is associated with a cascading pattern of social exclusion beginning at school but which persists across the lifespan (Whitehouse et al., 2009), and which often culminates in the prison system (Christle et al., 2005; Snowling et al., 2000). Indeed, an assessment of prison literacy levels in Ireland found that 70.6% of the prison population fell at the lower range of functional literacy or could not complete the survey (Morgan and Kett, 2003, The adult prison literacy survey). Consequently, early reading instruction forms an "important nexus between oral language competence, the transition to literacy, and psychosocial wellbeing across the school years and beyond." (Snow, 2016, p218). This societal perspective makes clear the critical importance of ensuring learners to reach their reading potential. However, despite the critical role of literacy in providing learners with the necessary skills to engage with society in a democratic manner, the provision of conclusive evidence-based approaches to literacy instruction has proved challenging (Morais, 2017; Morais and Kolinsky, 2005).

1.1 Teaching letters or teaching words

The long-running debate around how best to teach children to read has tended to focus on either phonics or whole-language teaching methods. Briefly, whole-language methods argue that the primary goal of reading is to extract meaning from text, and so early reading instruction should be focused on content that is meaningful and relevant to young children. Providing extensive exposure to meaningful print would therefore allow children to learn to recognise whole words, and over time learn the letter-sound mappings of print. The focus of instruction in whole-language approaches then is lexical; teaching children to recognise whole words. In contrast, phonics is concerned with the relationships between letters and sounds. The primary focus of instruction therefore is sub-lexical; teaching children about the component parts of words. Advocates of phonics-based methods argue that learning the componential relationship between letters and sounds is the key skill to be learned by children as novice readers (Byrne, 1998; Ehri, 2005; Share, 1995). The early stages of literacy acquisition are thus concerned with teaching how graphemes (one or more letter clusters) map to phonemes, e.g., the grapheme *ch* maps to the phoneme /tʃ/ in 'church'.

In recent years the education literature has settled upon phonics as the only evidence-based method of teaching reading (Ehri et al., 2001; Rayner et al., 2001; Torgerson et al., 2006; Wyse and Goswami, 2008). Systematic literature reviews by Ehri et al. (2001) and Torgerson et al. (2006) agree that children receiving phonics instruction achieve better reading outcomes than those who do not. Meanwhile in the UK, the Rose Review (2006) recommended that synthetic phonics, which involves explicit instruction in letter-sound decoding and blending, should underlie early reading instruction.

However, the strength of evidence for the primacy of phonics instruction does not negate the importance of wider language skills. Indeed, proponents of phonics acknowledge the value of lexical knowledge in reading, noting that the goal of reading is to identify words and ultimately access meaning (Nation, 2017; Perfetti, 2007). Accordingly, the ability to decode letter-sound mappings is necessary but not sufficient for fluent reading. For instance, oral vocabulary can influence reading comprehension independently of decoding ability (Gough and Tunmer, 1986; Nation and Snowling, 2004; Ricketts et al., 2007). Likewise, the review by Torgerson et al. (2006) found no association between phonics instruction and improved comprehension. Furthermore, knowledge of whole words can also support sub-lexical learning; children not only use letter-sound mappings to learn about whole words, they also use knowledge of whole words to infer the letter-sound mappings within those words (Stuart et al., 1999; Thompson et al., 1996). Together these studies paint a complex picture of reading development where lexical and sublexical knowledge develop in an interrelated and interdependent manner. Consequently, debate about the optimal approach to teach phonics is ongoing. For example, Shapiro and Solity (2016) compared the efficacy of two phonics programmes in the UK, including multiple schools and multiple years of assessment. McArthur et al. (2015) meanwhile investigated whether whole-word instruction should take place in parallel to phonics instruction for early readers. While differences in the overall efficacy of programmes can be highlighted in this way, it can be difficult to understand the mechanisms that promote efficacy.

Identifying the relative contributions of lexical and sublexical knowledge during reading acquisition is therefore an important part of improving reading instruction. For example, how and when does knowledge of written and spoken whole words support the decoding process? Is a firm basis of phonics required before recognition of sight words can be beneficial or do both develop in tandem? The difficulty in tracing these developments has been noted in the literature:

> "The existing literature provides little insight into exactly what is required for development of an optimal expert strategy for word identification in reading. The outcome of making the transition to this stage of development is assumed to be an "autonomous lexicon" (Share, 1995) which allows "automatic word identification", but how does that happen? What changes in the representations or processes underlying reading behaviour to afford these outcomes?" (Andrews and Scarratt, 1996, p141)

Indeed, this point is amplified by Castles and Nation for the relative development of sublexical and lexical reading skills:

> "The key point is that we need to uncover the role played by different factors, *as individuals progress from alphabetic decoding to skilled recognition of new words*, perhaps at an item-based level, rather than examining their influence after such a transition has occurred." (Castles and Nation, 2006, p173)

In both cases the authors note the that the relative developmental trajectories are key. Rather than examining the outcome of learning we should examine how representations change. In order to further inform our understanding of reading we should be less concerned then with oppositional perspectives between phonics and whole-word instruction, but rather how and when these two aspects of reading support each other in the learning trajectory (Nation, 2017).

Despite the vigorous debates around these issues in research and practice, conclusive evidence about the relationship between lexical and sublexical knowledge in reading has proved surprisingly difficult to provide. This is because both principled and practical challenges exist for naturalistic studies of reading acquisition. Classrooms are (literally and figuratively) noisy environments and so reliably tracking the trajectory of learning is difficult. The success of an individual reader will depend not only on that individual's cognitive abilities but also their home situation, teacher resources and expertise, classroom size, individual motivation, and other factors unrelated to the strength of the pedagogical approach (Diamond, 2007). See Morais and Kolinsky (2005) for a summary of the multiple confounds specifically associated with reading research. While population- and classroom-level studies provide the ultimate measure of whether a reading instruction policy has succeeded, Wyse and Goswami (2008) note that very few naturalistic studies comparing different methods of reading instruction meet rigorous experimental standards. Thus, not only are classroom interventions difficult, costly, and disruptive, they may also be of restricted scientific value due to the poor reliability allowed by the experimental setting. Moreover, because literacy develops over a long time frame, even when strong evidence is presented that evidence may not capture the rapid changes that take place.

This thesis offers a complementary perspective on reading development through a series of laboratory-based studies that provide fine-grained insight into the earliest stages of learning to read. In doing so we explore issues raised by computational and neuroscientific accounts of reading through the lens of memory consolidation. Adopting this approach offers a perspective that is difficult to achieve in naturalistic settings: a tightly controlled experimental framework in which the trajectories of different reading skills can be tracked. Although this focus on experimental control sacrifices the ecological validity found in classroom environments, the findings nevertheless offer a complementary perspective on classroom-based studies. While artificial reading studies in isolation may not constitute strong evidence for classroom practice, the approach used here may be adapted to test specific hypotheses quickly (unlike longitudinal studies) and with experimental rigour, thus supporting attempts to improve literacy instruction more broadly. To this end the thesis goes some ways to validating the use of artificial orthographies as a tool. Additionally, artificial reading studies offer an exciting domain in which to apply recent findings on the role of sleep in memory formation. By drawing together these two experimental approaches, artificial reading studies and memory consolidation studies, this thesis seeks to cast light on the early stages of learning to read. Exploring how different features of early reading change with and without the benefit of sleep may thus tell us about how these features support each other during learning.

Chapter 2

Literature Review

The first two sections of the following review highlight the most relevant issues that have emerged from computational and neuroscientific accounts of reading. Focusing on the most relevant topics from the perspective of reading instruction reveals that providing empirical evidence for the precise trajectory of reading development has proved difficult in both domains. Differences in classrooms, instructional methods, etc., mean it is difficult to make firm conclusions about early reading. The third section reviews cases in which laboratory-based reading paradigms have been adopted in order to investigate reading acquisition in an experimentally rigorous and replicable manner. Together these three sections highlight a number of unresolved issues about the trajectory of early literacy acquisition. The final area of the review covers the role of memory consolidation in learning. In particular, this section focuses on evidence for consolidation-related changes in processing recently learned linguistic information, and the potential opportunities allowed by adopting overnight consolidation as an experimental manipulation. This section concludes that the processes of memory formation revealed by the sleep consolidation literature have potential to inform both cognitive models of reading and practical reading instruction.

2.1 Computational models of reading

Computational models of reading attempt to directly specify the steps by which visual inputs can be used to access spoken language, and/or meaning. Committing to a specific computational account means that each model makes direct predictions about reading that can be empirically tested. The success or failure of different models to explain different aspects of reading can, therefore, inform our understanding of the reading process. We do not focus on psychological theories of reading development such as Share's self-teaching hypothesis (Share, 1995, 1999) or Ehri's phases of reading development (Beech, 2005; Ehri, 1995). These accounts of reading have not been implemented into a concrete model and so may be less informative with regard to the mechanics of orthographic learning. Instead, this section reviews the points of convergence and divergence between different computational models of reading in order to identify the aspects most critical to the process of learning to read. Notably, the distinction between recognising whole-words and decoding letter-by-letter in the educational literature is mirrored to a large extent by findings from computational models reading. Although these models are not the focus of this thesis, they help us to understand the types of knowledge that support reading, and which might be impacted by sleep consolidation.

2.1.1 The Interactive Activation model

An early model of reading, the Interactive Activation Model (IAM; McClelland and Rumelhart, 1981; Rumelhart and McClelland, 1982), highlights some of the features of reading that any model must account for: visual inputs and letters, whole word representations, and some way for these to interact. The model demonstrates the complementary roles of letters and words in reading by successfully accounting for the word superiority effect. The word superiority effect shows that participants are more accurate to report briefly presented letters when those letters appear within familiar words rather than within a scrambled string (McClelland and Rumelhart, 1981). This finding suggests recognition of whole words plays an important role in letter-recognition; an apparent reversal of the standard view that recognising letters is a prerequisite for recognising words.

The IAM proposed a hierarchical process for recognising written words, with visual features leading to recognition of letters and, from there, lexical forms. Critically, the IAM included cascading and interactive connections between visual-, letter-, and lexical-layers in the model. Knowledge of letters and knowledge of words interact with each other during processing. Thus, the highly-activated letter representation is more easily identified within a familiar word. With this structure, the model (largely) succeeds in accounting for the word superiority effect seen in behavioural studies of word recognition (Grainger, 2008).

In addition to highlighting the reciprocal relationship between letters and words, the model posited the existence of an orthographic lexicon. The orthographic lexicon refers to the idea that readers have learned representations of whole written words. These representations are distinct from the letters that make up those words, and further, are distinct from the phonological lexicon of spoken words that exist in spoken language. A major debate for later models of reading has to do with whether there is, in fact, an orthographic lexicon or whether the whole-word effects seen in reading can be explained without the need for readers to store written word forms (Harm and Seidenberg, 2004; Plaut et al., 1996; Rastle, 2007; Seidenberg and McClelland, 1989).

Although the word superiority effect shows that letters are recognised more easily in the context of a familiar word, the effect extends to show beneficial effects of a viable pseudoword (McClelland and Johnston, 1977). These pseudowords cannot have a whole-word representation and yet they facilitate recognition of single letters embedded within the context of the pseudoword. This finding suggests that intermediate levels of processing exist between the letter and whole-word levels of representation, perhaps made up of familiar bigrams or letter clusters. Later research has borne out the impact of intermediate level orthographic representations in reading, suggesting that the hierarchical levels by which letters are mapped to higher level representations include features such as syllables (Carreiras and Perea, 2002), rimes (Taft, 1992) and morphemes (Rastle et al., 2004).

The issues raised by the IAM remain relevant to current models of reading: the role of learning in the model, the relationship between visual inputs and letters, the relationship between letters and whole words, the question of whether written words have a stored representations, and the role of intermediate levels of orthography (e.g., bigrams).

2.1.2 The Dual Route Cascade model

One of the most prominent models for how written text is mapped onto spoken language is the dual route cascade model (DRC; Coltheart et al., 1993, 2001). The 'dual-route' element of the name refers to the two routes by which the model can read a spoken word, a lexical semantic route involving whole-word forms, and a sublexical route translating letters-to-sounds. As with the IAM, the model is 'cascaded' in the sense that information cascaded through the model, rather than discrete elements of the model operating in a serial manner (Figure 2.1, p10).

The sublexical route uses a rule-based system that converts letter representations to sounds in a systematic manner. In this way, unknown words can be decoded and read aloud. Letters are processed in a serial manner, each letter, in turn, being translated to sound (Rastle and Coltheart, 1999, 2000). The rules for converting between letters and sounds are taken from Rastle and Coltheart (1999) and were chosen based on the statistical properties of language. The serial nature of reading unknown words in this model helps to explain the word-length effects seen in reading. For example, Weekes (1997) showed that the naming latency of nonwords increased with the number of letters; nonwords with more letters take longer to read aloud. For high-frequency words, however, the naming latency did not increase as a function of letter-length, suggesting the whole-word form was being recognised and so serial decoding of the letters was unnecessary.

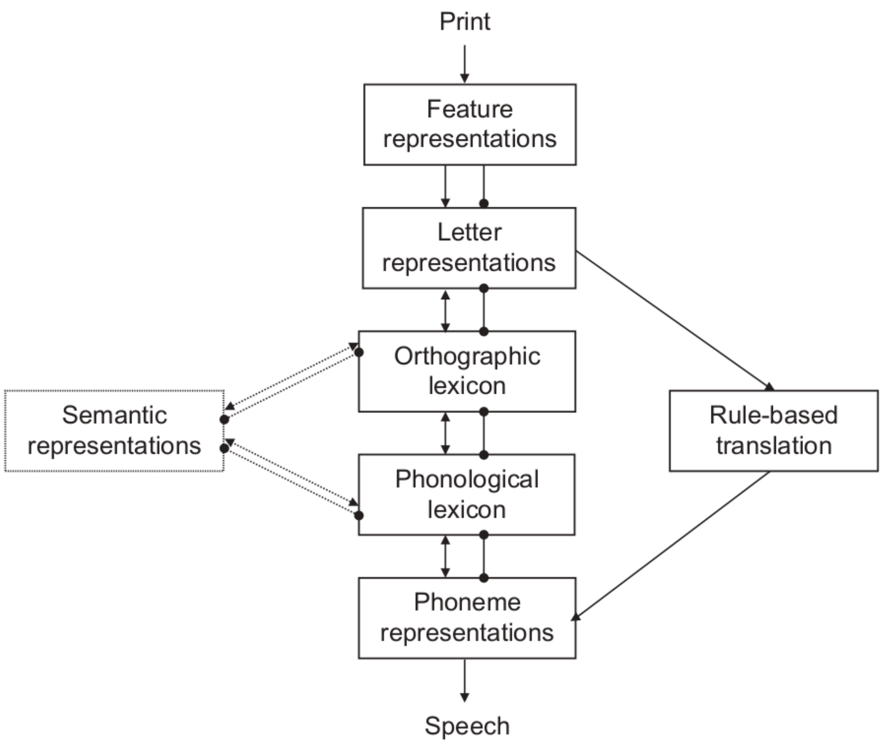


Fig. 2.1 DRC model of word recognition adapted from Hickok and Small (2015), Neurobiology of Language.

The lexical semantic route consists of three elements: an orthographic lexicon containing knowledge of written words, a phonological lexicon containing knowledge of spoken words, and a semantic system. The reciprocal connections between these elements mean that the orthographic lexicon entry of a word can be accessed in two ways. Once the entry for a word in the orthographic lexicon has been accessed, that entry can directly access the phonological information associated with that word. Alternatively, recognition of the lexical orthographic entry may allow direct retrieval of the semantic entry associated with that word, and from there to phonology (Rastle, 2007). This architecture means that for regular words both lexical and sublexical routes would converge on the same spoken output. For irregular words, by contrast, the sublexical route would produce inaccurate phonological outputs. Only the lexical semantic route can successfully read irregular spellings of familiar words, making use of semantic and phonological knowledge to recognise the word. Through the reciprocal connections between orthographic, semantic, and phonological elements of the model it is able to account for idiosyncratic spellings, e.g., that *pint* does not sound like *mint*. These routes mean that the DRC model largely succeeds in explaining a wide variety of behavioural patterns seen in reading including lexical decision tasks, effects of regularity, frequency, orthographic neighbourhood, and priming effects (Coltheart, 2005, 2006; Perry et al., 2007; Zorzi et al., 1998).

The contrast between lexical and sublexical processing to some extent parallels the discussions of phonics versus whole-word instruction methods mentioned in the introduction, but importantly the DRC model aims to describe fluent readers. Whereas educational debates ask whether early literacy efforts should attend to letters or words, the DRC is largely silent on how lexical and sublexical reading routes are acquired. One suggestion comes from Jackson and Coltheart (2001) who suggest children simultaneously develop all elements of the reading system laid out in the DRC model. Thus, novice readers can access words either lexically or sub-lexically from the very beginning, with changes reflecting their skill with different elements in the process (Coltheart, 2006). Despite this proposal the model as implemented is hard-wired and includes hand-coded rules to translate letters to sounds; the model does not directly specify the mechanism by which learning takes places, either of the rules to translate letters to sounds or to recognise whole words. Moreover, while the DRC model largely succeeds in capturing behavioural effects of reading irregular words in English, English orthography may not be representative of orthographic processing in general. Instead, Share (2008) suggests that as most orthographies are more regular than English then “a single rule-based mechanism should be adequate for pronouncing all (or nearly all) letter strings.” (p588). If most orthographies contain simple mappings between letters and sounds then perhaps whole-word recognition receives undue attention. Ziegler and Goswami (2005), by contrast, argue that English orthography is simply at one point on a continuum from transparent (e.g., Italian) to opaque (e.g., Chinese). The DRC model itself does not speak to cross-linguistic differences, and therefore it is potentially less informative for the purposes of the current research.

2.1.3 The Triangle model.

The triangle model of reading aims to capture the essential processes of reading in the relationship between three types of knowledge: orthography (*O*), phonology (*P*), and semantics (*S*) (Harm and Seidenberg, 2004; Plaut et al., 1996; Seidenberg and McClelland, 1989). The model is a connectionist model (see Elman et al., 1996 for an overview of the connectionist framework) and so does not include symbolic representations of rules or words. The model instead extracts the statistical regularities between the three elements during training, with knowledge captured by the changing ‘weights’ of connections. The connection weights therefore simultaneously store the knowledge and perform the processing in the model. The model can access the meaning of a written word either by directly recognising a word and so

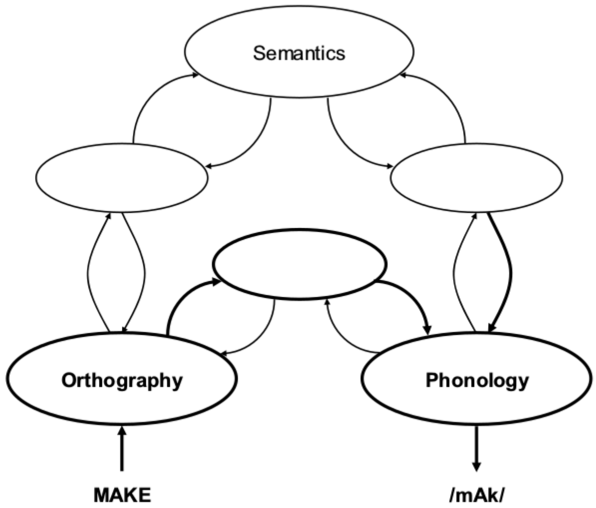


Fig. 2.2 Triangle model of reading adapted from Woollams, Lambon Ralph, Plaut, and Patterson, 2007.

mapping from orthography to semantics $(O \mapsto S)$ or indirectly by decoding the letter sound mappings and then using oral language to access meaning $(O \mapsto P \mapsto S)$.

Connectionist models are inherently models of learning as well as processing. These models are appealing as they demonstrate *in principle* that complex learning can take place purely on the basis of statistical learning. Connectionist models become sensitive to whatever regularities exist in the mappings between the inputs, i.e., the consistency between inputs and outputs. Consistency in spelling, then, refers to whether a written form has only one correct spoken form, or whether more than one pronunciation is possible (Vousden et al., 2011). Consistency might take the form of the mappings from letters to sounds, from orthographic word forms to spoken words, or of intermediate-level regularities such as bigrams or contextual aspects of words (e.g. onset/rime distinction; Treiman et al., 1995). Because learning and processing are two aspects of the same architecture, the triangle model may be informative not only for established reading but also for understanding the factors that are involved in learning. Coltheart et al. (2001), however, question the benefit of these models as it is unclear to what extent the 'backprop' learning mechanism of a connectionist model reflects human learning (see also Coltheart, 2005). Nevertheless, even without biological plausibility connectionist models have value for demonstrating *in principle* whether sufficient statistical structure is contained in an input to learn the output, without the need for explicit training. Experimental choices in the structure of the training input do impact on what is learned by the model, e.g. consistency effects may be learned as a consequence of the structure of the training items. Even with this in mind the explanatory power of statistical learning in connectionist models remains intact. For instance, Powell, Plaut, and Funnell (2006) trained a connectionist model of reading in an environment more akin to childhood reading: a reduced lexicon, incremental training, and training grapheme-phoneme correspondences. In this training context the learning pattern of the model mimics that of children to a much greater extent. The explanatory power of connectionist approaches depends on the possibility that the same simple learning mechanism could accomplish the learning task, given the correct architect and inputs.

One of the most important aspects of the triangle model in the context of this thesis is how it relates spoken language with orthography. In the Harm and Seidenberg (2004) implementation, although the model uses distributed processing throughout the network to represent orthographic, phonological, and semantic knowledge, the model is pretrained on 'spoken words'. Note however, that while the model contains representations of familiar words in the form of distributed connection weights between orthography, phonology, and semantics; it does not contain explicit, symbolic lexical representations in the same manner as the DRC. While words are represented in a distributed manner across the system, there is no specific orthographic lexicon.

Before training the model on orthographic knowledge, mappings from phonology to semantics are trained, thus introducing statistical structure akin to lexical knowledge, i.e., knowing the arbitrary relationship between the meaning of a word and its phonological form. In this way, the model more closely resembles a novice reader who has already assembled a large store of spoken words. During initial learning the consistent mappings between orthography and phonology are learned rapidly and so the model initially reads by the route ($O \mapsto P \mapsto S$), similar to phonological decoding. With more training, however, the model becomes sensitive to the 'lexical' knowledge in the system. Direct mappings from ($O \mapsto S$) are learned and so the 'division of labour' in the system becomes less reliant on the phonological pathway. Thus even when learning the systematic mappings from letters to sounds, the lexical knowledge in the model is critically involved.

While the triangle model provides an appealing framework in which learning and processing rely on the same architecture, the model conversely struggles to explain the role of explicit reading instruction. Where the DRC model contains explicit rules to map between letters and sounds, learning in the connectionist triangle model is dependent on extraction of statistical regularities. It is not clear how to reconcile this perspective with the instructional methods used in most classrooms.

2.1.4 Themes and questions

These models each capture different aspects of reading. Yet, the patterns described by these computational models display a remarkable overlap with the educational questions noted previously. Briefly, both phonological and semantic aspects of spoken language may play a critical role in learning to read; lexical knowledge supporting orthographic learning. Letter-level learning meanwhile may play a reciprocal role with spoken language in forming orthographic lexical (whole-word) representations. Learning the mappings between letters and sounds may be precursor to learning whole-word representations, or it may be that visual recognition of 'sight words' develops in parallel. Despite the contrasting perspectives offered by the computational accounts of reading reviewed above, similar themes emerge from each. It is specifically to these themes that much of the work of this thesis will be directed.

Learning orthographic units. Learning to recognise letters and to map letters to sounds is perhaps the most fundamental part of learning to read. Nevertheless, neither the triangle model nor the DRC goes into detail on how letters are initially learned. In the DRC the letters are coded directly, while in the implemented triangle model orthographic inputs were represented by vectors of input units. These *slot-based* input units serve their purpose in simulating the presence of an orthographic unit appearing in a written word but do not directly address the question of how orthographic inputs are learned. Indeed one of the consequences of this implementation decision was that positional and alignment information was not learned, e.g., *trust* would not be similar to *mistrust* (Plaut et al., 1996; Sibley et al., 2010). The triangle model as implemented by Plaut et al. (1996) and Harm and Seidenberg (2004) therefore does not capture information about the statistics of orthographic units themselves instead switching units on or off. Extending the triangle model to recode orthographic information points to one mechanism by which the triangle model could learn orthographic information.

One computational account of how orthography might be learned comes from the connectionist architecture of Sibley et al. (2008). This approach used a 'sequence encoder' capable of learning the mappings between letters and sounds as well as recoding words of variable lengths to an equivalent representation (Figure 2.3, p15). When trained on a set of 75,000 words the model was able to generalise its learning to read pseudowords that adhered to orthotactic and phonotactic constraints, but not to illegal strings of letters. Thus the model has learned something of the distribution of letter combinations in the language. This approach demonstrated that connectionist architectures could learn representations of common orthographic structures. Distributional information in a connectionist framework can, therefore, support learning of both lexical and orthographic structure independently.

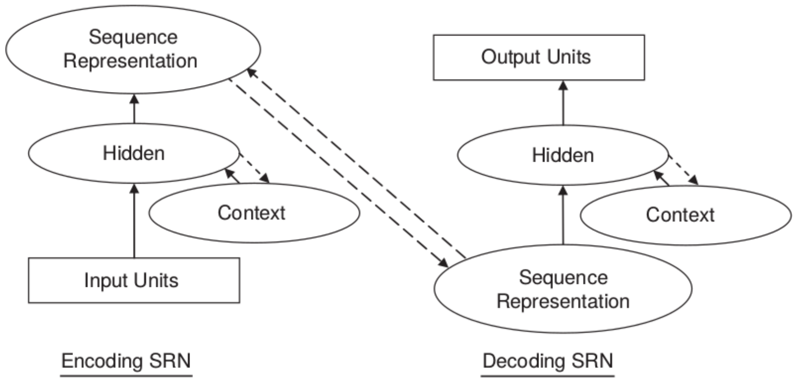


Fig. 2.3 Sequence encoder architecture adapted from Sibley, Kello, & Seidenberg, 2010. Letters are presented sequentially and encoded to a single representation; the sequence representation. The sequence representation is used to generate phonological output units. The error in the output units is backpropagated to the encoding SRN.

Critically, while there are many pre-specified models for how letters might be coded (for review see Grainger, 2008; Grainger, J. and van Heuven, W.J.B., 2004; Ziegler et al., 2010), the approach used by Sibley et al. (2008) learns about orthographic structure directly. The implementation of the model in this sense is less relevant than the insight that written words contain enough statistical structure for a model to learn orthotactic information. Note that all of the computational models summarised here make pragmatic choices for how to code letters, with the goal of modelling the reading process as a whole. Either the DRC or the triangle model could implement an equivalent letter-learning system.

The role of spoken language in reading. One crucial area of agreement across accounts of reading is that spoken language plays a role in supporting reading development. A large amount of evidence suggests spoken language plays a pivotal role in predicting children's later reading ability (Castles et al., 2006; Nation, 2017; Nation and Cocksey, 2009; Nation and Snowling, 2004; Ricketts et al., 2007; Stuart et al., 1999). These findings echo the conclusions of the educational literature that language and literacy are tightly linked. Not only is "literacy is parasitic on language" (Snowling and Hulme, 2012, p597) but moreover, language and literacy may be mutually supportive with each taking "turns to piggy-back on each other during the school years and beyond." (Snow, 2016). This interdependence reflects the sequence of learning in later implementations of the triangle model; only after the model is able to identify spoken words (i.e. $P \mapsto S$) is orthography introduced (Harm and Seidenberg, 2004). A recent connectionist model goes further, attempting to directly investigate the role of spoken language on different types of reading instruction (Chang et al., 2017). This model was trained using mappings that emphasised either ($O \mapsto P$) or ($O \mapsto S$)

mappings while systematically varying the amount of spoken language in the model before training (i.e. $P \mapsto S$). The ($O \mapsto P$) and ($O \mapsto S$) focussed training can be thought of as approximating the phonics and whole-word approaches to reading instruction respectively. In this way, the model sets out to provide an empirical basis for the complex trajectories of learning that can occur under different approaches to reading instruction. The results suggest that the pre-trained spoken language had the biggest impact on training that focussed on learning the ($O \mapsto P$) mapping, paralleling behavioural results using a similar training regime (Taylor et al., 2017).

Despite evidence that spoken language plays an important role in literacy, it remains unclear when and how spoken language supports reading acquisition (Castles et al., 2006; Nation, 2009). Simply, when and where does spoken language interact with written language during learning? The model of Chang et al. (2017) would suggest that the roles of spoken and written language in reading are 'interrelated and interdependent' across development (Diamond, 2007). The experimental approach taken by this thesis, using laboratory-based learning to study the role of memory systems on early reading processes, may thus inform our understanding of points at which spoken and written language interact, e.g., learning the mappings from letters to sounds may be influenced by phonological knowledge whereas learning to link written whole words to meaning may depend on spoken vocabulary.

More than one route to reading. These computational accounts of reading also converge with educational debates in recognising that both lexical and sublexical processing can lead to successful word reading. Proficient readers can decode letter-by-letter to access sound in a systematic manner, or they may be able to recognise whole words. Where the DRC model explicitly captures this process in its two routes, the triangle model accomplishes the same across a single distributed route (Coltheart, 2005). Nevertheless, familiar words in the triangle model may be identified through direct ($O \mapsto S$) mappings, whereas an unfamiliar word might rely on ($O \mapsto P \mapsto S$) mappings. A key question then is how these two processes relate to each other over the time course of learning.

The contrast between learning parts and learning wholes across these models highlights the impact of memory systems in guiding reading strategies. In the DRC model these two approaches are characterised as independent and so we might ask whether both routes develop in parallel, or does one route take priority? What grain size is represented in each route? We might additionally ask whether spoken language plays a role in determining these questions? Critically, the DRC does not explicitly address how learning takes place. It may be the case that children develop whole word representations from the beginning of reading and that the orthographic lexicon grows in scope alongside learning the grapheme-phoneme correspondence rules of the sublexical route (Coltheart 2006; Jackson & Coltheart, 2001).

The triangle model inherently captures features of learning in changing connection weights of the model. In particular, the sequence of training in later iterations of the model highlights the tensions between learning highly correlated systematic mappings (e.g., letters and sounds) and learning arbitrary mappings (e.g., written words to meaning). While the systematic mappings between orthography and phonology allowed the model to rapidly learn the $(O \mapsto P \mapsto S)$ mappings, the arbitrary mappings between O and S took longer. Nevertheless, the $(O \mapsto S)$ mapping was beneficial in disambiguating homophones such as *bare* and *bear*. Despite capturing the relationship between lexical and sublexical processing, the triangle model struggles to explain serial processing during reading (Coltheart, 2006). Moreover, recent attempts to model behavioural performance on statistical learning task using connectionist networks have highlighted a tension between learning parts and learning wholes. When learning parts embedded within a larger whole (visual object) there was a trade-off in the network between representing the whole object, or the component parts (Plaut and Vande Velde, 2017).

In summary, then, none of the computational models of reading provides a complete account of reading in terms of: developmental (learning), behavioural effects (e.g. transposed letter effects, serial word reading effects), and intermediate representations of letter-sound mappings. Rather the key issues identified by these models focus our efforts to understand the role of memory consolidation in the earliest stages of learning to read. Most notably these models highlight (1) the distinction between lexical and sublexical reading strategies. In addition, there is a clear need to (2) clearly track the trajectory by which memory systems interact with different aspects of reading developments, e.g., letter identification, whole word learning, or bigram learning. Most computational evidence places emphasis on the outcomes of learning to read, rather than the process of learning. From an educational view, and to distinguish between models, we should better understand the interdependence of lexical, phonological, visual, and orthographic knowledge in the earliest stages of learning to read.

2.2 Neural basis of reading

In order to understand the process of learning to read it is important to understand the neural systems that support reading. Reading is a culturally transmitted skill that draws upon the pre-existing capacities of the brain, and so learning to read depends on the ability to adapt existing brain functions to learn a visual representation of language (Dehaene and Cohen, 2007; Dehaene et al., 2005). Although computational models of reading do not explicitly attempt to map to the neural underpinnings of reading, there may be some informative parallels between computational accounts and the activation patterns seen in fMRI studies of reading (c.f., Taylor et al., 2013, and see Figure 2.4, p19). In particular, the distinction between letter-by-letter decoding and whole-word recognition may rely on different neural pathways. Hence, changes in activation of these pathways may be an important index of changes in reading strategy during learning.

2.2.1 Neural basis of skilled reading

Hierarchical visual processing in the vOT. The ventral occipitotemporal cortex (vOT) has taken on an important role in the literature concerning the neural basis of reading. This is perhaps unsurprising given the role of the vOT in vision more generally. Projecting from primary visual cortices the ventral and dorsal visual streams are often known as the 'what' and 'where' pathways respectively (Mishkin and Ungerleider, 1982; Ungerleider and Haxby, 1994). In this view the ventral 'what' pathway is involved in the identification of visual inputs, while the dorsal pathway has to do with localisation of visual objects in space. More recent accounts suggest that both ventral and dorsal visual streams capture information about the size, spatial location, orientation, etc., of visual stimuli but differ according to visual perception, and visual control of action (Goodale and Milner, 1992; Milner and Goodale, 1993, 2008). Through a hierarchical process the ventral visual stream maps from the basic visual features represented in the primary visual cortex to higher level visual representations that are independent of size, orientation, brightness, etc. (Kravitz et al., 2013). Thus, the ventral stream can represent partially abstracted visual representations such that the same face viewed from different angles would be represented in the same manner. This hierarchical perspective on visual object recognition has parallels with the lexical/sub-lexical contrast in reading; recognition of parts leads to recognition of wholes.

The special role of the left vOT in reading. The left ventral occipitotemporal (vOT) cortex, including posterior and anterior fusiform, inferior temporal, and lateral occipital regions play an important role in the visual processing of orthographic information (Cohen et al., 2000, 2002; Dehaene et al., 2002). Consistent with the general view of the vOT as a hierarchical visual stream, a variety of evidence suggests reading processes are hierarchically organised. Thus, posterior temporal and occipital regions process early componential representations of individual letters and letter sequences while anterior regions process holistic representations of whole words (Dehaene et al., 2005; Taylor et al., 2013). In addition, Mechelli et al. (2005) found that posterior fusiform activation was greater for pseudowords than for irregular words such as 'pint', whereas anterior fusiform showed the reverse profile.

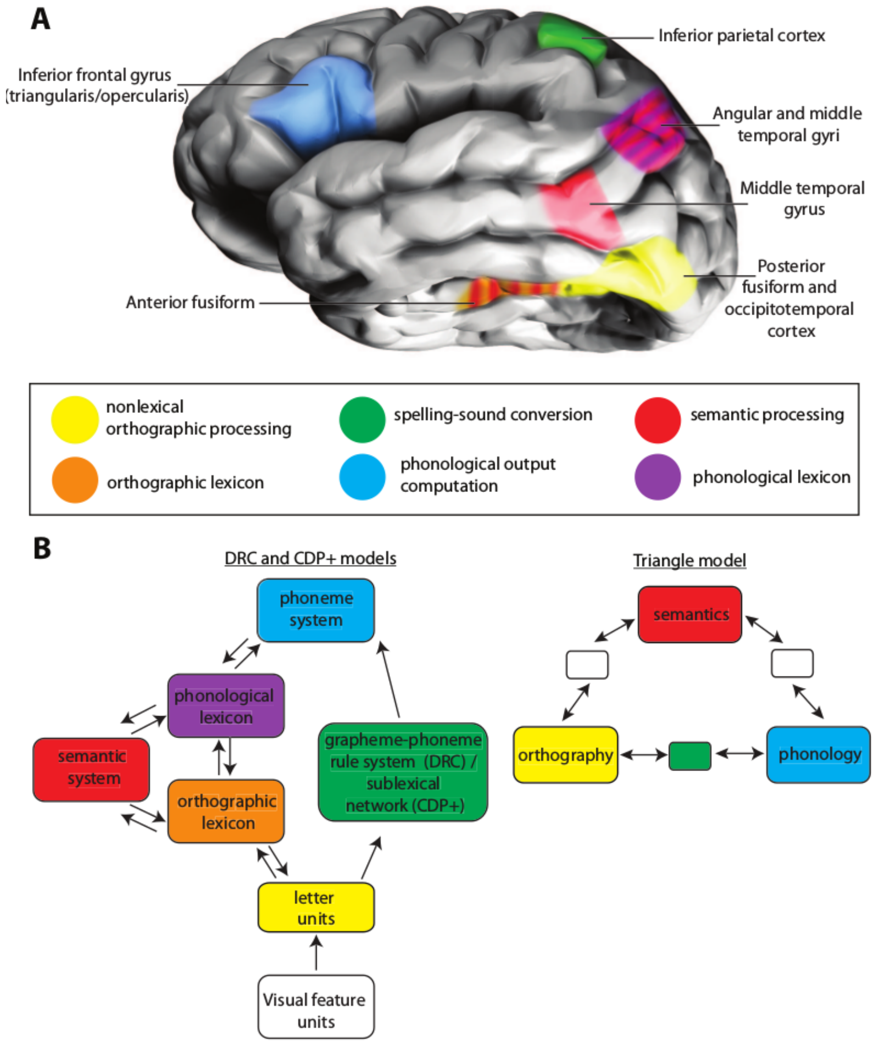


Fig. 2.4 Proposed mapping between cognitive models of reading and findings from a meta-analysis of fMRI studies of reading. Adapted from Taylor, Rastle, & Davis, 2013

Vinckier et al. (2006), meanwhile, showed a hierarchy of neural representations for different types of letter strings in vOT: posterior vOT activated for all visual stimuli (including consonant strings and false fonts), whereas mid- to anterior-fusiform regions were only activated for letter sequences that contained familiar letter combinations. Seghier et al. (2008) found that adult readers who were slower at reading pseudowords than irregular words showed additional activation in both left inferior parietal and left posterior occipitotemporal cortices, reflecting increased effort in componential reading processes. In contrast slower reading of irregular words was associated with increased activation in left anterior occipitotemporal and left ventral inferior frontal regions. These findings support the idea that posterior fusiform and occipitotemporal cortex process parts of words whereas anterior fusiform processes whole-word forms.

Despite acceptance of the important role of left vOT in reading, debate continues concerning whether this vOT hierarchy includes brain regions that uniquely contribute to reading (Dehaene and Cohen, 2011), or are shared with other domains in which visual and phonological information is associated, e.g. object naming (Price and Devlin, 2011a). The so-called visual word form area (VWFA) has been put forward as a region that adapts during the process of learning to become specialised for reading (Cohen et al., 2000, 2002; Dehaene et al., 2002). The neuronal recycling hypothesis suggests that since reading is a cultural tool and so cannot be an evolved trait, regions of cortex with existing functional specialisation become adapted to reading over the course of development (Dehaene and Cohen, 2007). Collectively these studies offer strong evidence that this region of the ventral visual stream responds reliably to a range of orthographies (Bolger et al., 2005), more or less transparent orthographies (Mei et al., 2013a), and that activation in the area is robust against variation in size, case, font etc., (Dehaene, Cohen, Sigman & Vinckier, 2005). Additionally responses in this area are sensitive to lexical and sub-lexical properties of the stimuli, such as bigram frequency (Cohen and Dehaene, 2004). However, despite the consistency of results showing that the so-called ‘visual word form area’ responds to orthographic stimuli there has been substantial disagreement to the claim that the area is specialised in some way to orthographic stimuli. Instead, Price and Devlin (2003, 2011a) propose that this area mediates between bottom-up visual stimuli and top-down linguistic and semantic inputs, explaining why consistent activation may be seen in this region without the need to appeal to functional specialisation for orthography. (Price and Devlin, 2011a) argue that since accessing phonology for written words is automatic (Hagoort et al., 1999; MacLeod, 1991; Price et al., 1996; Xue et al., 2006), passive viewing tasks impose a top-down phonological demand for word reading but not for object naming (Song et al., 2010a; Twomey et al., 2011; Yoncheva et al., 2010). As several studies (Kherif et al., 2011; Mano et al., 2013) have shown that responses in fusiform regions are sensitive to the task demands, any conclusions about brain activation in this area must consider these task demands carefully.

Parietal involvement in reading. In the dorsal visual stream, the parietal cortex plays an important role in reading, though the precise nature of this role is debated. Parietal regions show involvement in a range of tasks involving allocation of visual attention and when combining inputs from different modalities (Bitan et al., 2006, 2007c; Cohen et al., 2008; Shafritz et al., 2002). Similarly, Lobier et al. (2012) argue that parietal visual attention processes may support pre-orthographic character recognition. Using a task designed to isolate activation associated with identification of multi-character elements from different orthographic systems showed increased activation in superior parietal regions, consistent with suggestions that this region supports binding of multiple elements. Additionally, Cohen et al. (2008) degraded visual aspects of 4-, 5-, and 6-letter words by parametrically altering the rotation, letter-spacing, and lateral position of the words relative to the fixation. Parietal activation increased according to the level of degradation in the written word, but moreover, a word-length effect was found only for stronger levels of degrading. This moderation of the word length effects by visual familiarity highlights the importance of parietal regions to reading specific processes. Finally, Vinckier et al. (2006) identified a patient with bilateral occipitoparietal lesions who was unable to read familiar words if some visual features were changed, e.g., rotated letters.

Atypical parietal activation in reading has been found in groups with reading disorders (Peyrin et al., 2011, 2008). Children with dyslexia showed decreased parietal activation when compared to age-matched controls on a visual attention task (Peyrin et al., 2011). This group difference was only apparent on conditions that involved processing of multiple characters simultaneously leading the authors to suggest poor visual attention may be a contributing factor to reading deficits. Extending this research to adults with dyslexia Reilhac et al. (2013) showed reduced activation in superior parietal lobules (SPL) when detecting letter identities in a string of characters. One interpretation of these findings is that differences in parietal activation for poor readers may simply be a consequence of poor visual attention, rather than a contributory factor that leads to poor reading (e.g., Boros et al., 2016). A contrasting interpretation might be that allocation of visual attention is a key process in the early stages of reading. Thus, although altered parietal activation is typically seen in relation to poor readers in adulthood, it may play a critical role when learners are still engaged in serial decoding of unfamiliar written forms. The following section (Section 2.2.2, p22) expands on this idea with research showing parietal activation when children learn to map letters to sounds.

Carreiras et al. (2014) suggest the left parietal cortex contributes not only to identification of letters but also to coding of letter position within a word. Participants judged whether two consecutive four-character strings were the same or different, with 25% of items containing a character replacement, and 25% contain a character transposition. Characters could be either letters, digits, or symbols. Results showed increased activation in left parietal regions for letters relative to digits or strings. Thus these findings indicate that letter-position coding in the parietal cortex is not merely a reflection of general visual processing. Other evidence for parietal involvement in letter position coding comes from Friedmann and Gvion (2001) who reported 2 patients with parieto-occipital lesions who were impaired in encoding letter position, e.g., reading *board* as *broad*. Other than the letter-position impairment, however, the patients were successful at many other reading tasks, e.g., lexical decision and nonword reading. Convergent evidence from Ossmy et al. (2014) combined voxel classification with functional connectivity analysis to explore the impact of letter-position processing on brain activation. The whole brain classification analysis pointed to a parietal region (the intraparietal sulcus) where activation patterns encoded sufficient information to discriminate to 80% between two Hebrew words that differed only by letter position. In addition, the functional connectivity analysis found connections between this region and left vOT cortex. Cross-cultural support for parietal involvement in reading comes from Bolger et al. (2005) who collated brain imaging studies of reading across different writing systems, highlighting increased activation in left inferior parietal cortex in alphabetic languages compared to logographic languages. Finally, a meta-analysis by Taylor et al. (2013) of 36 fMRI studies of word and pseudoword reading showed parietal activation increased for pseudoword reading and is interpreted as reflecting the process of converting letters to sounds.

Inferior frontal gyrus contributions to reading. A third region that plays an important role in reading is the left inferior frontal gyrus (lIFG). The IFG has been implicated in both semantic and phonological selection (Bookheimer, 2002; Devlin et al., 2003; Fiez, 1997; Gough et al., 2005) and studies of reading have shown a similar pattern for word reading (e.g., Price and Mechelli, 2005). This region was also highlighted in the Taylor et al. (2013) meta-analysis as showing higher levels of activation both when reading pseudowords compared to words, and when reading irregular compared to regular words. Taylor et al. suggest that the lIFG plays a role analogous to the phonemic units of both the DRC and triangle models by resolving phonological ambiguity. More recently Pattamadilok et al. (2017) compared detection of visual (symbol detection), phonological (rime detection), or semantic content of written words using a go/no-go task. Activation in IFG pars triangularis and IFG pars opercularis was greater for the semantic and phonological judgements than for the visual task. This region might then be thought of as mediating the convergence of two routes to reading, and so there is increased activation for irregular compared to regular words as these two routes produce different pronunciations in the case of irregular words.

2.2.2 Neural perspectives on learning to read

There are two broad methods by which neuroscientists have studied the brain changes associated with the emergence of literacy (see Dehaene et al., 2015, for a review): tracking children at different stages of learning to read and comparing literate, illiterate, and ex-literate adults. Turning first to children, activation in vOT to words has been shown in young children in tasks involving sub-lexical processing such as single letter naming (Turkeltaub et al., 2008) and associating letters with sounds (Brem et al., 2010), but also for lexical tasks such as single word reading (Church et al., 2008). Furthermore, a meta-analysis of 40 imaging studies showed that both child and adult readers showed activation in left vOT, inferior frontal, and posterior parietal regions (Martin et al., 2015). However, there were also age-related differences: activation was more consistently observed in posterior fusiform regions for adult than child readers, possibly reflecting increased sensitivity in adults to the differences between letters and control stimuli. Tracking neural changes in a single group of children over four years, Ben-Shachar et al. (2011) showed that the sensitivity of left vOT to written words increased as reading improved and that this was correlated with sight word naming accuracy but not with measures of pseudoword reading. Furthermore, the spatial extent of the cortical region sensitive to visual words increased as children got older before decreasing until reaching adult level. This changing response may reflect the region initially becoming more engaged for orthographic inputs before later becoming more efficient as specialisation takes place, following an inverted-u shaped profile (Ben-Shachar et al., 2011; Price and Devlin, 2011b). Taken together, these results suggest that mid-vOT regions become more sensitive to orthographic information with increased age/proficiency but it is not clear whether this change is linked to holistic or componential reading processes.

Parietal activation in children has primarily been shown in tasks involving mappings between visual words and sounds, (e.g., Bitan et al., 2007a, 2006, 2007c; Cao et al., 2006; Hoeft et al., 2007). For example, children making spelling (orthographic) or rhyme (phonological) judgements about visually presented words showed increased activation in bilateral inferior/superior parietal lobules for spelling compared to rhyme judgements (Bitan et al., 2007a). Likewise Hoeft et al. (2007) found that activation in left inferior parietal lobes correlated with composite behavioural measures of phonics ability in children. Further evidence that parietal regions support the componential aspects of reading early in development comes from Cao et al. (2015) who compared adult and child English and Chinese speakers in a visual word rhyming task. Reading skill in English speaking children was correlated with activation in left inferior parietal lobule. The same was not true for Chinese speaking children, lending support to the idea that early reading in English, with its reliance on componential letter-sound mappings, engages left parietal regions more than logographic reading in Chinese readers. One problem with studies comparing children and adults is that it can be difficult to distinguish neural changes due to increased proficiency from changes due to maturation. Converging evidence shows a positive correlation in children between nonword reading and activation in left inferior parietal and left inferior frontal gyrus (McNorgan et al., 2011).

Ben-Shachar et al. (2007) find increasing sensitivity in vOT to words hidden in varying levels of visual noise, but also to false font stimuli and objects. By adopting a measure of the changes in activation within condition according to varying degrees of visual noise this study attempts to minimise the effect of phonological processes on BOLD response. Critically responses to vOT increase even for false-fonts as the visual noise decreases, suggesting the changing activation is due to visual changes, rather than phonological or semantic retrieval. They conclude that there is indeed specialisation for the visual inputs of words but that this specialisation was not to the exclusion of other inputs, e.g. visual objects. Extending this line of research to track the changes that occur naturally over development Ben-Shachar et al. (2011) followed a group of children over four years, tracking the changes in cortical sensitivity to written words. BOLD responses to a word-visibility task where words were hidden in parametrically modulated visual noise were compared to behavioural measures of reading. Results provide strong evidence that the sensitivity of the vOT to visual word inputs does change in time with improving reading.

While Ben-Shachar and colleagues (2007, 2011) interpret these data as supporting the view that vOT does change in its sensitivity to visual inputs, the authors clarify that the involvement of vOT in visual shape extraction is not limited to words, responding to other stimulus categories also. Although the approach of comparing sensitivity within category in varying levels of noise is a useful approach for exploring the visual sensitivity of this area, the approach also rules out the effects of the task on activation. In addition to sensitivity to visual stimuli, it may be the case that there is additional activation due to these task demands.

One way to circumvent this problem is to study functional and structural changes in adults who learned to read later in life. Dehaene et al. (2010) found that, compared to illiterate adults, both adults reading from childhood and late-learners showed greater activation to written words in left fusiform gyrus as well as language regions such as left superior temporal sulcus and left inferior frontal gyrus. Furthermore, adults who learned to read later in life showed increased grey matter in angular gyri, dorsal occipital, middle temporal, supramarginal, and superior temporal gyri, in comparison to illiterate adults (Carreiras et al., 2009).

In summary, evidence from studies of beginning readers and ex-illiterate adults has shown increased contributions of vOT regions with increased reading skill. It remains unclear, however, whether these contribute to holistic or componential processing of written words. Evidence for componential processes seems to point to inferior parietal regions which might play a preferential role in initial stages of acquisition. This might be taken as consistent with the componential, phonics-based educational literature introduced at the outset which similarly suggests that initial stages of teaching should focus on componential decoding skills. One possible challenge, however, is that these studies with children and adults have only explored relatively late stages of reading acquisition. It would be extremely difficult to attempt to scan children in their first months of literacy learning (in the UK, this would require scanning 4-year old children since reading instruction begins at that age). Therefore, in the present work, we explore the initial stages of reading instruction for adults learning to read in an artificial orthography. To the extent that changes in vOT and parietal brain activity for holistic and componential learning parallel activation seen during reading development we may be confident in attributing neural changes to the balance of these two underlying processes.

2.3 Artificial reading studies

Artificial reading studies teach people different aspects of reading in an experimental setting that allows precise control over the learning conditions. The benefits of adopting artificial learning approaches have been noted in relation to learning spoken words (Wonnacott et al., 2017), syntax (Wonnacott and Newport, 2005; Wonnacott et al., 2008), morphology (Merkx et al., 2011), and reading (Takashima et al., 2016; Taylor et al., 2011). Indeed using artificial learning may be particularly informative for a developmental process such as reading where it is difficult to disentangle the interrelated and interdependent factors that support reading (Wonnacott et al., 2017). Nation (2009) argues that the combination of experimental methodology with a developmental perspective offers an important avenue to advance our understanding of reading.

Laboratory-based learning paradigms offer a valuable method to address questions about reading, allowing a degree of experimental control impossible to achieve in naturalistic learning situations. Reading in a naturalistic context is influenced by a wide range of psycholinguistic features: word frequency (Forster and Chambers, 1973), regularity/consistency of spellings (Seidenberg and McClelland, 1989; Waters and Seidenberg, 1985; Zevin and Seidenberg, 2006), and phonological consistency (Ellis and Hooper, 2001; Ziegler and Goswami, 2005; Ziegler et al., 2001) amongst others. The combined effect of these factors over the time course of development means that it is difficult to untangle how each influences reading development. Tracking reading acquisition in an experimental setting allows insight into these processes, e.g., the interaction of word frequency and spelling consistency during learning (Monaghan and Ellis, 2002; Taylor et al., 2011). For example, McKay et al. (2008) used novel words written in a familiar orthography to show that semantic information plays an important role in reading even recently learned words. Participants were faster and more accurate to read novel words that were accompanied by a definition than for novel words in isolation. This semantic advantage was greater when the novel words had inconsistent spellings, lending support to the view that semantic information supports reading in cases where letter-to-sound mappings may fail. Moreover, Taylor et al. (2011) note that generating artificial stimuli makes it possible to avoid some of the issues that come about when using a restricted set of naturalistic words while trying to control for all relevant psycholinguistic factors. In addition, Hirshorn and Fiez (2014) highlight the strengths of using artificial orthography studies to make comparisons between different writing systems. While the Hirshorn and Fiez focus on the value of artificial orthography approaches to understand cross-linguistic differences the same benefits apply here, when identifying the factors that influence learning to read. Castles et al. (2006), meanwhile, note the apparent circularity of conclusions when trying to separate sublexical from lexical orthographic knowledge in many studies. Lexical knowledge may influence sublexical learning, and vice versa, making clear conclusions difficult. Artificial reading studies, therefore, allow traction on these questions by allowing (1) control over the stimuli themselves, and (2) control over the sequence of learning itself. Thus it may be possible to track the transition from serial letter-by-letter decoding to whole-word recognition, or in the neural domain to investigate the emergence of orthographic specialisation in the ventral visual stream.

2.3.1 Behavioural studies using artificial orthographies

While there are relatively few behavioural studies that have used artificial orthography paradigms, these studies are particularly informative for this thesis. Two early experiments explored the ability of participants to implicitly learn letter-sound mappings of novel written symbols (Bitan and Karni, 2003, 2004). Participants learned to recognise three sets of six nonwords through either an explicit letter-sound instruction condition, an implicit alphabetic condition (where each letter mapped to a single phoneme), or an arbitrary mapping condition. Although participants learned to recognise the whole words in all three conditions, they could not generalise the letters well in any condition. The ability to map from letters to sounds when reading an unfamiliar word is a key feature of reading. This outcome may reflect the stimuli rather than a failing of artificial orthography in principle; the symbols used in the experiment were all made up of three elements, e.g., the letter **f** mapped to the symbols **)<-** while **d** mapped to **)-<**. The complexity of these symbols may have prevented participants from picking up on the mappings from symbol to sound as they had to visual parse 9 elements (three elements in each of three symbols). Readers of English are familiar with complex graphemes, e.g., the two graphemes ‘th’ and ‘igh’ in the word ‘thigh’ map to the phonemes /θ/ and /ai/ (Coltheart, 2005). Nevertheless, learning the very complex mappings of Bitan and Karni (2003) may not reflect how children learn to read.

The ability to finely control stimulus characteristics allows examination of otherwise obscured aspects of learning. For example, Chetail (2017) adapted an artificial orthography paradigm to a statistical learning framework. Participants were exposed to a stream of 320 written word forms each made up of 5 characters. Importantly, participants did not learn the letter-sound mappings of the visual words. Instead, the focus was on how visual statistical learning leads to word recognition. To this end the artificial words had embedded bigram pairs occurring either always in onset position, or in the middle of the words. After only a few minutes of exposure to the visual stream of words was enough for participants to become sensitive to orthographic features such as bigram frequency and letter position. Even without the inclusion of a phonological element to training, orthotactic structures can be learned. These findings echo the pattern seen in children where sensitivity to orthographic structure emerges rapidly (Cassar and Treiman, 1997).

These studies highlight the importance of stimulus characteristics in determining learning. Whereas Chetail (2017) taught 320 words made up of 22 letters, Bitan and Karni (2003) taught only 6 items. Perhaps not surprisingly the six items of Bitan and Karni did not give participants enough varied exposure to the component letters for those letters to be learned independently of the lexical item; participants recognised the whole word forms but could not use the letters to read. By contrast, the 320 items used by Chetail, combined with the lack of corresponding spoken word forms, means that participants would struggle to recognise items at the whole word level. A theme of throughout the thesis is the tension between componential knowledge (mapping from component letters to sounds) and holistic knowledge (recognising whole words) in reading. We are interested in how knowledge of letters leads to knowledge of lexical forms, and vice versa.

Taylor et al. (2011) meanwhile demonstrated the strengths of adopting artificial learning paradigms in allowing sufficient experimental control to tease apart the complex interactions between different factors that determine word learning. This series of experiments showed participants could acquire and generalise novel letter sound mappings in order to read unfamiliar words. Moreover, participants' learning was sensitive both to the frequency and consistency of words; vowels that involved inconsistent letter-sound mappings required more exposure to enable word-specific learning. The addition of semantic and phonological manipulations further demonstrated semantic but not phonological benefits when reading novel words. This finding seems at odds with phonics-based accounts of word learning. The study by Taylor et al. (2011) most directly links to the question of how word-specific knowledge interacts with the letter-to-sound mappings during the learning process.

A number of studies have additionally used artificial orthography paradigms alongside manipulations of overnight consolidation. Memory consolidation offers an interesting perspective on learning and one that will be key to this thesis. The following section of the literature review (Section 2.4: Memory Consolidation, p30) will address how artificial learning paradigms have informed the consolidation literature.

2.3.2 Brain studies using artificial orthographies.

Tracking how the brain changes when learning to read in a natural context is challenging; it would be extremely difficult to acquire scanning data from 4-year-olds as they first encounter text. Studies of adult reading meanwhile tend to capture established reading, rather than the learning process. Training a novel orthography in context of an fMRI study, by contrast, allows insight into the neural changes that take place.

For example, Xue et al. (2006) tracked changing patterns of fusiform activation as different types of learning took place. Over several weeks, participants were trained to read a logographic orthography using a visual familiarity task, followed by phonological and semantic training. Scanning before, during, and after training showed a complex relationship between visual and phonological familiarity. Reductions in vOT activation to increasing visual familiarity suggestion visual orthographic learning is taking place. Greater visual familiarity with the orthography led to reduced vOT activation, perhaps due to emerging visual representations. However, subsequent phonological training led to increased activation, consistent with views that vOT engages phonological representations during learning. The experience-dependent changes seen here are compatible with both views of vOT function reviewed earlier; developing visual specialisation and engaging top-down phonology (Dehaene and Cohen, 2011; Price and Devlin, 2011a). Artificial reading studies demonstrate that vOT activation is not merely visual in nature but reflects additional phonological processing. Hashimoto and Sakai (2004) trained participants to associate Hangul characters with either speech sounds or tones. Activation in posterior Inferior Temporal Gyrus was increased only in the condition where the characters could be mapped onto speech sounds.

Taking advantage of the ability to control the sequence of learning allowed by artificial learning tasks, Mei et al. (2013a) taught two groups of participants to read the same stimuli as either alphabetic or logographic (word-to-sound) mappings over the course of eight days. Both of these conditions involve retrieving phonology in a visual-verbal association but differ in terms of the orthographic transparency. Imaging results showed that over the course of training the more transparent orthography became more left-lateralised, particularly in posterior fusiform regions. Thus, the same visual inputs led to different patterns of activation according to the type of 'reading instruction' given. Complementary findings have reported similar left lateralised fusiform responses for componential as opposed to holistic learning of artificial orthographies using ERP measures (Yoncheva et al., 2015). The convergence of the imaging findings with the patterns seen for reading real words and pseudowords (e.g., Vinckier et al., 2006) highlights the contrast of holistic and componential reading strategies.

Taylor et al. (2014a) directly investigated holistic and componential processing of visual-verbal mappings by comparing behavioural and neural responses while participants learned to name novel objects and to read novel words written in a novel orthography. While learning to name objects and learning to read words both involve associating visual and verbal representations of items, the underlying processes differ in ways that inform our understanding of reading. The relationship between letters and sounds in an orthographic writing system is systematic and componential; readers can decode unfamiliar words, e.g. ‘spape’, using their knowledge of letter-sound-mappings. By contrast, it is not possible to deduce the name of an unfamiliar object by examining its visual appearance. When we refer to object recognition as ‘holistic’ therefore we are addressing the fact that the whole object must be identified, and that an unfamiliar object cannot be named by recognising its constituent parts. In this sense, naming objects has parallels with recognising a familiar whole word without having to decode its component letters. This study is of particular relevance to the thesis as Experiment 1 adapts the experimental paradigm used here to an overnight consolidation design.

Imaging results from learning runs show activation in bilateral parietal cortices when learning the componential letter-sound mappings in reading while mid- to anterior-fusiform regions support learning of whole items for object naming. More posterior fusiform regions showed no difference between words and objects during learning. Additionally, activation in a left mid-fusiform ROI was predictive of successful object learning, but not word learning.

Notably, the testing runs from Taylor et al. (2014a) used a novel imaging strategy to isolate the process of retrieving holistic and componential phonological information. During scanning two trial types appeared, a see-think trial and a see-speak trial. During see-think trials participants were instructed to covertly recall the phonological form but not to articulate it aloud. During the see-speak trials, participants named the item aloud. If an item appeared in a see-speak trial, the same item would have appeared in the see-think condition on the previous trial. There were several reasons for this interaction analysis. First having just recalled the item’s name allowed participants to name the item aloud in the gap between scans. More importantly, this approach allowed the separation of the neural activation associated with the retrieval of phonology, from that associated with the visual differences of the stimuli, and the processes of articulation (discussed in more detail in Experiment 1) by examining the interaction between the two conditions. The consequence of this approach is that this experimental paradigm allows a close examination of the process of retrieval of componential and holistic visual-verbal mappings.

During the testing run clusters were only found for words more than objects, in bilateral inferior and superior parietal cortices and left superior frontal gyrus. It is important to note however that since the majority of scanning involved learning this testing run may be relatively underpowered in comparison to the training runs. These findings offer an interesting perspective on results previously seen in vOT regions by comparing completely novel orthography and objects during initial learning. The use of the interaction analysis allowed the processes associated with retrieval of phonology to be isolated while minimising the effect of visual differences between stimuli. Instead, the results suggest that the location of activation in the ventral visual stream depends on whether that item is being processed in a componential manner, looking at the elements within it, or in a holistic manner by taking in the item as a whole.

The research reviewed above suggests artificial orthographic research has an important role to play in allowing investigation of how specific linguistic properties impact on the learning and processing of writing systems. In particular, this approach allows manipulation of the factors identified in the review of computational models of reading, and of fMRI studies of reading: knowledge of letters, knowledge of whole-word orthographic forms, knowledge of spoken words, and consistency of intermediate-level aspects of written words (e.g. bigrams). The next section reviews the impact of memory systems in learning novel linguistic information. We suggest the memory consolidation framework may provide a way to track the initial acquisition of orthographic representations.

2.4 Memory Consolidation

Research on memory consolidation forms the second theoretical strand of this thesis. This literature reveals the impact of human memory constraints on learning and, therefore, provides perspective on learning to read. Of particular interest are findings on the role of memory consolidation on learning language; although the largest acquisition of language happens in the early years of childhood, new words join the lexicon throughout adulthood. The following review highlights a range of evidence that memory consolidation supports the acquisition of new linguistic information, but also highlights the distinct lack of research on how this process might inform our understanding of reading acquisition.

To frame the importance of this theoretical perspective for this thesis, the experimental work of this thesis adapts the memory consolidation literature to explore how written knowledge is learned, drawing upon sleep consolidation as a key experimental factor in order to explore the impact of memory systems on learning to read. The literature review which follows shows that consolidation is not only a critical element of learning, but can also tell us about the structure of what is being learned. Using sleep consolidation to track the relative emergence of lexical and sublexical knowledge in reading may thus provide a means by which we can link artificial reading research with naturalistic reading.

Hebb's seminal work on the formation of new memories characterised learning as the result of changes in the brain (Hebb, 1949). The simple mechanism proposed to account for these changes was a coincidence detection rule in which the activation or firing of one neuron will stimulate activation of a neighbouring neuron, triggering metabolic changes in both over time. As well as motivating the phrase 'neurons that fire together, wire together', these insights have been born out by modern neuroscience in the form of spike-timing-dependent plasticity (see Caporale and Dan, 2008, for review). Evidence from neuropsychology has suggested that memory formation relies on more than one system of learning, however. Henry Molaison, known as 'patient HM' has been used for many years as an example of the separability of memory formation from memory recall. After losing his hippocampus and surrounding medial temporal lobe structures in a surgery to correct epilepsy Molaison developed severe anterograde amnesia and was unable to form new memories (Scoville and Milner, 1957). Despite this impairment, he was capable of recalling information learned up until sometime prior to his operation. Critically, memories acquired in the run-up to the operation were also lost. This 'temporally-graded retrograde amnesia' indicated that some qualitative shift in representation distinguishes older memories from recently acquired ones. These long-term changes in the representation of knowledge that take place subsequent to initial learning have been termed 'consolidation'. McClelland et al. (1995) note, however, that this is a descriptive term that does not explain why the phenomenon takes place.

2.4.1 Complementary Learning Systems

Complementary learning systems theory (CLS, McClelland et al., 1995) attempts to explain the mechanism of consolidation effects through the interactions of two complementary learning systems: a rapid, context-dependent memory system for encoding specific episodes, and a context-independent system that learns more slowly, but allows for overlapping representations which can be generalised outside of the context of learning. Evidence for this account emerges from the parallels between neuropsychological evidence and patterns of learning in connectionist networks (McClelland et al., 1995).

Connectionist models have the ability to extract the shared structure from many instances of learning, capturing this knowledge in the connection weights across the network. The explanatory power of these models lies in their simplicity. Each 'neuron' in the network activates according to the inputs it receives from connections to other 'neurons' in the network, and the weight accorded to each connection. Learning takes place over time, with small changes in connection weights encoding information from multiple learning instances. In this way both learning and knowledge representation could take place within the same network.

Early connectionist models highlighted undesirable consequences of both learning and knowledge representation taking place within the same network architecture; learning that takes place later in time tends to have deleterious effects on information learned previously. As knowledge is stored in the connection weights between 'neurons', changing these connection weights risks interfering with previous learning. This 'catastrophic interference' between old and new learning posed a challenge for these models to account for human learning and memory (McCloskey and Cohen, 1989). Models of learning that attempt to capture human behaviour should show plausible patterns of both learning and forgetting (French, 1999).

While various network architectures have been proposed to minimise catastrophic interference (e.g., French, 1992), McClelland et al. (1995) argued that human neuropsychology solves this challenge through interactions between hippocampus and cortical networks. New information is encoded initially in a short-term hippocampal representation (Frankland and Bontempi, 2005; Squire and Alvarez, 1995) (Figure 2.5, p33). Over a period of days, weeks, or months, interactions between the hippocampus and cortical structures gradually transfer the memory representation to a distributed cortical system. Over time, the reactivation of memories promotes strengthened connections between neocortical representations (Hebb, 1949; Marr, 1970; Squire and Alvarez, 1995), forming stable long-term representations that are independent of the hippocampus. Connectionist modelling of memory formation suggests this process may play a protective role for existing memories by preventing 'catastrophic interference' between new and existing memories (French, 1999; McClelland et al., 1995; McCloskey and Cohen, 1989). These memories systems therefore serve complementary roles, allowing rapid acquisition of novel memories without disruption to existing knowledge (Dudai, 2004).

Offline consolidation and the special role of sleep. The reactivation and consolidation of newly learned information can take place both during wake and during sleep. For the purposes of this review, we focus primarily on offline (sleep-based) consolidation both because sleep is one of the primary mechanisms by which this reactivation of memories is thought to take place, but also because it can easily be controlled in an experimental design. Although the full process of consolidation plays out over days and weeks, experimental evidence suggests

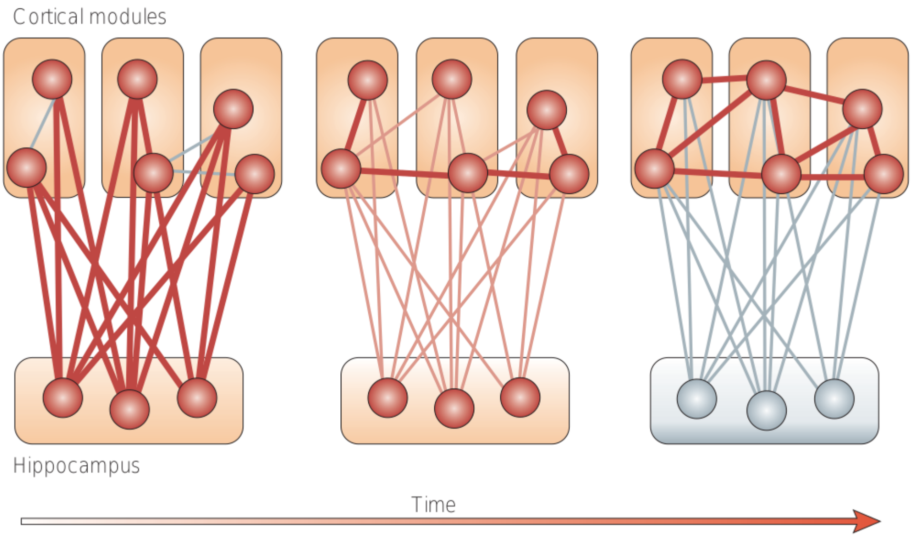


Fig. 2.5 Model of memory consolidation reproduced from Frankland & Bontempi (2005). Initial hippocampal encoding of new memories are gradually transferred to cortical networks. Over time repeated reactivation of the memory (for example during sleep) promotes formation of intra-cortical connections independent of hippocampus.

that the largest changes occur in the day after initial learning (see Diekelmann and Born, 2010; Feld and Born, 2017; Walker and Stickgold, 2004, for detailed reviews).

The impact of sleep on memory has been shown through a variety of experimental paradigms with results typically showing improvements due to sleep (Gais et al., 2002; Plihal and Born, 1997; Tucker et al., 2006) or protection against interference (Ellenbogen et al., 2006; Korman et al., 2007). Recording of biological markers of sleep stages provides more direct evidence for sleep-related changes in memory representations. For example sleep spindles, a characteristic pattern of sleep architecture, increase when learning new words (Gais et al., 2002; Tamminen et al., 2013) while targeted reactivation of newly learned words during sleep mediates the relationship between REM sleep and later memory performance (Tamminen et al., 2017) (Figure 2.6, p34).

Animal studies have provided important electrophysiological evidence for offline consolidation, linking brain and behaviour directly. In rats, hippocampal reactivation during sleep has been shown location-specific cells firing both during learning and in subsequent sleep (Wilson, 2002; Wilson et al., 1994). Electrophysiological evidence from rats has shown linked activity between hippocampal and neocortical regions during slow wave sleep, with the pattern of activation in both regions matching that seen during initial learning of spatial information (Qin et al., 1997; Sutherland and McNaughton, 2000). Likewise Ji and Wilson (2007) showed similar finding between the hippocampus and visual cortex in rats after a

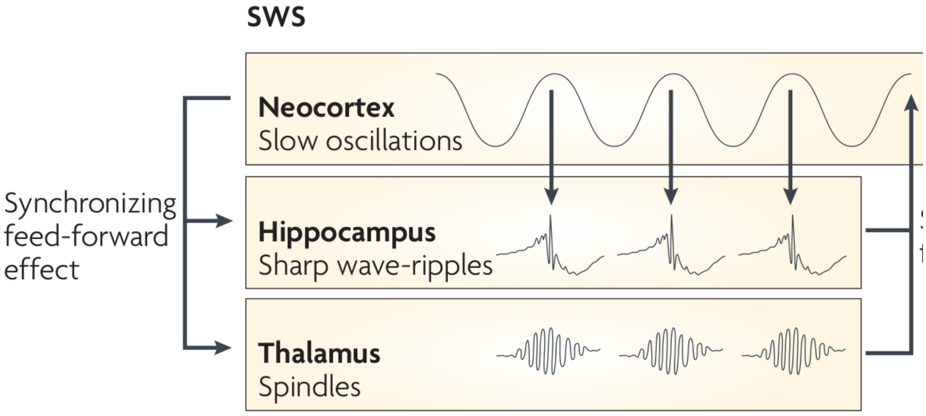


Fig. 2.6 Oscillatory activation for memory reactivation during sleep, reproduced from Diekelman & Born, 2010. Oscillatory activity during slow wave sleep drives syncronised activation between thalamus, hippocampus, and neocortex. These coincidental activations promote synaptic plasticity, and thus learning.

period of visual learning. Moreover role of sleep consolidation in animals has been observed to impact both on behaviour and neural activation: offline reactivation of the hippocampus after novel learning predicts later ability to retrieve memories in mice (van de Ven et al., 2016) while conversely dendritic spines have been observed to form during sleep relative to the performance of the animal in learning the behavioural task (Yang et al., 2014). The elements of sleep architecture indirectly identified through the behavioural and polysomnographic study of human learning, therefore, find parallels in the biological basis of sleep in animals.

2.4.2 Spoken word learning in the CLS

Davis and Gaskell (2009) adapt the complementary learning systems (CLS) account to explore how memory systems may contribute to the acquisition of novel spoken words. Combining the CLS perspective of learning with the distributed cohort model of spoken word recognition (Gaskell and Marslen-Wilson, 1997) allowed the authors to make predictions about the mechanisms of learning novel words, as well as the neural and behavioural changes that may occur as a new word enters the lexicon. The CLS account of word learning suggests that initial sparse word representations are encoded in hippocampal/medial temporal regions. In this way, the episodic context in which the word was learned supports recognition of that word. However, at this stage these novel words have not yet been incorporated into the broader lexicon of spoken words and so will not enter into competition with other spoken words (as predicted by the distributed cohort model). By allowing for a division of labour between fast and slow learning systems, the CLS model provides a way to avoid this 'catastrophic interference' by interleaving new information with old and allowing a gradual integration process (French, 1999; McClelland et al., 1995; McCloskey and Cohen, 1989). As discussed above, a primary theoretical motivation for consolidation of newly learned information is to avoid interference in a connectionist network. The interference between old and new information in connectionist models has been suggested to be incompatible with how learning takes place in children; learning new words does not impede or degrade knowledge about words children have learned previously (Coltheart, 2006; Dufau et al., 2010). This is an important and valid concern. Connectionist models are appealing in their ability to demonstrate that the interaction of many simple 'neurons' in a network architecture can learn and represent complex statistical relationships without reliance on explicit training, symbolic representations, or hard-coded rules. Nevertheless, if the properties of such networks are fundamentally at odds with how learning occurs in children then the model may have little value. Consolidation of new information in dual-system accounts of learning offers a mechanism by which connectionist networks can combine flexibility to learn new information with sufficient stability to avoid catastrophic interference.

A large body of behavioural evidence has emerged showing that novel linguistic information is integrated into the lexicon over several days to weeks, often using artificial learning approaches (Dumay and Gaskell, 2007; Gaskell and Dumay, 2003; Merkx et al., 2011; Tamminen et al., 2012). For example, Gaskell and Dumay (2003) demonstrated evidence that a period of offline consolidation is required before newly learned words can engage in lexical competition with existing words. Words such as 'cathedral' can be identified quickly as once the utterance reaches 'cathed' there are no alternative candidate words. Gaskell and Dumay taught pseudowords that would influence this uniqueness point, for example, 'cathedruce', with the rationale that if the learned word had entered the lexicon it should slow down the recognition of 'cathedral'. While the novel words did not engage in lexical competition immediately after training, the effect was found upon testing 5 days later. Consistent with the CLS account of word learning, the emergence of these competition effects was distinct from recognition memory. Dumay and Gaskell (2007) extended these findings to directly test the role of sleep consolidation by training and testing at 12-hour intervals beginning at either 8 am or 8 pm. In this manner, the time between learning and testing was kept constant while only one of the groups had a period of sleep. Lexical competition effects appeared after 12 hours only for the group with sleep. Critically, however, the study included a further test after 24 hours. As predicted by the CLS account of word learning the competition effect appeared for both groups after they had a chance to sleep.

In addition to the behavioural evidence, some fMRI results have supported the claim that overnight consolidation supports changes in the neural representation of words. Davis et al. (2009) adopted an efficient scanning procedure which taught participants novel spoken words across two days. Following training on the second day a single scanning session compared neural responses to both the day 1 (consolidated) and day 2 (unconsolidated) spoken words. A further condition of untrained words that participants heard for the first time in the scanner was also included. FMRI responses in the STG revealed no differences between the unconsolidated and untrained words. Trained but consolidated words, by contrast, showed different activation to both unconsolidated and untrained words, instead showing activation more similar to existing words.

An important contrasting perspective on the role of sleep in learning novel words comes from a series of experiments from Kapnoula and colleagues. These studies showed lexical competition between recently learned spoken words could engage in lexical competition with existing words (Kapnoula and McMurray, 2016; Kapnoula et al., 2015). Using a visual world paradigm in which the recently learned words appeared alongside a real-word competitor and a real-word target revealed interference immediately after training and without the need for consolidation.

These findings provide evidence that overnight consolidation plays a pivotal role in learning novel spoken words. Moreover, the findings inform cognitive accounts of spoken word recognition. The natural focus of these experiments has been on the lexical level of learning – learning spoken words. Returning to the question of lexical or sublexical *orthographic* learning, we might expect that overnight consolidation would support the acquisition of whole written words. Nevertheless, the relative sparsity of research on the sublexical learning in spoken language does not preclude consolidation of sublexical written language.

2.4.3 Orthographic consolidation

In light of the strength of evidence that overnight consolidation supports the learning of novel spoken words, it is notable how little research has asked whether overnight consolidation might support learning of novel orthographic forms. As noted above, reading depends on a multiple cognitive skills working quickly in tandem to translate visual inputs to language. A novice reader must learn to identify visual characters, translate those characters to sounds, map those sounds to meaning, and scan across the text fluently. A core question of the thesis will be whether overnight consolidation can shed light on the how and when these cognitive skills come online. Identifying how memory systems impact on the early stages of reading acquisition may consequently inform cognitive accounts of reading, as well as improving the delivery of early reading instruction.

The triangle model will play a key role in framing any effects of overnight consolidation on learning to read. The majority of research into the lexical acquisition and sleep has focused on spoken word learning and has framed consolidation in light of the distributed cohort model, a connectionist framework for spoken word recognition (Gaskell and Marslen-Wilson, 1997). When thinking about the early stages of literacy acquisition the role of sleep consolidation is less clear. The triangle model, however, inherently captures the computational motivation for two-stage learning models; avoidance of catastrophic interference in a connectionist network. The networks of connections within each element of the triangle, as well as the connections between them, therefore provides a more relevant context in which to understand any overnight changes that might take place. Note that it is possible that localist models of reading could incorporate some aspect of consolidation but, as mentioned above, these models do not speak to learning. In addition, evidence for consolidation is not evidence for distributed processing. Although the triangle model offers a compatible framework in which to consider consolidation of written language, overnight changes in brain or behaviour would not be evidence for one model over another.

Wang et al. (2016) note how few studies have directly explored the impact of overnight consolidation on the process of learning to recognise written words. Of these, Bowers et al. (2005) showed apparent effects of overnight consolidation in the written domain by training participants to read and type new words. Following the approach used in the spoken word consolidation literature the novel words were similar to words with few neighbours, e.g., ‘banana’ and ‘banara’. A semantic categorisation task measured whether responses slowed to existing words, and thus whether lexicalisation of the novel word had taken place. Response times slowed to words that had acquired a new neighbour when compared to words that remained isolated, with the effect increasing on the second day of testing. In this case participants were learning a new word in the context of a familiar alphabet, and indeed with novel words that were similar to familiar orthographic forms. The study design included a second training session and so cannot distinguish training and consolidation effects. Addressing this issue, Wang et al. (2016) replicated this study with the inclusion of a between-group AM/PM manipulation. In this way competition effects for the recently learned words could be compared with and without sleep, holding time delay between training and testing constant. Here, the PM group showed competition effects 12 hours later (i.e., after sleep) while the AM group only showed this effect after 24 hours (also after sleep), thus supporting the role of sleep in orthographic learning. Bakker et al. (2014) meanwhile assessed whether learned spoken lexical items would prompt competition in written forms and vice versa, thus testing whether lexical consolidation would transfer across modalities. Training on both spoken and written word forms led to within-modality competition effects, i.e., written words engaged in competition in an orthographic task. Spoken words additionally led to competition in the written modality, a pattern that was not found for the reverse. Instead, learning written word forms showed competition in spoken modality only after one week, suggesting both the primacy of spoken language and that the processes of orthographic consolidation may follow a distinct pattern compared to spoken word learning. These data echo the importance of spoken language in both the DRC and triangle models of reading. Here, again, the findings were in the context of a familiar orthography. The experiments in this thesis go a step further by addressing the role of consolidation in learning a novel orthographic system.

Studies combining artificial orthography and consolidation. Of particular interest for the current thesis are instances where studies have taught artificial orthographies using designs that may encourage overnight consolidation. We know of no studies that have investigated the role of sleep in learning a novel orthography. Nevertheless, a number of studies running over longer timeframes have informative outcomes. For instance, Takashima et al. (2014b) investigated the formation of lexical and sublexical orthographic representations over the course of one month. Participants learned to read Greek graphemes mapped to Dutch phonemes. Critically, although participants could always decode the words letter-by-letter, the mappings from words to sounds included consistency at the level of syllables. Over time participants learned these intermediate representations and were faster to read familiar syllables even in the context of untrained words. Behavioural and fMRI measures at the end of the month suggest three levels of representations had been learned: letters, syllables, and words; participants could read novel words letter-by-letter, recombined words with familiar syllables, and trained whole words. The authors concluded that holistic word representations have emerged following a month of training. In a follow-up study Takashima et al. (2016) provided further evidence that intermediate representations may form a stepping-stone from letter-sound decoding to whole-word recognition. The study replicated the general pattern of findings from Takashima et al. (2014b) but focused more directly on promoting the emergence of holistic word representations through repeated presentations of specific letter sequences during training. Additionally, the study asked whether changes would occur to the lexical representations in the days following initial learning, with fMRI scans on days 1 and 5, as well as an additional testing session on day 30. Behavioural results showed clear effects of learning at both the whole-word and syllable levels. That is, participants were faster and more accurate to name trained words than novel words with familiar syllables which were, in turn, faster than completely novel combinations of letters. Brain imaging data from day 5 meanwhile showed differences in activation between trained and untrained items. However, as the study conflates training and day of learning it is not possible to distinguish whether changes are due to familiarity with the items or to sleep-related changes. In a similar study, Mei et al. (2013a) compared fMRI responses before and after 8 days of training. The design taught the same written stimuli in either alphabetic or logographic mappings, demonstrating reduced activation for logographic compared to alphabetic mappings in posterior fusiform cortex in the later scanning session. This outcome is consistent with the formation of holistic representations but cannot distinguish between the alternative explanation of training effects.

While none of these studies tells us about the role of sleep in the acquisition of a novel orthography they are less informative about how the cognitive processes that support literacy interact with memory systems. The approach adopted throughout this thesis instead takes a fine-grained approach to identify changes in processing that may occur due to overnight consolidation when reading a novel orthography. In this manner, we hope to identify the role of overnight consolidation in learning novel letters and words.

Consolidation leads to better generalisation. A key prediction of the CLS account of learning is that consolidation may allow linguistic information to be generalised outside of the context in which it was learned. When learning a novel orthography this may be extremely important. As discussed above one of strengths of an alphabetic writing system is the ability to generalise a small number of letters to read a large number of words (Tamminen et al., 2015). Will it be the case that a period of consolidation after learning new letters will improve participants' ability to read unfamiliar words? Tamminen et al. (2012) taught participants novel morphological affixes, e.g. *-nule*, in context of familiar words, e.g., *buildnule*. A spoken shadowing test immediately after training showed participants were faster to repeat trained affixes compared to untrained affixes. Critically, this was not the case when the novel affix appeared with an untrained stem, e.g., *sailnule*. When tested two days later participants were faster to repeat trained affixes regardless of whether the stem had been trained or not. Thus, a context-independent representation had emerged. In the early stages of reading it may be that letters are learned in the context of a frequent sight word, and that extracting and generalising letter-sound mappings requires a period of offline consolidation.

Consolidation of visual information. Changes in the visual representations of letters is another route by which overnight consolidation might play a role in learning to read. While a large body of evidence suggests that sleep consolidation is beneficial in the learning of novel spoken words and affixes (Davis et al., 2009; Dumay and Gaskell, 2007; Gaskell and Dumay, 2003; Merkx et al., 2011; Tamminen et al., 2012), there is a wider literature showing evidence for the involvement of sleep in the acquisition of a range of motor and perceptual skills (e.g., Gais et al., 2000; Karni and Sagi, 1993; Korman et al., 2007; Stickgold et al., 2000). Many of the overnight changes identified in the visual processing literature are potentially relevant for literacy acquisition. For example, sleep consolidation has been shown to improve performance in visual discrimination tasks (Karni and Sagi, 1993; Stickgold et al., 2000) with effects persisting even after controlling for changes in circadian rhythm (Gais et al., 2000). Moreover, disruption of sleep cycles impacts on sleep-related improvements in visual learning (Karni et al., 1994). These findings are consistent with dual-process theories of learning; an electrophysiological study of rats by Ji and Wilson (2007) showed coordinated interactions between hippocampus and visual cortex not only while learning to navigate a maze but also during subsequent sleep. Given that these findings suggest overnight changes in visual pattern discrimination and saccadic sequence learning we might expect improvements in the processing of letters rather than whole-words, as is the case in studies of spoken word consolidation. Overnight changes have also been demonstrated for learning saccadic eye movements (Meital et al., 2013). Similarly, McDevitt et al. (2014) showed that after a period of sleep consolidation participants were better able to segment and recognise complex visual objects, a similar process to identifying letters embedded within an unfamiliar word. Taken together these findings of overnight visual consolidation suggests exploring overnight changes in visual orthographic processing is a fruitful direction of research.

Consolidation may support statistical learning. Behavioural findings show that learners make use of statistical information across a range of modalities and learning situations (Conway and Christiansen, 2005; Fiser and Aslin, 2001; Maye et al., 2002; Saffran et al., 1996). This is particularly relevant in the context of the triangle model of reading which suggests statistical learning links orthographic, semantic and phonological knowledge. Meanwhile, the literature on statistical learning has largely focused on how distributional statistics allow learners to “discover the relevant units of perception” (Plaut and Vande Velde, 2017, p1). In reading the ‘relevant units of perception’ may depend on the transparency of the orthography, with learners becoming sensitive to the grain size (letter, syllable, or word) that most consistently maps between letters and sounds (Ziegler and Goswami, 2005). Expanding on the issue of grain size, statistical learning may help to explain the apparent tension between componential and holistic approaches to reading. Plaut and Vande Velde (2017) developed seven connectionist models that attempted to capture the tension between learning parts and wholes in human statistical learning tasks. For example, Simulation 1 modelled a study by Giroux and Rey (2009) where participants heard a stream of auditory ‘words’ that each contained either two or three syllables. The trisyllabic words had an embedded bisyllabic pair that was equivalent to the two-syllable words. Results of these simulations showed that participants can learn structure at multiple scales simultaneously, but that holistic learning inhibits discovery of the internal structure by the network (simulations 1 and 2).

Overnight consolidation may play a role in learning and abstracting statistical structure. For instance, Durrant et al. (2011) exposed participants to tonal sequences with an embedded statistical structure. The combination of two tones in a probabilistic sequence predicted the probability of the subsequent tone, thus all structure was second order in nature. Two experiments measured learning using a 2-alternative forced choice task (2-AFC) where participants chose which of two sequences sounded more familiar. In both experiments, participants in the sleep conditions were more likely to select the sequence that adhered to the statistical structure from training. A more intriguing finding from the perspective of learning to read is that the same paradigm transfers across modality. Using the same auditory tone sequences as from the previous study Durrant et al. (2016) showed that following sleep participants were able to generalise the auditory sequence to a visual context. This statistical perspective on consolidation is particularly relevant to the process of learning to read. The mappings between letters and sounds in a language are complex, with English writing containing a mix of consistent and inconsistent mappings; sleep consolidation might support the acquisition of these inconsistent mappings. A study by Warker and Dell (2006) on the role of statistical structure in acquiring implicit phonotactic constraints incidentally showed apparent consolidation-related learning. A follow-up study by Gaskell et al. (2014) directly tested the role of consolidation in this process by training participants to repeat aloud sequences of phonemes containing an artificial constraint. After training participants either napped for 90 minutes or stayed awake. Tests of learning and generalisation indicated that only the sleep group had acquired the novel phonotactic constraints. Children develop sensitivity to orthotactic constraints early in reading development (Cassar and Treiman, 1997), extrapolating the statistical cues of written language. Notably, no single written word contains enough information to learn these rules. Instead, the information must be extracted across multiple instances (Djonlagic et al., 2009). Taken together these perspectives on consolidation suggest it may play an important role in the early stages of learning to read.

2.5 Summary and thesis overview

The perspectives provided by computational models of reading and neuroscientific evidence converge on several key themes, in particular, the contrast between learning orthographic representations at the level of letters or whole words. These questions also appear as open questions in educational debates about reading instruction. However, research in these areas also highlighted a number of factors that influence literacy acquisition, yet are difficult to address in a naturalistic context: the nature of orthographic lexical representations; the role of frequency, semantics, and phonological knowledge on decoding; and how intermediate-level representations are learned. Studies using artificial orthographies to track the emergence of reading offer an important perspective on these questions.

In tandem, we have seen evidence that consolidation may support the acquisition of spoken or written whole-word representations, learning novel visual representations, and the abstraction of statistical information. These perspectives offer compelling reasons to explore the role of consolidation in the learning of parts and wholes in relation to learning to read. Each of the models of reading summarised above highlight the contrast between systematic letter-sound decoding and holistic word recognition. By combining experimental designs that utilise sleep-consolidation as an experimental variable this thesis sets out to explore the early stages of learning to read a novel orthography. From the literature reviewed above a number of specific questions emerge:

1. To what extent does learning a novel orthography draw upon the same neural systems as reading real words and pseudowords, and to what extent might this change with consolidation?

2. Will whole-word orthographic representations emerge in the context of learning to read an artificial orthography?

3. How will phonological, lexical, and semantic information impact on the ability to use and generalise an artificial orthography?

4. Will sleep consolidation impact on recently learned orthographic knowledge, and if so, will consolidation occur at a lexical or sub-lexical level?

Thesis overview. This thesis addresses the issues highlighted above through six experiments combining behavioural and fMRI measures. In Chapter 2, Experiment 1 compared holistic and componential processing of recently learned words and objects in the context of an overnight consolidation paradigm. This study validated the artificial orthography paradigm by showing that recently learned orthography activated similar neural regions to reading pseudowords written in familiar orthography. Furthermore, the convergence of results between (1) activation when reading and naming the artificial words and objects; (2) activation when reading and naming familiar words, objects, and pseudowords in a functional localiser; and (3) the regions highlighted for holistic and componential processing in a meta-analysis of fMRI studies of reading strongly suggests that further study of holistic and componential orthographic learning is worthwhile. In Chapter 3 Experiments 2, 3, and 4 investigated the role of overnight consolidation on the formation of representations of novel letters and orthographic whole-word forms. All three experiments show overnight improvements in the ability to use and generalise knowledge of letters, but not words. Experiments 3 and 4 assessed whether overnight consolidation played a role in the formation of bigram representations, while additionally providing evidence that participants have consolidated some representation of the trained words on day 1 by means of a recognition memory task. Experiments 5 and 6 explored the role of spoken word knowledge in reading and recognising words in an artificial orthography while simultaneously attempting to replicate previous findings of overnight consolidation of novel spoken words. These studies failed to replicate previous findings of spoken word consolidation. Possible reasons for the lack of replication are discussed.

Chapter 3

Neural basis of holistic and componential visual-verbal mappings

- This chapter combines behavioural measures and fMRI to compare learning and generalisation of novel visual-verbal associations in the form of artificial objects and words.
- FMRI results showed increased activation in posterior ventral occipitotemporal (vOT), parietal, and frontal cortices when reading an artificial orthography compared to naming artificial objects, and the reverse profile in anterior vOT regions.
- The comparison between neural activity for artificial words and objects showed extensive overlap with systems differentially engaged for real object naming and English word/pseudoword reading in the same participants.

The results described in this chapter have been published as:

Quinn, C., Taylor, J. S. H., & Davis, M. H. (2017). Learning and retrieving holistic and componential visual-verbal associations in reading and object naming. *Neuropsychologia*, 98, 68-84. https://doi.org/10.1016/j.neuropsychologia.2016.09.025

3.1 Introduction

In recent years the education literature has settled upon phonics as the only evidence-based method of teaching reading (Torgerson et al., 2006; Wyse and Goswami, 2008). Indeed, in the UK, the Rose Review (Rose, 2006) recommended that synthetic phonics, which involves explicit instruction in letter-sound decoding and blending, should underlie early reading instruction. This approach provides children with the primary skill of being able to translate print to sound. Whole-word methods of reading instruction instead argue for the primacy of meaning in reading, with knowledge of letter-to-sound mappings being acquired through exposure to meaningful text. In this case the primary skill of reading should not be translating print to sound, but instead print to meaning. Correspondingly the focus of early learning in whole-word reading schemes is to recognise whole 'sight words', rather than decoding the letter-sound correspondences within each word. Thus, many are sceptical of whether, in line with the Rose Review, phonics should be taught "first and fast". While experimental data has an important role to play in this activity, Wyse and Goswami (2008) note that very few naturalistic studies comparing different methods of reading instruction meet rigorous experimental standards. In this paper we consider whether laboratory studies of holistic and componential visual-to-verbal learning may offer a way to address educational questions in a controlled manner.

The distinction between recognising whole-words and decoding letter-by-letter in the educational literature is mirrored to a large extent by findings from cognitive research on reading. Cognitive models of reading, such as the Dual Route Cascaded (DRC) and triangle model, reflect the distinction between holistic and componential processing by suggesting that the meaning of a written word can be accessed in more than one manner (Coltheart et al., 2001; Plaut et al., 1996). For example, in the DRC model, words can be read componentially by decoding letter-by-letter (sub-lexical route), or can be mapped onto their pronunciations and meanings directly by recognising the whole word form (lexical route). It is the componential relationship between visual and phonological forms in alphabetic languages that means we can read pseudowords, e.g. 'spape', using our knowledge of letter-sound mappings. In contrast, to read an irregular word (e.g. 'pint') we must have whole-word knowledge to know that it does not sound similar to words that share the same orthographic components ('mint', 'hint', etc.). In the triangle model (Plaut et al., 1996) the mappings between written (orthographic) and spoken (phonological) forms are componential; this model does not contain whole-word, or lexical, representations of this information. However, in this model, the relationship between a familiar word's written form and its meaning is more holistic and item-specific, since the form-to-meaning mapping cannot be broken down into sub-components, at least for monomorphemic words (i.e. most monosyllabic words).

Furthermore, this item-specific knowledge is proposed to be important for irregular word reading, helping them to be pronounced differently from similarly spelled regular words. Thus, both the DRC and the triangle model propose that reading involves both componential and whole-word knowledge, with the former being more important for pseudowords or less familiar words, and the latter more important for words, in particular those with irregular spellings.

Although both componential (sub-lexical) and holistic (lexical) processes are involved in skilled reading it is not clear how the relative importance of these skills changes as we learn to read. The goal of the present study was to advance our understanding of the initial stages of reading acquisition by exploring the neural basis of componential and holistic processing. To do so we compared learning to read an artificial alphabetic orthography with systematic symbol-to-sound mappings with learning names for novel objects with arbitrarily associated names.

3.1.1 Neural bases of holistic and componential processes in reading

To briefly reiterate the literature review, ventral occipito-temporal (vOT) cortex processes both sublexical and lexical orthographic representations in a hierarchical manner (Cohen et al., 2000, 2002; Dehaene et al., 2002; Mechelli et al., 2005; Taylor et al., 2013; Vinckier et al., 2007). This lexical/sublexical distinction largely reflects the contrasts found in the educational literature; componential reading (e.g., decoding pseudowords) shows more activation in posterior vOT, while holistic recognition of familiar words shows more activation in anterior vOT regions (Mechelli et al., 2005). As with the educational literature however there may be a difference between established reading and the process of learning to read. Thus, the focus on vOT contributions to established reading processes may have shifted attention away from other brain regions that may play a more active role during reading acquisition; e.g., children learning to map symbols and sounds show increased parietal activation (Bitan et al., 2007a, 2006; Cao et al., 2015; Church et al., 2008).

Artificial reading studies have offered valuable perspective on these questions. For example Mei et al. (2013b) show changes in vOT activation for a novel orthography over the course of 8 days while Takashima et al. (2014b) show changes in parietal activation for holistic word representations over the course of one month of training. Where these studies showed the effect of learning over time, Taylor et al. (2014a) explored the process of initially acquiring letter-sound-mappings by comparing behavioural and neural responses while participants learned to name novel objects and to read novel words written in a novel orthography. Although both learning to name objects and learning to read words involve associating visual and verbal representations of items, the underlying processes differ in ways that inform our understanding of the neural basis of reading. The relationship between letters and sounds in an orthographic writing system is systematic and componential; readers can decode unfamiliar words, e.g. 'spape', using their knowledge of letter-sound-mappings. By contrast, it is not possible to deduce the name of an unfamiliar object by examining its visual appearance. When we refer to object recognition as 'holistic' therefore we are addressing the fact that the whole object must be identified, and that an unfamiliar object cannot be named by recognising its constituent parts.

The Taylor et al. (2014a) study used a novel imaging strategy to isolate the process of retrieving holistic and componential phonological information. During scanning two trial types appeared, a see-think trial and a see-speak trial. During see-think trials participants were instructed to covertly recall the phonological form but not to articulate it aloud. During the see-speak trials participants named the item aloud. Critically, if an item, e.g. /bup/ appeared in a see-speak trial, the same item would have appeared in the see-think condition on the previous trial. There were several reasons for this interaction analysis. First having just recalled the item's name allowed participants to name the item aloud in the gap between scans. More importantly, this approach allowed the separation of the neural activation associated with the retrieval of phonology, from that associated with the visual differences of the stimuli, and the processes of articulation (discussed in more detail in the methods section) by examining the interaction between the two conditions. The consequence of this approach is that this experimental paradigm allows a close examination of the process of retrieval of componential and holistic visual-verbal mappings.

Imaging results from learning runs show activation in bilateral parietal cortices when learning the componential letter-sound mappings in reading while mid- to anterior-fusiform regions support learning of whole items for object naming. More posterior fusiform regions showed no difference between words and objects during learning. Additionally, activation in a left mid-fusiform ROI was predictive of successful object learning, but not word learning. During the testing run in contrast clusters were only found for words more than objects, in bilateral inferior and superior parietal cortices and left superior frontal gyrus. It is important to note however that since the majority of scanning involved learning this testing run may be relatively underpowered in comparison to the training runs. These findings offer an interesting perspective on results previously seen in vOT regions by comparison of completely novel orthography and objects during initial learning. The use of the interaction analysis to isolate the processes associated with retrieval of phonology while minimising the effect of visual differences between stimuli. The results suggest rather that where in the ventral visual stream we see activation for an item depends to a large extent on whether that item is being processed in a componential manner, looking at the elements within it, or in a holistic manner by taking in the item as a whole.

The componential/holistic distinction identified by Taylor et al. is compatible with cognitive models of reading. The dual-route cascade model of reading, for example, suggests words can be reading either lexically, by recognising the whole-word, or sublexically, by decoding letter-by-letter. Indeed, a meta-analysis of word reading imaging studies (Taylor et al., 2013), suggests both of these processes take place. One of the key contrasts here was between words and pseudowords, where the key difference is whether a lexical (holistic) representation exists. In particular the parietal regions identified by Taylor et al. (2014a) as being involved in learning to read words are similar to those regions identified in the meta-analysis as showing common activation for both real words and pseudowords. The comparison of words and pseudowords (and thus holistic versus componential representations) by contrast revealed greater activation for words compared to pseudowords in anterior regions of the fusiform gyrus, while pseudowords compared to words showed greater activation in more posterior regions.

The outcome in Taylor et al. (2014a) that there is more activation in mid- and anterior fusiform areas for novel objects than novel orthography contrasts with imaging studies of real reading (Dehaene and Cohen, 2007, 2011; Szwed et al., 2011). This finding may therefore be due to the artificial orthography being processed differently to real reading, with no representations for the novel orthographic units. Alternatively it may be because the novel words being associated with the orthography do not yet have a lexical (holistic) representation and so do not engage the typical mid-fusiform areas. One approach to tracking the emergence of orthographic- or lexical-specific activation may come from research into the role of sleep consolidation in learning.

3.2 Experiment 1

Here we set out to track the changes in neural activity that occur over the first two days of learning to read and name artificial words and objects. In doing so we build on previous work on the neural systems previously identified in learning holistic and componential visual-verbal associations. We extend this earlier work by asking: (1) do we see a similar neural dissociation during retrieval of holistic and componential visual-verbal associations. Furthermore, we ask whether neural activity associated with retrieval of written words offers evidence of the emergence of whole-word representations, (2) in the context of overnight consolidation and (3) by comparison with untrained words. Finally, we assess: (4) the extent to which neural activity associated with reading and naming artificial words and objects overlaps with regions involved in holistic and componential processing of real words, objects, and pseudowords. As we will explain below, these four elements substantially extend the findings of Taylor et al. (2014a) while using similar methods.

In the current study, we trained participants outside of the scanner to read artificial words written in an artificial orthography, and to name artificial objects. Critically, the artificial written words had a componential and systematic mapping between the visual (letters) and verbal (sounds) forms, whereas artificial objects had a holistic and arbitrary relationship between the visual and verbal forms. That is to say, if participants successfully learn the componential letter-sound mappings of the written words then they should be able to generalise this knowledge in order to read unfamiliar words. By contrast because the relationship between the visual and verbal forms of an object is arbitrary and holistic; it is not possible to name an unfamiliar object.

Whilst Taylor et al. (2014a) focused on measuring neural activity during the learning of holistic and componential visual-verbal mappings by training participants during scanning, here we focus on activation during retrieval of knowledge previously acquired outside of the scanner. This design allowed us to dedicate more time to testing neural activity associated with reading or naming of trained items during scanning. We were also able to adopt an event-related fMRI design in which written words and objects were intermixed, in contrast to the blocked design used by Taylor et al. (2014a). This may be more sensitive to detecting activation differences between words and objects (Josephs and Henson, 1999) and increases the likelihood that any activation differences between reading written words and naming objects are due to the immediate demands of processing of componential and holistic visual-verbal associations, and not due to longer-term differences in the strategy adopted in different testing blocks.

In order to investigate the time-course over which holistic representations of written words may emerge we adapt the train twice, test once design used by Davis et al. (2009) to explore the neural effect of overnight consolidation for spoken words. Following the method used by Davis et al., half of the words and objects in the current study were learned on day 1 and the other half on day 2, with scanning taking place following day 2 training. This design allows an opportunity for overnight consolidation of day 1 but not day 2 items. In line with complementary learning systems accounts (Davis and Gaskell, 2009) we might anticipate differences between neural responses for items scanned following a night of sleep as compared to items learned on the same day as scanning. Using this design does not allow us to distinguish whether any consolidation effects come about due to the processes of sleep or due to time elapsed since learning. Nevertheless the design allows us to compare initial and longer-term changes during the earliest stages of learning to read. It will be for future research to determine the underlying cause(s) of any changes observed.

In addition to reading or naming all the trained words and objects from days 1 and 2, participants also read three sets of untrained artificial words during scanning. These conditions therefore permitted a comparison of trained and untrained words – allowing us to assess activation differences that might parallel those seen between English words and pseudowords. Two sets of these untrained words were the written forms of the object names learned on days 1 and 2. The spoken forms of these items were therefore familiar (trained) while the orthographic forms remained unfamiliar (untrained) (Figure 3.1, p56). This manipulation allowed us to determine if familiarity with the phonological form of a word (in the context of object naming) results in differential activation as compared to the third set of untrained words which were completely unfamiliar. For these comparisons, we can therefore assess the possibility (raised by research on spoken word learning, Davis and Gaskell, 2009) that holistic lexical representations are enhanced during overnight consolidation and hence that these effects may differ for words learned on day 1 or day 2. These comparisons would be very difficult to achieve in a naturalistic setting.

In order to validate the use of this laboratory-learning approach in relation to the brain systems ordinarily engaged for word reading and object naming we also included a functional localiser scan in which participants read English words and pseudowords and named familiar objects. This allowed us to ask whether the neural systems that support reading of artificial words and naming of artificial objects are the same as those used for real word reading and object naming. This may help us in ascertaining whether a neural distinction between holistic and componential processing applies equally in artificial and real reading and object naming.

In summary then we address four major questions in this study: 1) Are the same neural systems involved in the retrieval of holistic and componential visual-verbal associations as previously shown to support learning? 2) How does overnight consolidation impact on neural representations of recently learned words and objects? 3) What do comparisons of reading trained and untrained words suggest about neural systems for whole word representation and generalisation to pseudowords? 4) To what extent do the neural systems involved in reading and naming artificial words and objects overlap with the systems involved for real words and objects?

3.3 Methods

3.3.1 Participants

25 right-handed native English speaking adults aged 18–40 took part in a study approved by the Cambridge Psychology Research Ethics Committee. No participants reported having dyslexia, speech, or language impairment or any pre-existing neurological condition that would preclude participation in functional MRI. As critical comparisons involved novel words and objects learned prior to scanning, five participants were excluded due to poor performance reading/naming these items during scanning (< 20% correct). An additional participant did not complete the functional localiser run and so the localiser analysis reports the results from 19 participants.

3.3.2 Materials

Artificial stimuli

Three sets of 36 monosyllabic consonant-vowel-consonant (CVC) pseudowords were used in the experiment ("pæg", "zɒn", etc.). Each pseudoword set was assigned to one of three conditions in a counterbalanced manner over participants: objects, words, and an additional set of untrained words (Table 3.1, p53). These pseudowords were constructed from the same set of 12 consonants (b, d, f, g, k, m, n, p, s, t, v, z) and four vowels (æ, ɛ, ɒ, ʌ) as used in Taylor et al. (2014a). Segment position was matched across stimulus groups with each consonant appearing three times in onset position and three times in coda while each vowel appeared 9 times in each set. Each of the three stimuli sets were further split into two groups of 18 items to be trained on days 1 and 2. Unfamiliar visual symbols (artificial letters) were mapped to the 16 phonemes in a one-to-one manner meaning that the written forms of items had consistent letter-sound mappings (See Figure 3.1, panel D, p56 for examples). Spoken forms of pseudowords were recorded by a female native English speaker in a soundproof booth and digitised at a sampling rate of 44.1kHz. In order to allow comparison of recognition memory in a later task we constructed a further set of 28 untrained foil items that followed the same CVC structure and contained the same letters as the trained items but without the constraint of matching letter frequency in onset and coda positions.

Before going further a note on terminology is important. One of our main questions involves the process by which the spoken and visual representations of words become lexicalised. As a result calling items pseudowords is potentially ambiguous. The spoken forms of all of our items start as pseudowords but some of them are trained to explore whether they will become more like words than pseudowords. To avoid the confusion that naturally

Table 3.1 Artificial stimuli sets (A, B, and C) were assigned in a counterbalanced manner to trained words, trained objects, or untrained words. Trained items were further split into counterbalanced sets for days 1 and 2.

Set A		Set B		Set C	
Part 1	Part 2	Part 1	Part 2	Part 1	Part 2
mAf	dAs	bAv	fAm	vAd	vAk
bAb	fAk	pAz	vAg	dAt	pAf
pAg	zAm	nAm	fAp	gAk	zAt
bAz	pAb	zAd	sAb	tAm	kAs
fAv	zEt	tAf	bEz	nEg	pAv
dEm	gEn	sEg	kEt	bEm	kEb
kEs	dEz	bEp	nEf	tEz	sEp
zEk	vEg	vEd	dEv	fEg	bEv
tEp	gEb	sEk	tEk	mEk	gEd
fQm	mQt	gQp	dQv	pQn	gQm
nQs	sQf	zQn	pQs	kQv	fQd
vQg	bQv	fQt	tQn	zQp	dQz
sQn	pQf	mQn	zQs	nQf	vQn
tQz	kVv	pVm	mQd	dQb	tVs
sVt	tVd	dVs	gVk	mVs	nVg
nVd	vVp	vVt	mVb	fVb	mVn
mVp	nVk	gVb	kVz	sVf	zVt
gVd	kVn	kVg	nVf	bVp	sVz

arises from this we refer to the trained words written in a novel orthography 'artificial words'. Trained objects are 'artificial objects' and untrained novel words that participants read for the first time during testing are referred to as untrained words. Instead we refer to pseudowords only in the localiser task below where we compare reading of familiar English words to reading of pseudowords written in a familiar orthography.

Stimuli for the functional localiser were 120 monosyllabic items were chosen from the updated Snodgrass and Vanderwart item set (Magnié et al., 2003) and were randomly assigned on participant-by-participant basis to appear either as object pictures (60) or written as words (60). This prevented potential priming effects that would occur if all items appeared as both words and objects for each participant. 98 of the 120 items (81.6%) had grapheme-to-phoneme correspondences that are classified as regular according to the DRC model of reading (Rastle and Coltheart, 1999, Table 3.2, p55), and the mean log frequency of the items based on the Zipf scale (Van Heuven et al., 2014) was 4.47 ($SD = 0.56$), i.e. relatively high frequency. 120 monosyllabic pseudowords were generated from the ARC nonword database (Rastle et al., 2002) and were pairwise matched to these Snodgrass and Vanderwart items for letter length and orthographic neighbourhood size. (see Table 3.2, p55; Words: $mean\ length\ (SD) = 4.09(0.82)$, $orthN\ mean\ (SD) = 8.30(5.14)$; Pseudowords: $mean\ length\ (SD) = 4.20(0.75)$, $orthN\ mean\ (SD) = 8.27(4.39)$). Snodgrass and Vanderwart items and pseudowords were further matched pairwise for initial phoneme as this factor has been reported to have the most impact on reading and naming latencies (Rastle et al., 2005). 60 of these pseudowords were randomly selected for each participant. Localiser items (words, pseudowords, and objects) were therefore matched at a group level but not for each individual participant. This decision prevents potential priming effects that might occur if the same item appeared as both word and object for an individual.

3.3.3 Procedure

The experiment used a 'train twice, test once' design in which behavioural and neural responses to items learned on day 1 and day 2 (hereafter day 1 and day 2 items) can be compared in a single scanning session performed on day 2 (Figure 3.1, panel A, p56). Participants learned items over two days and then completed a combined fMRI and behavioural testing session following training on the second day. As a consequence of testing only once this efficient scanning design removes effects of practice on neural responses (e.g., for a longitudinal design) and avoids the neural variability that would be caused by testing on two occasions or scanning two different groups of participants (Davis and Gaskell, 2009). Participants' responses were recorded and scored offline.

Table 3.2 Experiment 1: Word and pseudoword stimuli

Real words			Pseudowords		
ant	duck	peach	abe	dutt	paunt
arm	ear	pear	arn	eash	pim
axe	egg	pen	ath	eck	plean
ball	eye	pig	bame	elt	pouse
bath	fence	pipe	barse	farb	proo
bear	fish	plug	bave	feach	pudge
bed	flag	pot	beb	fedge	purf
bee	flute	ring	berve	fet	rass
bell	fly	rope	bim	fick	remp
belt	foot	saw	bive	fiss	sadge
bird	fork	screw	bleer	flum	scark
book	fox	seal	bolve	fuke	shig
boot	frog	sheep	bouth	fune	shub
bow	glass	shirt	breb	gamb	sile
bowl	glove	shoe	brime	garge	skine
box	goat	skirt	brip	geg	snay
bread	grapes	skunk	browl	gope	snud
broom	gun	sled	bue	gow	sog
bus	hair	snail	burl	hame	solk
cake	hand	snake	cang	heek	souch
cap	harp	sock	chag	helt	spab
car	hat	spoon	ched	hong	spale
cat	heart	star	chemp	hoose	spape
chain	hook	stool	chone	hosh	spen
chair	horse	stove	clag	hount	stonk
church	house	sun	clope	hudge	strang
clock	key	swan	cly	kib	strine
cloud	kite	swing	cose	kinge	swart
clown	knife	tail	crad	kouse	tark
coat	lamp	thumb	crame	lerge	teb
comb	leaf	tie	crench	lomp	thep
couch	leg	toe	crub	loof	tilm
cow	lips	top	culk	losh	torb
crown	lock	train	cupe	lun	touth
cup	map	tree	cusk	modge	turb
deer	moon	tree	dack	moise	tute
desk	mouse	vase	dag	mooth	vone
dog	nail	watch	dalve	netch	womp
doll	nose	wheel	dask	nink	wug
door	nut		detch	nolt	
dress	owl		dod	orn	

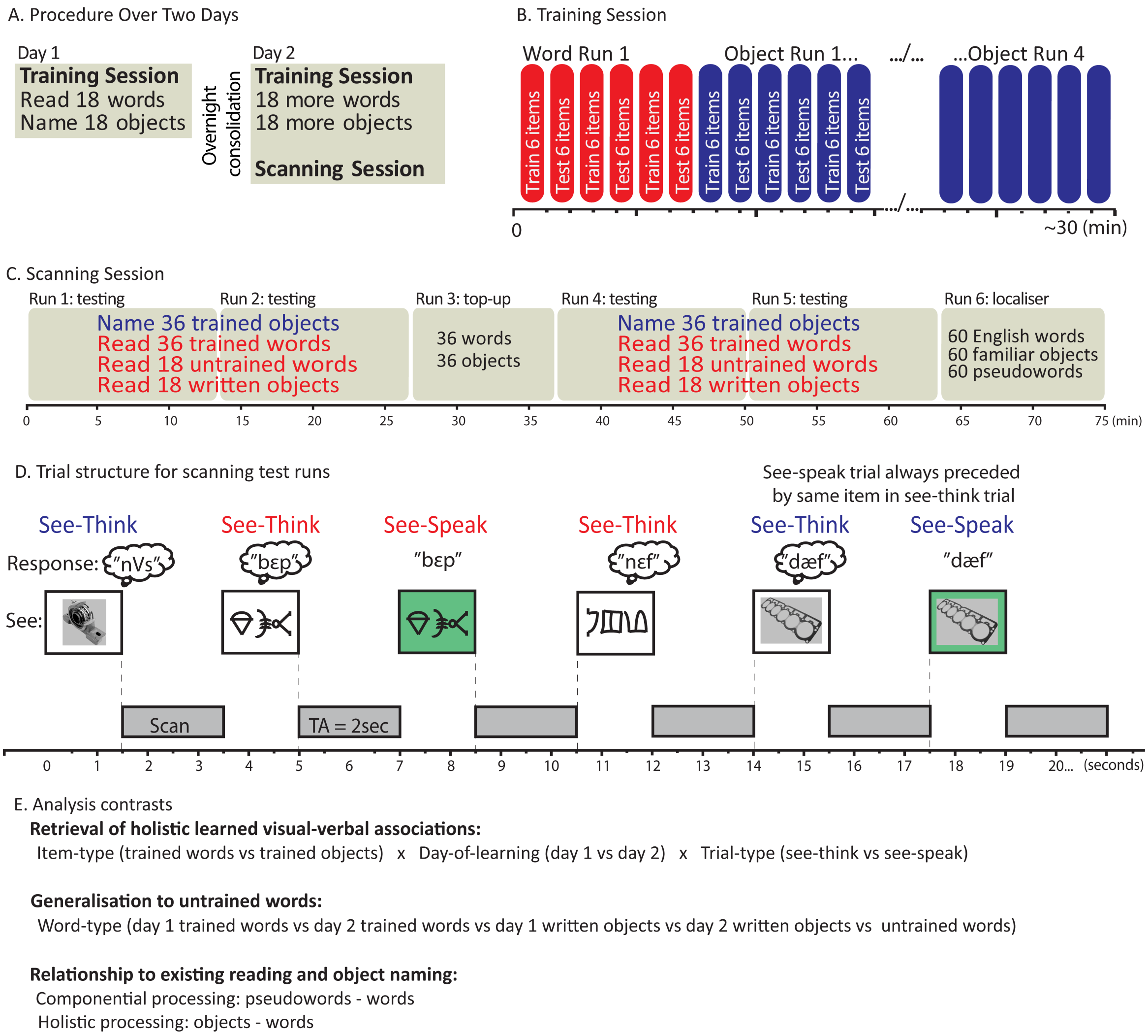


Fig. 3.1 (A) Timing of training and testing procedures on the two days of the study. (B) Train/test structure used during the training sessions with alternating periods spent learning and being tested on words and objects. Within each run participants were trained then tested on 6 items at a time until all 18 items had been tested. (C) Timeline of fMRI scanning runs including testing runs, top-up and localiser runs. In testing runs 1 and 2 participants were presented with each of the trained items twice in a see-think trial and once in a see-speak trial. Half of the untrained items were similarly presented in these two testing runs. Testing runs 4 and 5 followed the same structure with all trained items presented again, and the other half of the untrained items. (D) Time line showing the structure of the two trial types presented during the scanning runs: see-think and see-speak trials. Word and object trials were presented in a random order for each participant with the constraint that see-speak trials always followed a see-think trial for the same item. However, see-think trials were also presented in isolation so that activity for these two trial types can be separated at the analysis stage. See-speak trials were identical to the see-think trials except that the green background cued participants to say the appropriate word/object name aloud rather than covertly. (E) Goal of each analysis and the comparisons used in each case.

Training (prior to scanning)

Training took place over two days with 36 spoken pseudowords being associated with 18 artificial written forms and 18 artificial object pictures each day (Figure 3.1, panel A, p56). Participants completed 8 runs of training each day (four each of word and object training) consisting of alternating word and object runs (Figure 3.1, panel B, p56). Within each run 18 items were presented across 6 blocks that alternated between training and testing. During trials in the training block participants passively viewed the visual forms of each item onscreen for 3500*ms* and then heard the spoken phonological form 500*ms* after the visual onset. Six items therefore appeared in each training block. During testing blocks, the same six items appeared in a different order. The visual forms were presented and participants read/named the item aloud during the 3500*ms* in which the item was onscreen. Responses were recorded and scored offline and were deemed correct if participants correctly reported all three phonemes in the target item. At the end of the training on day 2, participants completed a short practice session of the task used in the scanner with real words and objects.

MR Data Acquisition

Functional magnetic resonance imaging data were acquired using a 3T Siemens Trio scanner (Siemens Medical Systems, Erlangen, Germany) with a 32 channel head coil and Sensimetric headphones. Responses were recorded with a dual-channel MRI microphone (FOMRI II, Optoacoustics). Audio stimuli were processes using the Sensimetric EQ Filtering 2.1 software for presentation over Sensimetric S14 headphones in the scanner. Visual stimuli were presented using a monitor mounted at the rear of the scanner bore, viewed via an angled mirror attached to the head coil.

We used a rapid sparse imaging event-related design with a repetition time ($TR = 3500ms$) longer than the acquisition time ($TA = 2000ms$), which allowed a gap of 1500*ms* during which spoken responses could be recorded in the absence of scanner noise. This silent period between scans meant that participants could hear their own voice when speaking, and additionally reduced the impact of motion-induced artefacts on the acquired images (Peelle, 2014; Perrachione and Ghosh, 2013). Each of the four word/object test runs involved acquisition of 162 images. Image acquisition consisted of 32 transverse oblique axial slices, angled to avoid the eyes. Each slice was 3mm thick and consisted of a 64*x*64 matrix of 3*x*3*mm* voxels. There was a .75*mm* gap between adjacent slices such that the total image volume allowed for whole-brain coverage, except for a few cases in which the very top of the parietal lobe was not covered. To assist in anatomical normalisation, we also acquired a T_1-weighted structural volume using a magnetisation prepared rapid acquisition gradientecho protocol (repetition time = 2250*msec*, echo time = 2.99*msec*, flip angle = 9°, 1*mm* slice thickness, 256*x*240*x*192 matrix, resolution – 1*mm* isotropic).

Scanning Procedure

The scanning session consisted of 6 scanning runs lasting 72 minutes in total (Figure 3.1, panel C, p56). Four of these runs tested artificial word reading and object naming, the main focus of the study. After two of these runs, participants additionally completed a 'top-up' run where they were presented with both the visual and phonological forms of items in the same manner as the training session (i.e. paired presentation of visual and verbal forms of each item). This allowed another opportunity to learn the mapping between the visual and phonological forms of items, thus maximising the trials with correct behavioural responses. A functional localiser run that involved reading English words and pseudowords and naming familiar objects completed the scanning session. At the start of scanning a high-definition MP-Rage structural scan was acquired.

Participants completed four reading/naming runs of 11 minutes duration while in the scanner (Figure 3.1, panel C, p56). Each run contained 9 testing blocks of 63 seconds each with a rest period of 10.5 seconds between blocks. 18 trials appeared in each block of testing, made up of 12 see-think trials and 6 see-speak trials. All 36 trained words and 36 trained objects, along with half of the untrained words $(N = 18)$ and half of the written forms of objects $(N = 18)$ were split to appear across runs 1 and 2. The remaining half of the untrained words and written objects appeared in runs 3 and 4, along with a second presentation of all 36 trained words and 36 objects. Consequently untrained items were not repeated and so would remain novel, while participants had two opportunities to name each of the trained items. This design meant that each trained item (36 words and 36 objects) was associated with 6 scans: 2 presentations each with 2 see-think trials and 1 see-speak trial, 432 scans in total for the trained items. The untrained items (36 untrained words and 36 written objects) meanwhile were associated with 216 scans.

Critical to our design was that half of the see-think trials were followed by a see-speak trial in which the same item was presented (Figure 3.1, panel D, p56). During see-think trials the items appeared on a white screen and participants were instructed to recall but not articulate the spoken form. For the see-speak trials items appeared on a green screen and participants were instructed to say the phonological form of the item aloud. Each trial lasted 3500*ms* starting with a visual item presented for the first 1500 ms followed by a single functional brain volume being acquired in the remaining 2000*ms*. Including see-think and see-speak trials was important to the design for several reasons. First, the time between scans was not long enough for participants to read a novel word and say it aloud. As the see-speak trials always followed immediately after a see-think trial participants had already retrieved the item pronunciation on the previous trial and could articulate its spoken form in the short period between scans. Second, this design ensured that the majority of trials were not affected by head movements due to articulation (as there were double the number of see-think trials as see-speak trials), and prevented anticipation of articulation on the subsequent trial, since participants could not predict whether a see-think trial would be followed by a see-speak trial. Third, as articulation only took place on see-speak trials, subtraction of see-speak from see-think trials will remove activation associated with articulation, and reveal activation associated with covert phonological retrieval on see-think trials. Finally, it is possible that activation differences between words and objects may in part be driven by visual differences between these two types of stimuli. As the same visual form was presented on successive see-think and see-speak trials, subtraction of see-speak from see-think trials may also reduce the impact of these visual differences. We return to this issue in the discussion.

During the localiser task participants were presented with 60 items in each of three conditions: written English words, pseudowords, and real objects. The same event-related sparse imaging design was used ($TR = 3500ms$, $TA = 2000ms$). Items were randomised and appeared onscreen for the 1.5 s of silence between scans using the same block structure as above; a 63 s block containing 18 trials followed by 10.5 s rest. 186 EPI images were acquired (13 min scanning time).

3.3.4 Imaging analysis

MRI Preprocessing

Image processing and analysis of all EPI data were performed using SPM8 (Wellcome Trust Centre for Functional Neuroimaging, London, UK) in conjunction with AA software version 4 (Cusack et al., 2014). The first six volumes of each scanning run were discarded to allow for equilibration effects. Images for all scanning runs for each participant were realigned to the first image in the first scanning run (Friston et al., 1995) and the resulting mean image was co-registered to the T_1 structural image. Normalisation of structural images to standard MNI space was calculated using tissue probability maps (Ashburner and Friston, 2005), and these warping parameters were the applied to all functional images for that participant. Normaliser functional images were resampled to 2mm isotropic voxels and spatial smoothing was applied using a kernel full-width-half-maximum of 8mm. For the functional imaging analyses described below we used an event-related analysis implemented in SPM8 software. Accordingly, event times were convolved with the SPM8 canonical hemodynamical response function following the recommendations of Perrachione and Ghosh (2013). Movement parameters estimated at the realignment stage of preprocessing were added as regressors of no interest. All analyses used a voxelwise threshold of $p < 0.001$ combined with a cluster extent-based FWE-corrected threshold of $p < 0.05$ unless otherwise stated.

Functional localiser analysis

We used an overt naming task in order to overcome the challenge posed by Price and Devlin (2011a) that passive viewing of words may induce greater covert naming than passive viewing of objects. To reveal the neural systems involved in holistic as opposed to componential visual-verbal processing of real stimuli, we used the contrast $[objects - words]$. To ensure that this comparison revealed engagement of different representations, as opposed to differences in processing effort, this contrast was conducted after taking response time (RT) differences into account, using the approach proposed by Taylor et al. (2014b). See Figure 3.2 (p61) for response time differences to words, objects, and pseudowords. This approach involves building a regression model that includes one parametric modulator to model the effect of RT on BOLD signal (irrespective of condition), and additional parametric modulator(s) representing the different stimulus conditions. Activation associated with this second parametric modulator then reflects the differences in neural response between conditions over and above effects due to response time differences (See Taylor et al., 2014b, for full explanation).

The contrast $[pseudowords - words]$ was included to reveal the neural systems involved in componential as opposed to holistic mappings between visual and verbal representations. Unlike the contrast between words and objects, we do not partial out the effects of response time when comparing words and pseudowords. Following the framework of Taylor et al. (2014b) both words and pseudowords should engage the same neural systems, but pseudowords take longer to read aloud, and should therefore drive greater activity due to greater processing effort (c.f. Taylor et al., 2013, 2014b). Differences in response time between words and pseudowords should reflect the differences in effort. As this contrast was intended to tap processing effort, response time differences between words and pseudowords were not taken into account when conducting this contrast.

Reading familiar words is an automatic process that is relatively effortless compared to both naming objects and reading pseudowords, leading to much shorter response times for the former than either of the two latter conditions. Given this, it is perhaps not surprising that contrasts of both $[words - objects]$ and $[words - pseudowords]$ showed activation throughout the default mode network (Gusnard and Raichle, 2001; Raichle, 2015). We therefore chose not to use these contrasts as these regions are unlikely to contribute directly to word reading (see Table 3.4, p68 for details of these contrasts).

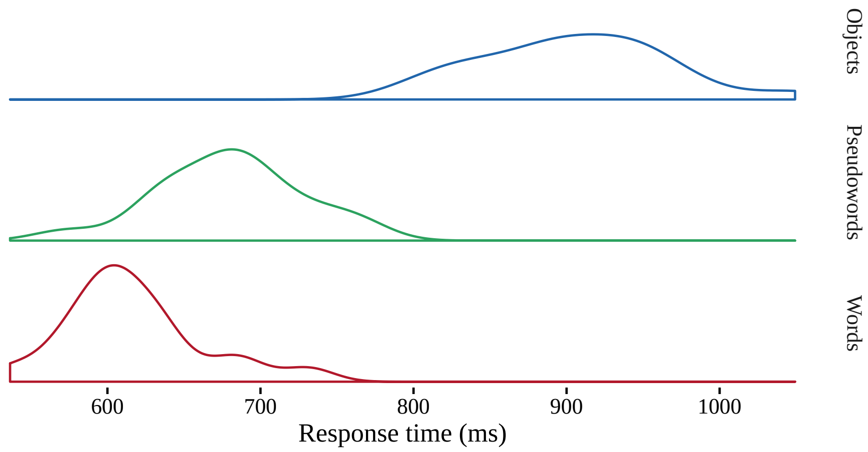


Fig. 3.2 Density plot showing response times to objects, pseudowords, and words during the localiser task. Incorporating a parametric modulator of RT across both words and objects allowed comparison of the differences between words and objects independent of response time differences.

Artificial words and objects analysis

First level models were constructed from all event types seen during testing runs (see-think and see-speak events for each of 7 conditions – words day 1, words day2, objects day 1, objects day 2, written objects day 1, written objects day 2, untrained words). The events were additionally split according to whether or not the response for each trial was correct or incorrect, leading to 28 event types in the first level model. Although all event types seen during testing were modelled, second level analyses focused only on trials in which participants responded correctly, to ensure a fair comparison between conditions even if differences in accuracy were observed. In order to assign accuracy to see-think trials (in which there was no behavioural response) we assumed that accuracy in the see-speak trials could be applied to the corresponding see-think trials for that same item.

A second-level model was constructed in SPM8 using conditions derived for trained items that were responded to correctly involving three factors: trial-type (see-think vs see-speak), item-type (words vs objects), and day of learning (day 1 vs day 2) (Henson & Penny, 2003, Technical Report). To assess neural responses for generalisation items (which were only presented in written form), a further second-level model was constructed in which we compared responses for trained words (day 1 vs day 1), with the written forms of object names (day 1 vs day 2), and untrained words leading to a 5-level factor. See-speak and see-think trials were collapsed in this analysis as one of the primary reasons for their inclusion was to deal with visual differences between words and objects. In this comparison of 5 word types the trial-types are of less importance. F-contrasts were constructed to identify significant main effects and interactions, and where significant effects were found, t-contrasts were used to explore the specific effects.

3.4 Results

3.4.1 Artificial words and objects during training

Due to problems with recording equipment data for 4 participants was not collected on all training runs. 16 participants had full data from both days 1 and 2. These 16 participants were entered into behavioural analysis of the training runs. Reading accuracy was better for words learned on day 2 than day 1, whereas accuracy at naming objects was similar for day 1 and day 2 items (Figure 3.3, p65; Table 3.3, p63). Data were entered in a repeated measures ANOVA. There was no overall effect of item type on performance, $[F(1,15) = 0.5,\ ns]$. The effect of day was significant with day 2 items showing higher performance than day 1 items, $[F(1,15) = 8.86,\ \eta_p^2 = 0.371,\ p < .01]$. Furthermore there was an interaction between item type and day, $[F(1,15) = 8.81,\ \eta_p^2 = 0.370,\ p = 0.01]$, with performance for objects remaining similar on both days but with improved word reading performance on the second day.

3.4.2 Results during scanning

Real words and objects in the localiser

Naming performance in all three conditions was very high (Table 3.3, p63) but response times were faster for words than objects, $[t_p(18) = 21.11,\ p < 0.001]$, and for words than pseudowords, $[t_p(18) = 8.32,\ p < 0.001]$. As described in the methods, the contrast $[objects - words]$ was conducted after taking between-condition differences in response time into account (see Figure 3.2, p61 and Table 3.3, p63 for details of response time differences). This was to ensure that we could localise regions more engaged by holistic as opposed to componential processing rather than regions that responded more strongly to object naming as it was more effortful than word reading. Objects showed more activation than words in bilateral inferior temporal gyrus as well as middle and inferior occipital gyri (Figure 3.4, p66, Blue; Table 3.4, p68). We next contrasted pseudoword with word activation, without accounting for response time differences, since both words and pseudowords should engage componential reading processes, but reading pseudowords is more effortful than reading words (see methods and Taylor et al., 2014a). Pseudowords activated bilateral middle and

Table 3.3 Experiment 1: Behavioural results

Accuracy on final run of training			
	Day 1 training *mean* (*SD*)	Day 2 training *mean*(*SD*)	
Naming objects	77% (13%)	79% (19%)	
Reading words	67% (13%)	86% (12%)	
Accuracy during scanning			
	Trained on day 1 *mean* (*SD*)	Trained on day 2 *mean*(*SD*)	Untrained *mean*(*SD*)
Naming objects	51% (50%)	63% (48%)	
Reading words	82% (38%)	82% (38%)	
Reading written objects	82% (38%)	83% (38%)	
Reading untrained words			80% (40%)
Performance on localiser task			
	Accuracy *mean* (*SD*)	Response times *mean*(*SD*)	
Naming objects	99.72 (3.43)	905 (238)	
Reading words	99.04 (2.67)	617 (113)	
Reading pseudowords	99.12 (2.19)	679 (158)	

inferior occipital gyri, left superior temporal gyri, left superior temporal gyrus, bilateral precentral gyri, bilateral postcentral gyri, left supplementary motor area, and left inferior frontal gyrus (triangularis), more than words (Figure 3.4, p66, Pink; Table 3.5, p69).

In order to compare activation for real words and objects with the artificial words and objects we constructed four spherical regions of interest based on peaks found in the functional localiser that corresponded most closely to peaks found in a meta-analysis of word and pseudoword reading (Taylor et al., 2013). The ROIs had a 10mm radius and were centred on: left posterior vOT (-42, -60, -8), left anterior fusiform gyrus (-28, -50, -16), left superior parietal lobe (-26, -60, 54), and left precentral gyrus (-54, 4, 26) (See white circles in Figure 3.4, p66). ROIs (2), (3) and (4) were chosen from peak locations for activation during pseudoword – word reading in Table 3.5 (including a sub-peak of the cluster in superior temporal and frontal regions). These coordinates are respectively 17.2*mm*, 7.48*mm*, 6.63*mm*, and 11.49*mm* distant from comparable peaks in Taylor et al. (2013).

Reading artificial words and naming artificial objects during scanning

Accuracy was higher for reading words than naming objects (Figure 3.3, p65; Table 3.3, p63). Whilst words from days 1 and 2 were read with equivalent accuracy, objects learned on day 1 were named less accurately than objects learned on day 2. A repeated measures ANOVA confirmed that accuracy was higher for reading words than naming objects, $[F(1,19) = 37.535,\ \eta_p^2 = 0.664,\ p < 0.001]$, and higher for items trained on day 2 compared to day 1 items, $[F(1,19) = 5.939,\ \eta_p^2 = 0.238,\ p = 0.008]$. Follow-up tests confirmed that naming accuracy for day 1 objects during scanning was significantly worse than for recently learned day 2 objects.

In the scanning session, in addition to reading the artificial words that were trained on days 1 and 2 (visually and phonologically familiar), participants also read the written forms of the object names trained on days 1 and 2 (visually unfamiliar, phonologically familiar), as well as a set of completely novel untrained words (visually and phonologically unfamiliar) during scanning. These 5 conditions were entered into a one-way repeated measures ANOVA, allowing us to ask whether visual learning, phonological learning, and/or a period of offline consolidation leads to improved reading performance. Although these data are grouped by factors of day, item-type, and training status, we do not analyse these factors separately. This decision is taken because the factors are not clearly discrete; could trained words from day 1 be treated the same as untrained words from day 1? The letters would be familiar and consolidated but the lexical forms would not. Consequently, the data would need to be analysed multiple times in order to address all potential groupings, with a consequent risk of inflating type-1 errors. Moreover, we do not have strong predictions for which of these factors would drive differences and so by analysing all factors in a one-way ANOVA we could explore any differences between conditions using post-hoc analyses. Accuracy in all of these conditions was very similar with no significant difference between any of the conditions, $[F(4,76) = .694,\ ns]$.

3.4.3 Whole-brain imaging analyses

To examine neural activity associated with reading trained words versus trained objects, and for items learned on days 1 and 2, we conducted a 2 x 2 x 2 repeated measures ANOVA to compare the effect of item type (word or object), day of learning (day 1 or day 2) and trial type (see-think or see-speak). We first examined the main effect of item type (words versus objects), collapsed across day of learning and trial type. The contrast $[words > objects]$ showed extensive activation in bilateral parietal cortices, as well as peaks in middle and inferior occipital gyri (posterior vOT regions), right supramarginal, bilateral precentral,

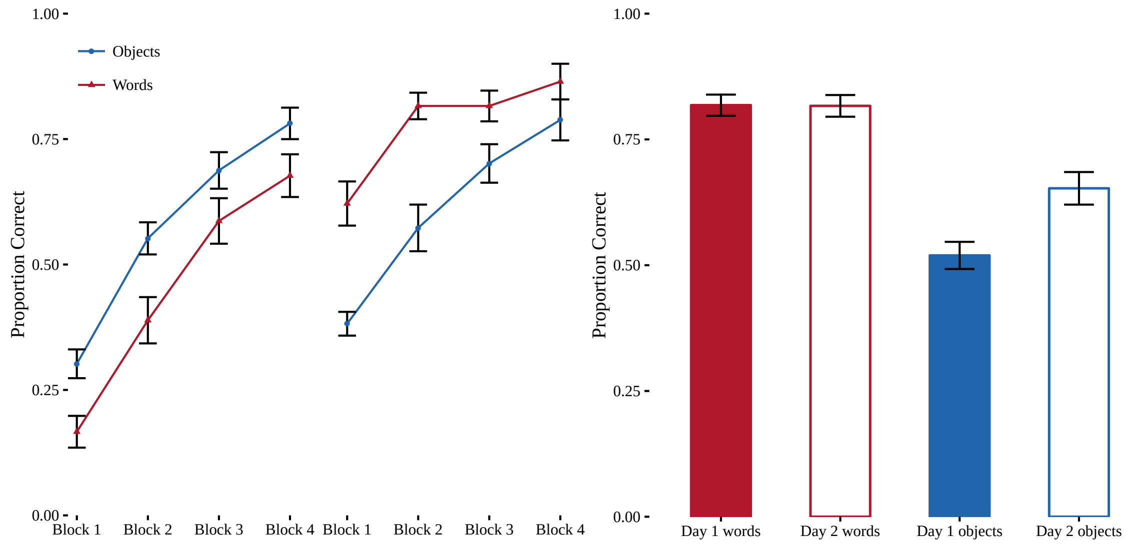


Fig. 3.3 (A) Mean accuracy for Object Naming (blue) and Word Reading (red) during the four blocks of training on each day, error bars show +/- 1 standard error of the mean after between subject variance removed suitable for repeated measures comparisons (cf. Loftus & Masson, 1994). (B) Accuracy in the scanner on day 2 while participants read/name items from both day 1 and day 2. Error bars as in panel 2A.

and bilateral middle frontal gyri, cerebellum, left supplementary motor area, and right hippocampus (Figure 3.5 in pink, p67; Table 3.6, p70).

The reverse contrast $[objects > words]$ revealed clusters in bilateral fusiform gyri (anterior vOT regions), bilateral angular gyri, left middle occipital cortex, left precuneus, left middle and anterior cingulate cortices, right inferior parietal cortex, bilateral superior and middle frontal gyri, left inferior frontal gyrus, left middle temporal gyrus, right cerebellum, right inferior temporal gyrus, right calcarine fissure, and left superior temporal pole (Figure 3.5 in light blue, p67; Table 3.6, p70).

As discussed in the methods, the main effect of word reading versus object naming does not rule out the possibility that low-level visual differences (such as differences in brightness or retinal extent) between words and objects may be driving differences in activation. We therefore examined the interaction between item and trial type to reveal changes in object versus word retrieval-related activity during repetition suppression (i.e. additional activity for see-think trials, compared to the see-speak trials that immediately followed and that contained the same visual form). The contrast $(words\ [see\text{-}think - see\text{-}speak] > objects\ [see\text{-}think - see\text{-}speak])$ revealed no clusters at an FWE cluster-corrected threshold of $p < .05$ (Figure 3.5, p67; Table 3.6, p70). However, the contrast $(objects\ [see\text{-}think - see\text{-}speak] > words\ [see\text{-}think - see\text{-}speak])$ revealed clusters of activation in bilateral fusiform gyri and

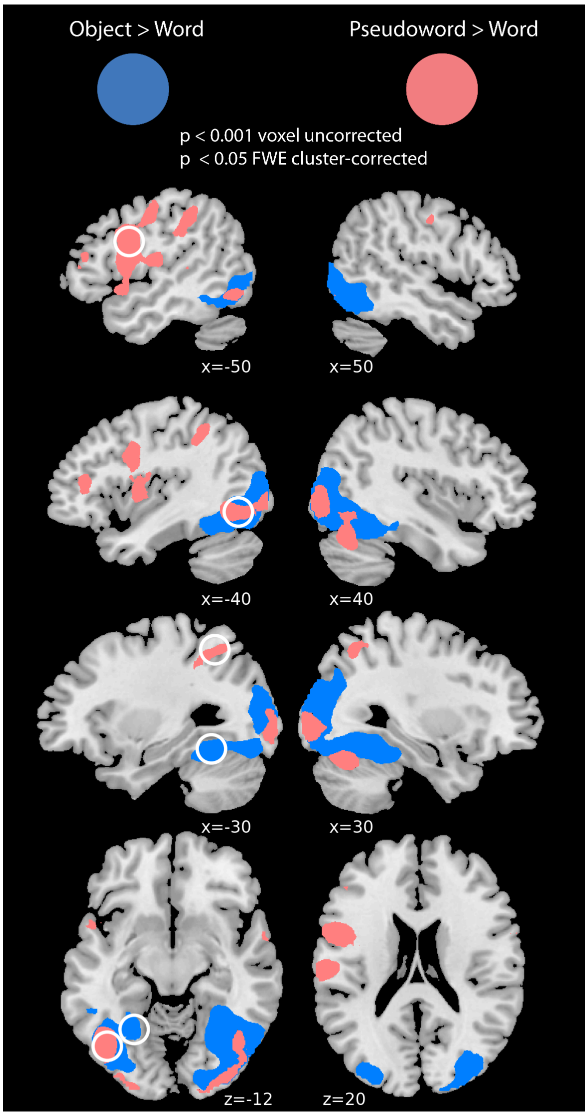


Fig. 3.4 Brain regions showing differential activation for contrasts of interest from functional localiser displayed on the MNI standard brain. Red = [pseudowords - words], blue = [objects - words]. Slices show activations at $p < 0.001$ voxel-wise uncorrected and $p < 0.05$ FWE cluster corrected. White circles indicate approximate locations of regions of interest defined from the functional localiser.

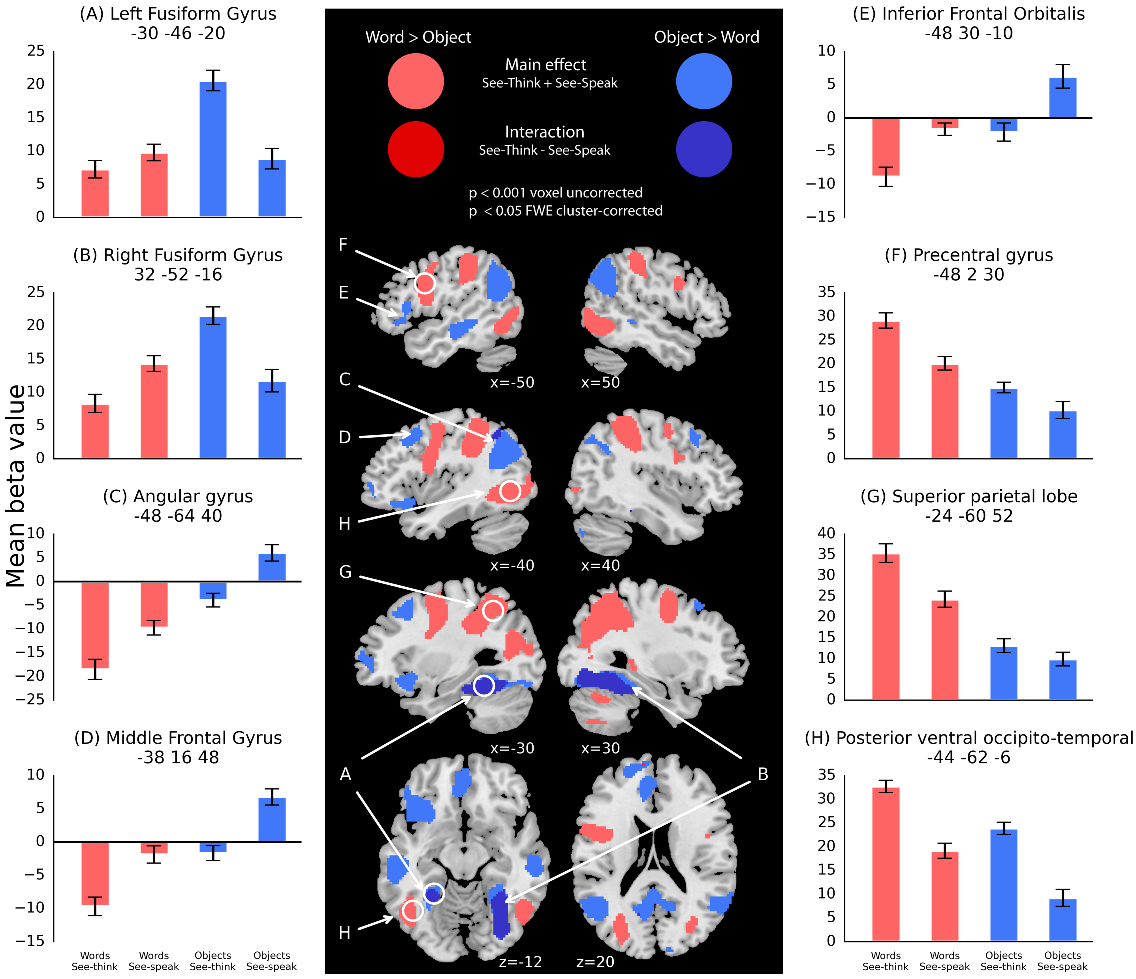


Fig. 3.5 Brain regions showing differential activation for contrasts of interest for the artificial words and objects in see-think and see-speak conditions. Pale red = [words [see-think + see-speak] - objects [see-think + see-speak]], pale blue = [objects [see-think + see-speak] - words [see-think + see-speak]], dark red = [words [see-think - see-speak] - objects [see-think - see-speak]], dark blue = [objects [see-think - see-speak] - words [see-think - see-speak]]. NB: Dark red activation does not reach statistical significance in this panel. Numbered plots show response profiles for contrasts versus the (unmodelled) resting baseline at peak locations labelled in the activation maps. White circles indicate approximate locations of regions of interest defined from the functional localiser.

left cuneus (Figure 3.5 in dark blue, p67; Table 3.6, p70). As shown in Figure 3.5A and 3.5B differential activity in left and right fusiform gyrus for objects compared to words is significantly greater during see-think than during see-speak trials. This might suggest that fusiform activity for objects is associated with initial identification and/or retrieving their names rather than being due to low-level visual differences. We will return to this point in the discussion.

The main effect of day showed no clusters surviving correction at whole-brain level. In addition, there was no interaction between day and item or trial type. The main effect of trial type was not of particular interest in this study, but is reported for similar, previous data in Taylor et al. (2014a).

Table 3.4 Brain regions involved in reading real words and objects in the localiser, thresholded at voxelwise $p < 0.001$ uncorrected, cluster extent corrected. The table reports the first three peaks more than 8 mm apart in each cluster.

Brain region (AAL)	Hemisphere	x	y	z	Voxels	Z Value	Cluster-level p value
Words - Objects (corrected for RT differences)							
Inferior parietal cortex	R	56	-50	42	9314	5.45	< 0.001
Supramarginal gyrus		64	-42	30			
Supramarginal gyrus		56	-38	44			
Cuneus	R	14	-70	36	7415	5.39	< 0.001
Precuneus		8	-52	36			
White matter		2	-38	24			
Cerebellum	L	-36	-66	-50	1013	5.19	< 0.001
Cerebellum		-8	-84	-40			
Cerebellum		-36	-78	-40			
Superior temporal gyrus	L	-48	-32	8	4985	5.16	< 0.001
Superior temporal gyrus		-58	-6	6			
White matter		-40	-30	8			
Inferior frontal opercularis	R	54	12	10	401	4.42	0.001
Insula		42	12	-12			
Insula		50	12	-4			
Middle frontal gyrus	L	-26	34	26	393	4.31	0.001
Middle frontal gyrus		-32	36	36			
Middle frontal gyrus		-28	28	32			
Objects - Words (corrected for RT differences)							
Inferior temporal cortex	R	54	-62	-12	5021	6.02	< 0.001
Middle occipital gyrus		36	-88	10			
Inferior occipital gyrus		46	-82	0			
Middle occipital gyrus	L	-24	-94	8	3395	5.27	< 0.001
Inferior occipital gyrus		-48	-62	-14			
Middle occipital gyrus		-36	-84	12			

Table 3.5 Brain regions involved in reading real words and pseudowords in the localiser, thresholded at voxelwise p < 0.001 uncorrected, cluster extent corrected. The table reports the first three peaks more than 8 mm apart in each cluster.

Brain region (AAL)	Hemisphere	x	y	z	Voxels	Z Value	Cluster-level p value
Words – Pseudowords							
Middle temporal gyrus	L	-56	-54	20	1893	5.50	< 0.001
Middle occipital gyrus		-40	-78	22			
Middle occipital gyrus		-38	-78	36			
Middle temporal gyrus	R	56	-58	14	1887	5.41	< 0.001
Middle temporal gyrus		42	-62	16			
Middle occipital gyrus		44	-62	28			
Cerebellum	L	-8	-56	-44	4312	5.40	< 0.001
Cerebellum		-14	-46	-42			
Middle cingulum		-8	-34	40			
Fusiform gyrus	L	-36	-36	-10	315	4.80	0.002
Fusiform gyrus		-28	-38	-16			
Fusiform gyrus		-26	-30	-24			
Middle frontal gyrus	L	-28	30	46	470	4.47	< 0.001
Middle frontal gyrus		-32	28	38			
Superior frontal gyrus		-24	38	44			
Superior frontal gyrus	R	24	34	46	663	4.45	< 0.001
Superior frontal gyrus		28	24	52			
Middle frontal gyrus		24	24	42			
Cerebellum	R	14	-46	-42	160	4.33	0.049
Cerebellum		-12	-52	-46			
Cerebellum		-4	-54	-44			
Medial frontal orbitalis	R	2	54	0	202	3.75	0.02
Medial frontal orbitalis		3	46	0			
Pseudowords - Words							
Cerebellum	R	36	-64	-22	1362	6.02	< 0.001
Middle occipital gyrus		44	-82	2			
Inferior occipital gyrus		36	-88	-2			
Superior temporal gyrus	L	-62	-18	6	3930	5.27	< 0.001
Precentral gyrus		-56	4	26			
Postcentral gyrus		-60	-18	20			
Inferior temporal gyrus	L	-42	-60	-8	802	4.60	<0.001
Middle occipital gyrus		-30	-94	-4			
Inferior occipital gyrus		-40	-74	-8			
Supplementary motor area	L	-4	2	60	205	4.30	0.019
Postcentral gyrus	R	66	-10	34	259	4.02	0.006
Precentral gyrus		62	4	28			
Precentral gyrus		46	-10	40			
Inferior frontal triangularis	L	-48	36	16	192	3.80	0.025
Inferior frontal triangularis		-40	36	14			
Inferior frontal triangularis		-42	34	6			

Table 3.6 Experiment 1: Brain regions involved in reading artificial words and naming artificial objects.

Brain region (AAL)	Hemisphere	x	y	z	Voxels	Z Value	Cluster-level p value
Words – Objects collapsed across Day and See-Think/See-Speak condition, correct responses only							
Superior parietal cortex	R	24	-62	54	6171	Inf	< 0.001
Supramarginal gyrus		46	-30	44			
Inferior temporal gyrus		50	-60	-8			
Superior parietal cortex	L	-24	-60	52	14011	Inf	< 0.001
Inferior parietal cortex		-40	-38	40			
Middle occipital gyrus		-28	-68	26			
Cerebellum	R	26	-70	-50	1817	7.13	< 0.001
Cerebellum		26	-64	-26			
Cerebellum		6	-70	-24			
White matter	L & R	22	4	-10	668	5.72	< 0.001
White matter		16	30	6			
White matter		-16	30	2			
Precentral gyrus	R	48	6	30	416	5.53	< 0.001
Objects - Words collapsed across Day and See-Think/See-Speak condition, correct responses only							
Angular gyrus	L	-48	-64	40	3799	Inf	
Middle occipital gyrus		-40	-72	36			
Middle temporal gyrus		-58	-24	-12			< 0.001
Angular gyrus	R	44	-62	38	1969	7.57	
Angular gyrus		54	-58	26			
Inferior parietal gyrus		54	-56	46			< 0.001
Middle cingulum	L	-2	-40	36	3890	7.34	
Precuneus		-8	-66	32			
Precuneus		-6	-60	16			
Middle frontal gyrus	L	-38	16	48	10405	6.98	< 0.001
Medial superior frontal gyrus		-4	42	30			
Medial superior frontal gyrus		-36	56	0			
Fusiform gyrus	L	-28	-46	-14	1177	6.38	< 0.001
Lingual		-28	-84	-14			
Lingual		-24	-96	-12			
Fusiform gyrus	R	30	-54	-10	1046	5.98	< 0.001
Fusiform gyrus		32	-74	-12			
Fusiform gyrus		28	-42	-14			
Cerebellum	R	18	-84	-34	446	5.45	0.001
Cerebellum		44	-60	-42			
Cerebellum		46	-68	-40			
Inferior temporal gyrus	R	58	-24	-16	687	5.26	< 0.001
Middle temporal gyrus		48	-36	-2			
Calcarine fissure	R	22	-100	-4	301	5.24	0.007
Cuneus		12	-98	16			
Calcarine fissure		14	-102	4			
Words[see-think - see-speak] - Objects[see-think - see-speak], collapsed across Day, correct responses only							
Medial superior frontal gyrus	R	12	58	30	648	4.23	< 0.001
Superior frontal gyrus		-14	58	30			
Medial superior frontal gyrus		10	64	14			
Objects[see-think - see-speak] - Words[see-think - see-speak], collapsed across Day, correct responses only							
Fusiform gyrus	R	32	-52	-16	1478	6.60	< 0.001
Fusiform gyrus		32	-70	-12			
Fusiform gyrus		36	-32	-22			
Fusiform gyrus	L	-30	-46	-20	1021	6.83	< 0.001
Cuneus	L	-12	-64	30	324	5.05	0.005
Inferior parietal cortex	L	-44	-54	46	223	4.59	0.026

3.4.4 ROI analyses based on real words and objects in the localiser

Table 3.7 Experiment 1: Region of interest analyses

Region	Centre of Mass			Item type (words vs objects)			Trial type (see-think vs see-speak)			Interaction	
	X	Y	Z	$F(1,19)$	η_p^2		$F(1,19)$	η_p^2		$F(1,19)$	η_p^2
Anterior fusiform	-28	-50	-16	33.92***	0.64	Objects > Words	0.45			22.15***	0.54
Left posterior vOT	-42	-60	-8	15.60***	0.45	Words > Objects	44.34***	0.70	See-only > see-speak	1.14	
Precentral gyrus	-56	4	26	28.43***	0.60	Words > Objects	12.94**	0.40	See-speak > see-only	0.01	
Superior parietal lobe	-26	-60	54	53.79***	0.74	Words > Objects	6.73*	0.26	See-only > see-speak	0.14	

Region	Centre of Mass			Reading generalisation conditions
	X	Y	Z	$F(1,19)$
Anterior fusiform	-28	-50	-16	0.709
Left posterior vOT	-42	-60	-8	0.364
Precentral gyrus	-56	4	26	0.319
Superior parietal lobe	-26	-60	54	0.063

* p < 0.05; ** p < 0.01; *** p < 0.001.

The ROIs defined in the localiser allow us to ask whether activation for trained artificial words and objects overlapped with that for real words and objects. Four regions of interest were defined based on the localiser, one in left anterior fusiform where objects show more activation than words, and three showing greater activation for pseudowords than words in left posterior vOT, left precentral gyrus, and left superior parietal lobe (the white circles in Figure 3.4, p66, shows where the ROIs were defined while the white circles in Figure 3.5, p67, show how the ROIs relate to the whole-brain activation for the artificial words and objects). In each ROI we conducted a 2x2 repeated measures ANOVA, with the factors item type and trial type, collapsed across day (Table 3.7, p71). In the anterior fusiform ROI where real objects showed greater activation than real words, artificial objects also showed greater activation than artificial words and this was more pronounced for see-think than see-speak trials (similar to the interaction profile plotted in Figure 3.5A). This demonstrates that object relative to word activation in this region was not purely driven by visual differences between the item types, but by retrieval of holistic as opposed to componential visual-verbal associations. In the left posterior vOT ROI (similar to Figure 3.5H), the precentral ROI (similar to figure 4F) and in the superior parietal ROI (similar to Figure 3.5G) we saw activation consistent with contributions to reading artificial words with a main effect of item-type (words > objects), and of trial type (see-think > see-speak) but no interaction between these factors. Our confidence that these left hemisphere activation differences were driven by processes involved in recalling spoken from visual forms, as opposed to purely visual differences, is greater for objects in anterior fusiform, than for words in posterior vOT, precentral gyrus, and superior parietal cortex.

We used these same four ROIs to examine whether visual or phonological familiarity influenced neural activity during word reading. In each ROI, we conducted a one-way repeated measures ANOVA with the five word reading conditions (collapsed across trial-type): trained artificial words from day 1 and day 2, the written forms of artificial objects trained on days 1 and 2, and a set of untrained artificial words. This comparison allows us to ask whether visual and phonological familiarity (as for the trained words), or purely phonological familiarity (as for the written objects), or a period of offline consolidation (as for the day 1 items) is necessary to support more word-like representations. There was no significant difference between the five conditions in any of the four regions of interest (Table 3.7, p71).

At the suggestion of a reviewer (for the paper Quinn, Taylor, & Davis (2017) Learning and retrieving holistic and componential visual-verbal associations in reading and object naming. *Neuropsychologia*) we additionally analyse these comparisons based on (1) whether the written words are visually familiar (trained words from days 1 and 2) or unfamiliar (written forms of objects from days 1 and 2, as well as untrained words), and (2) whether the written words are phonologically familiar (trained words from days 1 and 2 as well as written forms of objects from days 1 and 2) or unfamiliar (untrained words). Comparison of these conditions in each of the four regions of interest confirm no difference due to either visual or phonological familiarity (anterior fusiform: visual familiarity, $[t(19) = 1.314,\ ns]$, phonological familiarity: $[t(19) = 0.581,\ ns]$; posterior vOT: visual familiarity, $[t(19) = 0.358,\ ns]$, phonological familiarity: $[t(19) = 1.406,\ ns]$; superior parietal: visual familiarity, $[t(19) = 0.775,\ ns]$, phonological familiarity: $[t(19) = 0.961, ns]$; precentral gyrus: visual familiarity, $[t(19) = 0.329,\ ns]$, phonological familiarity: $[t(19) = 1.441,\ ns]$).

3.4.5 Hippocampal ROI analysis

Complementary Learning Systems (CLS) accounts of word learning (Davis and Gaskell, 2009) suggest a key role for the hippocampus, and indeed previous studies have shown changes in hippocampal responses to recently learned spoken words (Breitenstein et al., 2005; Davis et al., 2009; Takashima et al., 2014b). As there is strong a priori reason to expect an effect of day of learning on hippocampal activity, regions of interest analyses were conducted using two separate AAL masks for the left and right hippocampi. Activation values were combined across see-think and see-speak trials and entered into a 3-way repeated measures ANOVA with factors day of training (day 1 or day 2), item type (word or object) and lateralisation (left or right hippocampus). Hippocampal activation was greater for day 2 than day 1 items, $[F(1,19) = 4.758,\ \eta_p = .20,\ p = 0.042]$. There was no main effect of laterality, $[F(1,19) = 0.571,\ ns]$, or item type, $[F(1,19) = 0.049,\ ns]$, and no interactions reached significance (Figure 3.6, p73).

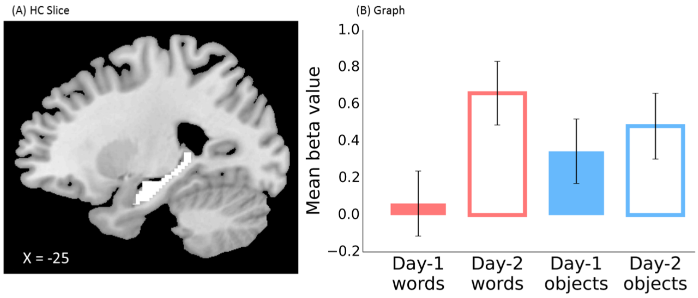


Fig. 3.6 (A) Location of a bilateral hippocampus ROI defined using the AAL template brain. (B) This region showed differential activation (graphed as the contrast of activation versus an unmodelled resting baseline) for novel words and objects learned on days 1 and 2, collapsed across see-think and see-speak conditions and left/right hippocampus. Red = words, blue = objects, solid bars = day 1 items, empty bars = day 2 items.

3.5 Discussion

Both educational and cognitive perspectives on reading have highlighted a critical distinction between holistic and componential processing, i.e. recognising whole-word forms and decoding words letter-by-letter. Here we have compared an artificial learning paradigm to object naming and word reading of familiar, real language stimuli in order to identify the neural systems that support holistic and componential visual-verbal mappings and their acquisition. Our findings combine to demonstrate both behavioural and neural dissociations between holistic and componential mappings, which we will discuss in turn. We will start by summarising behavioural evidence for this distinction as shown by differences in learning profiles and generalisation, before moving onto discuss ventral occipito-temporal, parietal, and frontal contributions to reading and naming of both real and artificial items. We will conclude by returning to the educational issues that we introduced at the outset and consider the broader implications of our findings.

3.5.1 Behavioural results highlight componential and holistic learning

Behavioural results during training confirm that learning to read artificial words involved componential learning whereas learning to name artificial objects involved holistic learning. Participants become better at naming objects across four runs of training on day 1 but when they return on day 2 to learn 18 more objects their learning profile was essentially the same as on day 1. This pattern is due to the holistic and arbitrary relationship between the visual and verbal form of an object; knowledge of object-names from day 1 does not help to name objects learned on day 2. In contrast, the componential and systematic relationship between the visual and verbal forms of written words means that letter-sound mappings learned on day 1 are also effective in supporting reading of items learned on day 2. Hence, reading performance at the start of day 2 is substantially better than at the start of day 1.

The distinction between holistic and componential learning is further borne out by behavioural performance during scanning. Participants were significantly worse at naming objects learned on day 1 compared to those learned on day 2, due either to forgetting of more distantly learned object names or interference from object names learned immediately prior to scanning on day 2. The present data does not distinguish these two possibilities (Anderson, 2003; Mensink and Raaijmakers, 1988). Whichever explanation we invoke, however, interference or forgetting arises from the holistic and arbitrary nature of visual-verbal mappings for object names; items learned on day 2 do not support, and might even interfere with, items acquired on day 1. In contrast, reading performance in the scanner was equally good for words learned on both days. Thus, the significant interaction between day and item type is again consistent with componential knowledge in reading words aloud since items learned on day 2 contained the same letter-sound correspondences, and therefore supported successful reading of words learned on day 1. Finally, the ability of participants to read untrained words accurately further demonstrates their ability to generalise this letter-sound knowledge to novel written words.

Hence our participants have acquired the ability to decode written words. This is the same skill as is taught to beginning readers through phonics instruction. These findings therefore support our use of functional neuroimaging of artificial language learning to explore the neural basis of learning to read. To the extent that neural activity overlaps for real and artificial items, we can further argue for parallels between the processes that support skilled reading/naming, and processes recruited for reading/naming our artificial materials. This will be the focus of the next two sections of discussion.

3.5.2 VOT contributions to reading and naming

A variety of evidence reviewed in the introduction suggests hierarchical organisation of visual processing in vOT regions. Results of the functional localiser to some extent support these proposals as the componential reading contrast [*pseudowords > real words*] showed activation in lateral and posterior vOT regions (replicating Mechelli et al., 2005; and others, see Taylor et al., 2013 for a meta-analysis). However, the holistic contrast of [*real objects > real words*] also showed activation throughout vOT, including posterior as well as mid- and anterior vOT regions. This finding appears more compatible with views of vOT specialisation that propose association with phonological representations as a key factor (Price and Devlin, 2011a) rather than specialisation for alphabetic forms (Cohen et al., 2000; Dehaene and

Cohen, 2007, 2011). Our use of an overt naming task might be critical in explaining this observation. Covert reading/naming may induce greater phonological processing for word reading than object naming since phonological access is automatic for written words (Hagoort et al., 1999; MacLeod, 1991; Price et al., 1996; Song et al., 2010b; Twomey et al., 2011). Nonetheless, we also observed that mid- and anterior vOT regions showed an additional response to objects than to words consistent with a contribution to processing holistic visual forms in hierarchically higher levels of the vOT.

Activation for the artificial words and objects matches the hierarchical organisation of vOT responses seen both in the localiser and in previous literature. We observed greater activation for reading artificial words than naming artificial objects in bilateral posterior vOT, whereas the reverse profile of greater activation for naming artificial objects than reading artificial words was observed in bilateral anterior vOT. Furthermore, this differential response in anterior vOT interacted with trial type, such that object > word activation was more pronounced for see-think than see-speak trials.

We preceded see-speak trials with a see-think trial for the same item primarily for pragmatic reasons: (1) it avoids articulation on the majority of trials, reducing head movement, because see-think trials occurred twice as often as see-speak trials (2) it would otherwise have been difficult for participants to produce item names in the short silent interval between two scans, (3) it permits separation and comparison of neural activity during covert (see-think) and overt (see-speak) articulation. That participants were able to respond fast enough on see-speak trials can be seen as a form of behavioural priming by which articulation of an item name is faster if it has been presented on an immediately preceding trial. Studies of neural repetition suppression have shown reduced activity for repeated items in a similar anterior fusiform region to that showing repetition suppression for artificial objects but not written words in the present study (Fig. 4A, cf. Glezer et al., 2009; Kherif et al., 2011). Previous work has argued that the reduction in activity on see-speak compared to see-think trials reflects the fact that phonological retrieval primarily occurs on see-think trials (Taylor et al., 2014a). Hence, we proposed that anterior fusiform plays a greater role in processing holistic object-name associations than componential written word pronunciations. However, at the suggestion of reviewers, some more careful consideration leads us to acknowledge that there may also be visual contributions to repetition suppression. It is possible that anterior fusiform regions contribute to visual configural processing unique to objects and that this process, instead of, or as well as, phonological retrieval, is reduced on repeated trials (James et al., 2002; Vuilleumier et al., 2002). While other studies were able to rule this out (for example, Kherif et al. (2011) showed repetition suppression following pairs of non-identical pictures with the same name), our paired presentations involved the same visual form as well as the same name. Further investigations using names for objects that are depicted in multiple different pictures, and/or written words in multiple fonts, might help assess whether anterior fusiform is primarily involved in holistic relative to componential visual configural processing, or in the retrieval of holistic rather than componential visual-to-verbal associations.

One effect that we failed to see for the artificial orthography, which we had anticipated based on findings for reading real words and pseudowords, was differential activation for trained versus untrained words. This null effect is notable considering the comparison involved 216 trials for trained words with 216 trials for untrained words (108 trials for the written forms of objects and 108 for completely untrained words), while the localiser showed differences between words and pseudowords with only 60 trials of each. This outcome, in conjunction with frontal and parietal activation for trained words relative to objects that we will discuss subsequently, might suggest that trained words were still being read componentially. It may be the case that our design included an insufficient number of training presentations (4 presentations of each word) to produce whole-word representations and that more intensive or longer-lasting training is required in order to generate holistic representations for written words in an artificial orthography. Future work will address this possibility. It might also be the case that adding irregular spelling-sound mappings would increase the necessity for these whole word representations. Note however, that cognitive models of reading, such as the DRC model, propose that whole-word representations develop for both regular and irregular forms.

If holistic word representations were to emerge with further training we might expect trained items to activate anterior fusiform regions more than untrained items, in a similar manner to real as well as trained objects relative to words in the current study. In contrast, untrained words might activate left posterior occipitotemporal, parietal, and frontal regions more than trained words, in a similar way to pseudowords relative to real words. This would support neuroimaging studies that, in line with cognitive models of reading, have suggested a distinction between sub- and whole-word processes in dorsal versus ventral brain regions (Taylor et al., 2013). Note however, that although the DRC model proposes lexical orthography-to-phonology mappings, the triangle model (Plaut et al., 1996) proposes that such item-specific mappings primarily emerge for the mapping between visual/phonological word forms and their meanings. Thus, this model might predict activation for trained relative to untrained items in anterior fusiform only if artificial words were trained with meanings.

3.5.3 Parietal and frontal contributions to reading and naming

In addition to vOT contributions, there is substantial evidence for frontal and parietal regions supporting componential reading processes. For example, we saw extensive activation of parietal and frontal networks for the contrast of [*pseudoword* > *word*] reading in the localiser scan. This replicates a large number of previous observations in the functional imaging literature (see Taylor et al., 2013 for a meta-analysis). We will discuss the implications of these findings separately, first for parietal and then for frontal regions.

The observed increased activation in inferior parietal cortices for word reading over object naming replicates results previously reported by Taylor et al. (2014b). However, our results go beyond these in two ways. First, by demonstrating substantial overlap between activation contrasts for artificial and real language stimuli (as scanned during a functional localiser). Second, by showing that differences in parietal involvement can be seen in an event-related design in which trials presenting artificial words and objects are randomly intermixed, rather than the blocked presentation used by Taylor et al. (2014a). This suggests that activation differences can be evoked solely by stimulus differences in the absence of the more strategic effects that are possible for blocked designs.

Unlike Taylor et al. (2014b), we did not find an interaction between item type (words vs objects) and trial type (see-think vs see-speak) in parietal regions. We cannot therefore be certain that parietal involvement in reading words reflects the retrieval of written word pronunciations, independent of the perceptual differences between words and objects. Indeed, some have primarily associated parietal activity during reading with perceptual processes. For example Cohen et al. (2008) using fMRI, and Rosazza et al. (2009) using EEG, argued that parietal regions are primarily active when readers are forced into an "attention-based serial reading strategy" (p 361) by changes in the visual form (orientation or degradation) of written words. A similar, serial visual processing, interpretation was offered by Protopapas et al. (2016) who obtained length effects in this region during pseudoword reading.

However other data argue against a purely visuo-spatial attention account of parietal activation during reading. Carreiras et al. (2014) demonstrated selectivity in left inferior and superior parietal regions for coding the identity and positions of letters, relative to symbols and numbers, suggesting a role for this region in processing stimuli that have linguistic associations. Parietal activation has also been shown to be greater when participants make judgements about spelling-to-sound mappings, over and above judgements about spellings or sounds in isolation, again implicating this region in cross-modal processing (Booth et al., 2003, 2007). Thus, in line with the proposal made by Taylor et al. (2013; 2014), we suggest that engagement of parietal regions for pseudoword relative to word reading, and for artificial word reading relative to object naming, reflects their role in the translation of component letters into sounds.

In our study, posterior frontal (precentral gyrus) and parietal regions largely show the same response profile. Both regions are activated in the localiser for pseudowords vs real words, and in the contrast of artificial words vs objects. Previous studies have implicated the frontal regions activated here (specifically the precentral gyrus) in the selection and assembly of phonological outputs (Bookheimer, 2002; Devlin et al., 2003; Gough et al., 2005). Left precentral gyrus not only shows increased activation for pseudoword compared to word reading (Taylor et al., 2013; 2014b), but also reduced activation following consolidation of new spoken words (Davis et al., 2009). Furthermore, a recent fMRI study has dissociated posterior frontal regions (such as the precentral gyrus) that contribute to phonological output from more anterior inferior frontal regions, which may contribute to phonological selection (this latter process is particularly engaged by reading irregular words, Taylor et al., 2014b). This distinction between phonological selection and phonological assembly is consistent with our finding of increased activation for naming artificial objects than reading artificial words in left IFG orbitalis, and the reverse profile in left precentral gyrus.

3.5.4 Emergence of holistic representations for consolidated items

As mentioned above, even after a night of consolidation for day 1 words we did not find differences between trained and untrained words either in behavioural performance or in the regions of interest defined from the functional localiser. Nonetheless, there was some evidence of whole-word learning since the hippocampus was differentially active for items learned on days 1 and 2. Consistent with the predictions of CLS accounts, there was more hippocampal activation for both words and objects learned on the same day as scanning as compared to the previous day. As the words from days 1 and 2 share the same letters, this reduced activation for day 1 words would not be possible unless a whole-word representation of some form existed. Hence, evidence that consolidation impacts on hippocampal activation provides evidence for some form of holistic representation for the day 1 words. However, given the lack of reliable activation differences between trained and untrained written words we cannot be confident about where these representations reside. Further studies should extend and adapt the time period of training and scanning to answer this question. Further work could then also determine the relative roles of sleep and time in the consolidation of orthographic and lexical representations. This question concerning the role of item-level consolidation in early stages of learning to read has implications for the role of consolidation in literacy learning more generally. Although lexical consolidation has been shown for school-age children (Henderson et al., 2013), this has primarily been in the context of learning spoken and not written words.

3.5.5 Validation of laboratory-based learning paradigms and implications for education

The extensive overlap of activation between the functional localiser and the artificial items, combined with the behavioural evidence, shows that laboratory studies of reading can engage holistic and componential learning mechanisms. This is a striking finding when we consider that we are comparing words and objects that have been used since childhood with artificial items that have been learned at most one day prior to scanning. This outcome suggests that laboratory-based learning studies can activate the same neural systems as engaged in more ecologically-valid paradigms (e.g., in studies of beginning readers). However, our artificial orthography studies have an advantage of maintaining strict experimental control. Neuroimaging results for real language stimuli may be sensitive to a wide variety of linguistic features: word frequency, age-of-acquisition, etc. which can be readily controlled in laboratory learning studies. Similarly, real language stimuli may be subject to individual differences in terms of language and literacy exposure that are again readily controlled in a laboratory setting. Finally, educational questions about early literacy acquisition are dependent on a wide range of external factors that may influence classroom outcomes. Consequently we suggest laboratory-based approaches offer an important complement to more naturalistic studies.

In education, the relative importance of holistic and componential reading strategies is much debated, with componential phonics-based approaches to reading instruction being dominant. Our results are consistent with the importance of componential learning during the earliest stages of reading acquisition: we see activation in posterior vOT, parietal, and frontal regions for the contrast of artificial word greater than object naming. However, in contrast to imaging findings with skilled readers and English words we see relatively little activation evidence for holistic representations of artificial written words despite abundant evidence for holistic representation of novel objects.

While fronto-parietal activation has been implicated in many studies of word reading, the relative contributions of dorsal and ventral regions across different stages of learning to read has remained relatively unclear, with several studies citing the need to further investigate the relationship between parietal and vOT contributions (Carreiras et al., 2014; Cohen et al., 2008; Price, 2012; Reilhac et al., 2013). Although fMRI studies of children have shown parietal involvement (Bitan et al., 2007b; Cao et al., 2006; Hoeft et al., 2007), children old enough to undergo scanning already have several years of exposure to written words. By tracking the very earliest stages of learning to read, our laboratory-based approach offers a way to identify the contributions of parietal and vOT regions over the very earliest stages of learning to read. That these areas show extensive activation prior to neural evidence for holistic representations speaks to the important role of componential mechanisms in early stages of acquisition. Given functional imaging and neuropsychological evidence for parietal contributions to spatial encoding of written strings (Carreiras et al., 2014; Cohen et al., 2008), our findings motivate further work to explore parietal contributions to successful and unsuccessful literacy acquisition (Peyrin et al., 2011; Reilhac et al., 2013).

Chapter 4

Lexical and sublexical orthographic consolidation

1. This chapter shows clear evidence for overnight consolidation at the level of letters, replicated across three experiments.
2. There was partial evidence for consolidation of written and spoken lexical forms.
3. Overnight consolidation at the level of letters may thus play an important role in the initial stages of learning to read.

4.1 Background

A particular benefit of adopting a tightly controlled experimental approach to track reading development is that it allows us to isolate and investigate competing cognitive processes. In order to fluently translate written text to spoken language (and ultimately meaning) learners must combine spoken language with knowledge of letters and whole written words. Tracking the ways in which knowledge of letters and words interact with one another during literacy acquisition has proved challenging in naturalistic contexts, in part because of the mutual interdependence of these skills during learning. This question has played out most clearly in the debate between the componential, letter-level focus of phonics instruction methods versus the lexical and meaning focus of whole word instruction methods. To what extent does knowledge of letters support learning whole word forms, and vice versa? Lab-based learning of orthography allows traction on these issues. Here overnight consolidation is used as an experimental factor to try to track the emergence of these skills in fine detail. Overnight consolidation has been shown elsewhere to influence the learning of spoken words and improvements in visual object recognition (Bowers et al., 2005; Davis et al., 2009; Dumay and Gaskell, 2007; Gaskell and Dumay, 2003). Thus the existing evidence may point to consolidation both at the level of letters and words.

This chapter aims to isolate whether overnight consolidation will impact on the learning of a novel writing system; and further whether any consolidation takes places at a lexical or sub-lexical level. While results from Experiment 1 suggest consolidation of written information may be taking place, the experimental design is not capable of identifying whether consolidation effects have taken place at lexical or sublexical levels. Across three experiments we train participants to read two distinct orthographies over two days, one on each day of training, before testing knowledge of both on day 2. Letters and trained words learned on day 1 therefore have had a period of consolidation. Thus, similar performance reading trained and untrained words in the day 1 orthography would indicate consolidation at a sublexical level; day 1 letters are shared by both trained and untrained words. In contrast, different performance for trained and untrained words would indicate consolidation at the level of the whole word.

4.1.1 Outstanding issues from Experiment 1

The findings of Experiment 1 demonstrated that participants could generalise their knowledge of orthographic information in order to read untrained words. The ability to generalise componential knowledge of letter-sound mappings to read an unfamiliar word is a critical feature of reading that differs from both spoken word recognition and visual object recognition.

One proposed benefit of sleep consolidation is that knowledge becomes independent of the specific context in which it has been learned and so can be generalised to novel situations (Frankland and Bontempi, 2005; McClelland et al., 1995; Tamminen et al., 2012). It is possible that improvements in the ability to generalise orthographic information may have taken place in Experiment 1; participants may have consolidated either visual knowledge of the letters or knowledge of the letter-sound mappings. The experiment ran across two days with generalisation tested during scanning on the second day (by reading untrained words). An important design choice of Experiment 1 was to use the same letters on days 1 and 2. We could therefore test whether participants could generalise the recently learned orthography in order to read the written forms of the objects. Although this design decision allowed us to test generalisation of the spoken words associated with objects from each day, the consistency of the orthography across the two days of the experiment meant that it was not possible to identify whether sub-lexical (orthographic) consolidation has taken place.

Two contrasting pieces of evidence from Experiment 1 suggest consolidation of some form may have taken place. First, the hippocampal ROI analysis demonstrated reduced responses to words and objects learned on day 1 relative to those learned on day 2. A second, indirect, suggestion of consolidation comes from the behavioural results that demonstrated a strong ability to generalise orthographic knowledge in order to read untrained words compared to a previous experiment using a similar design but without an overnight manipulation (65% in Taylor et al. (2014a) versus 82% in Experiment 1 of this thesis). While not directly comparable the differences between these findings raises the question of whether consolidation takes place at a lexical or sub-lexical level.

In addition to investigating whether consolidation takes place at a lexical or sub-lexical level we attempted to improve the experimental design and to address potential weaknesses from the first experiment. These weaknesses included: (1) a (relatively) unspeeded measure of learning in which participants were (2) trained and tested on the same task, using (3) highly confusable stimuli, and that showed (4) poor levels of learning.

The first of these potential reasons for the lack of consolidation effects is that the reading aloud task gave participants a relatively long window in which to respond (3500ms). It may be the case that this longer time window allowed participants to generalise their orthographic knowledge using episodic representations of the written words. Tamminen et al. (2012) suggest explicit reasoning processes may mask any overnight changes in the representations of recently learned knowledge, and that unspeeded tasks may encourage reasoning based on episodic memory. In addition, testing participants using the same task on which they were trained may not be a sensitive measure of generalisation. Instead, to test participants' ability to apply their learning in a novel task, and to encourage participants towards speeded responses we introduce a novel spoken-to-written matching task as the primary measure for Experiments 2, 3, and 4 (details below). Notably other studies investigating overnight consolidation have found changes that are primarily related to the speed of performance, e.g., Fischer et al. (2002). Adopting a timed measure of performance may therefore be more sensitive to overnight changes.

The third potential reason for the lack of evidence for lexical consolidation in Experiment 1 is the confusability of the stimuli. A strength of laboratory-based approaches is the ability to control potential confounds such as word frequency or age-of-acquisition effects. The pseudowords used in Experiment 1 exemplify this statistically controlled approach by ensuring each item set (trained day 1, etc.) contained each consonant three times in onset and three times in coda position. However, an unintended consequence of this statistically-balanced experimental decision is that the items learned on each day occupy an extremely dense phonological space (e.g., bAv, bEv, bVv, and so on). Items are therefore very confusable and potentially lacking distinctiveness. These experimental pseudowords differ therefore from natural language where words evolve to maximise the distinctiveness between items, and thus improve information transmission (Monaghan et al., 2011; Roy et al., 2015). The relatively low d' scores for trained words in the recognition memory task from Experiment 1 supports this perspective (lowest value of 0.94 for day 2 words and highest of 1.26 for day 1 objects). By contrast previous studies that have shown consolidation of novel spoken words have typically used lexical items with few neighbours (e.g., *cathedral* and *cathedruke* in Gaskell and Dumay, 2003) and correspondingly high distinctiveness. It may be the case that distinctiveness is an important part of developing a novel representation overnight. To address this issue the stimuli for Experiment 2 were constructed in such a way as to increase the distinctiveness of the pseudowords, while still maintaining statistical balance in the stimuli.

A final possible reason for the lack of evidence of overnight consolidation for the novel objects was that participants did not learn to a very high level ($\approx 50\%$ accuracy for object naming). It may be the case that overnight memory consolidation is dependent on the strength of the memory representation before sleep (Peigneux et al., 2004; Talamini et al., 2008). We adapt the training paradigm to encourage higher levels of learning to address this concern.

The primary question of Experiment 2 asks whether consolidation takes place at the lexical or sub-lexical level. A large body of evidence suggests that sleep consolidation is beneficial in the learning of novel spoken words (Davis et al., 2009; Dumay and Gaskell, 2007; Gaskell and Dumay, 2003; Merkx et al., 2011; Tamminen et al., 2012). It is not clear however whether these effects will extend to orthographic learning and consolidation. On the other hand, one manner in which sub-lexical consolidation of reading could occur is through overnight changes in the representation of visual knowledge. As noted in the literature review, there is a wider literature showing evidence for the involvement of sleep in the acquisition of a range of motor and perceptual skills (e.g, Korman et al., 2007; Stickgold et al., 2000). These improvements in visual knowledge may play a supporting role in learning and generalising recently learned letters.

4.2 Experiment 2

Experiment 2 builds on the findings of Experiment 1 by comparing the effect of consolidation on lexical and sublexical orthographic learning. The experiment runs over two days with participants learning two distinct orthographies, one on each day. The orthographies did not overlap; a phoneme (and corresponding grapheme) encountered on day 1 would not be included in the day 2 items. Thus consolidated and unconsolidated orthographies could be compared during a testing period following training on day 2. Furthermore, a set of untrained words was created for each day. By comparing participants' performance when reading trained and untrained words we compared the effect of lexical learning and consolidation independently from any impact of sublexical consolidation.

If sublexical, but not lexical, consolidation takes place then we should expect better performance for both trained and untrained words from day 1. Evidence of lexical learning in the absence of consolidation-related improvements would be indicated by better performance for trained words from days 1 and 2, relative to untrained words. Finally, if whole-word representations have been consolidated after a night of sleep we expect better performance only for the trained words from day 1 relative to all other item types.

We address the outstanding issues from Experiment 1 by: (1) developing a more distinctive set of pseudowords that would occupy a more sparse phonological space in the hopes of encouraging stronger memory traces. Furthermore we (2) modified the training procedure to improve performance on the training phase, again with the goal of strengthening representations of newly learned items. We (3) include two measures of learning: a reading aloud task and a spoken-to-written word matching task. In the matching task participants hear a spoken target and a see a written probe before pressing a button to indicate whether spoken and written words match. While the reading task is relatively slow, with participants constructing and articulating words, the goal of the matching task is to provide a more speeded measure that is not reliant on overt production. The matching task also allows a measure of how well participants can generalise their knowledge to a novel task.

Finally, one of the primary goals was to clearly identify whether lexical consolidation effects would emerge after just one night. In order to increase the sensitivity of the experiment to any lexical consolidation we (4) included a variety of measures that might capture improvements due to lexical learning. These measures include a repetition priming measure on the reading aloud task, as well as an onset/coda error detection measure and a manipulation of visual familiarity for trained words on the matching task. These measures are discussed in detail below.

4.3 Methods

4.3.1 Participants

21 native English-speaking adults aged 18 – 40 took part in a study approved by the Cambridge Psychology Research Ethics Committee. No participants reported having dyslexia, speech, or language impairment. Three participants were excluded from the analysis for failing to learn (below 30% correct on the reading task) leaving 19 participants in the analysis.

4.3.2 Materials

Four sets of 16 monosyllabic consonant-vowel-consonant (CVC) pseudowords were constructed to allow for a trained and an untrained set for each of days 1 and 2 (see Figure 4.1 for sample stimulus assignment). Two sets of the pseudowords were constructed using eight phonemes (b, f, n, v, s, θ, ʤ, k) and four vowels (æ, ɛ, ɒ, ʌ). The two remaining stimulus sets were constructed using eight different phonemes (d, m, p, z, t, ʃ, h, tʃ) but shared the same 4 vowels (æ, ɛ, ɒ, ʌ). Three of these phonemes mapped to digraphs in English orthography (/tʃ/ $\mapsto \langle ch \rangle$, /ʃ/ $\mapsto \langle sh \rangle$, and /θ/ $\mapsto \langle th \rangle$) but were included to allow for a wider item set containing more distinctive lexical items. For simplicity we refer to these as either 'vowels' or 'consonants'.

An outstanding issue from Experiment 1 was the possibility that the lexical items were too easily confusable. It may be the case that in order for consolidation to take place distinctive representations are required. The lexical items are thus much more distinctive compared to Experiment 1 despite the frequency of phonemes remaining balanced across the set (note that many of the spaces in Figure 4.1 remain empty). Spoken forms of the pseudowords were recorded by a female native English speaker in a soundproof booth and digitised at a sampling rate of $44.1kHz$.

Each of the four sets was assigned to trained or untrained conditions on either day 1 or 2 in a counterbalanced manner across participants. As with the previous experiment the consonant frequency was matched across stimulus groups with each consonant appearing twice in onset and twice in coda. Unfamiliar visual symbols were mapped to each of the

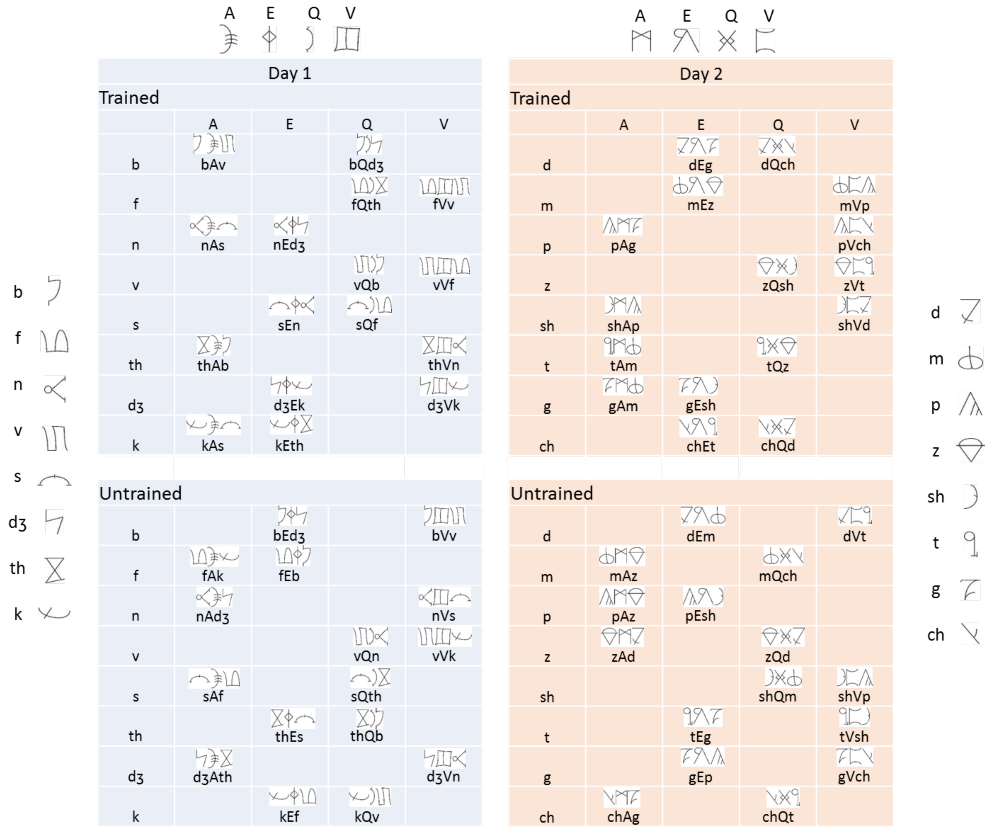


Fig. 4.1 An example assignment of the 64 pseudowords to be trained or untrained on day 1 or 2. 16 items are trained each day, retaining a further 16 untrained. The rows show the initial phoneme while the columns show the vowels. The vowel graphemes learned each day are shown at the top while the ’consonant’ graphemes are to the sides. Each ‘consonant’ appears twice in onset position and twice and coda position, while each vowel appears four times in each set. The four stimulus sets were rotated so that each appeared as trained/untrained and day 1/day 2 for different participants.

phonemes in a one-to-one manner (see Figure 4.1 for examples). The vowels were mapped to two visual symbols, one for each day.

4.3.3 Procedure

A full timeline of the experiment can be found in Figure 4.2 (p88). All tasks were presented using custom scripts in Psychopy, a python-based software package for psychophysics (Peirce, 2007, 2009).

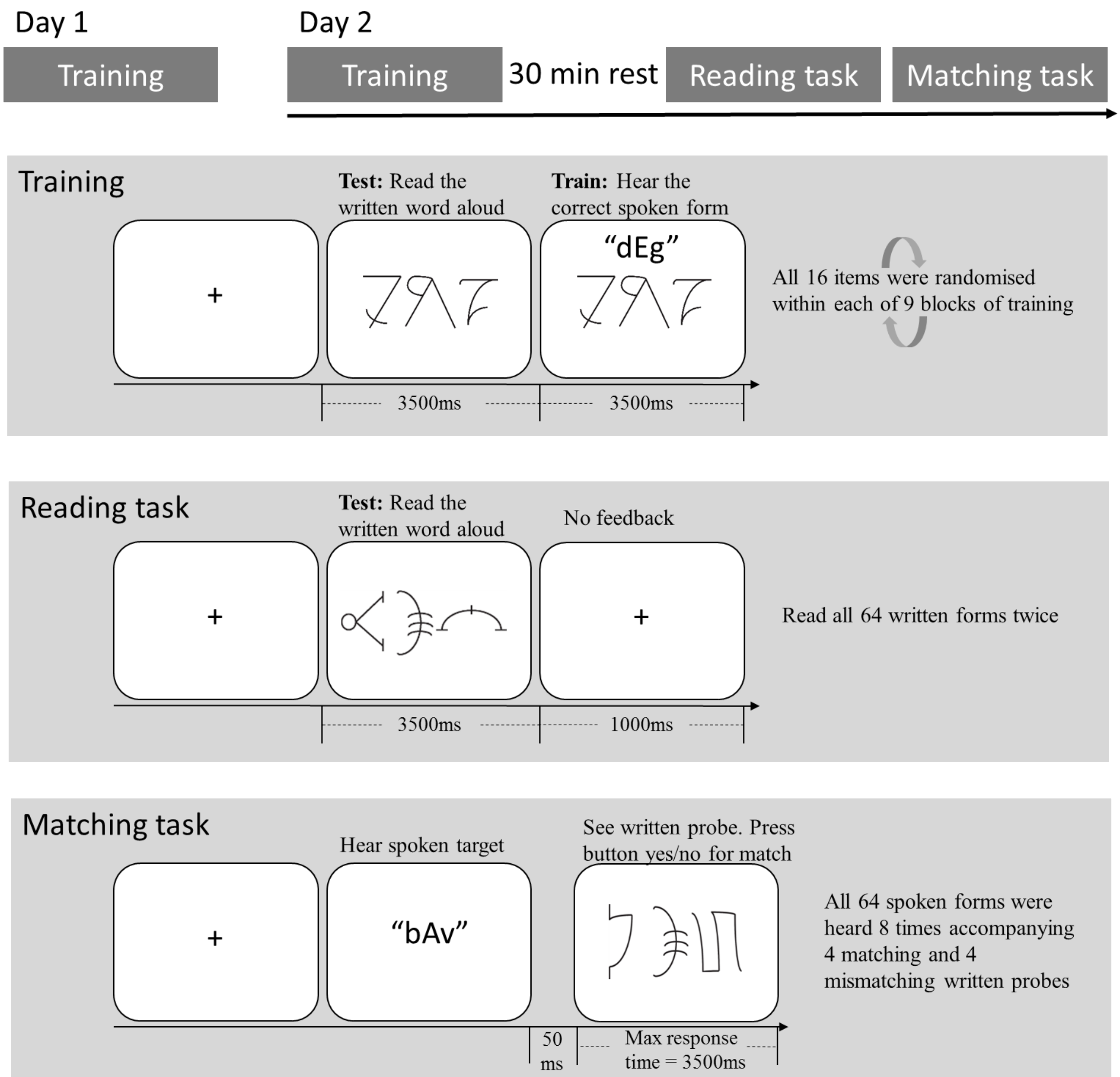


Fig. 4.2 (A) Timing of training and testing procedures over the two days of the study. (B) Timeline of a training trial. Participants are asked to read the item aloud before hearing the correct spoken form. Participants then have 3500ms to learn before the next trial. The 16 items to be trained were randomised within each run. (C) Timeline of a reading test trial. Participants were tested using a similar procedure to training, but did not hear the correct form of the written word and were given no feedback. (D) Timeline of a matching trial. The spoken target word was rapidly followed by a written probe onscreen. Participants had up to 3500ms to indicate whether the spoken target and written probes matched by pressing a button.

Training

Training took place over two days with 16 spoken pseudowords being associated with written forms on each day. The pseudowords were written in distinct orthographies each day and so participants learned a day 1 orthography and a day 2 orthography. During training participants were first familiarised with the items by seeing a block in which each item appeared on screen while they heard the spoken form. The trial structure combined training and testing to allow a continuous measure of learning (Figure 4.2B, p88). During each trial the written form appeared onscreen for 3500*ms* during which participants were instructed to read the item aloud. Participants then heard the correct spoken form of the item and had a further 3500*ms* to learn during which the written form remained onscreen. There was a 1000*ms* gap between trials, thus eight seconds per trial. Each item had nine presentations, randomised over three blocks. Responses were recorded and scored offline and were deemed correct if participants correctly reported all three phonemes in the target item.

Testing

Following training on day 2 participants had a 30 minute rest break. Immediately before the testing phase participants completed a reminder task in which they saw and heard all items from both days in a randomised order. This reminder minimised forgetting effects for items from day 1 while maintaining equivalent exposure to all items.

Reading task. Participants read aloud items from all four experimental conditions over two blocks: trained day 1, trained day 2, day 1 generalisation, day 2 generalisation. Items from the four conditions appeared in a fully randomised manner. Each item appeared twice (thus 128 responses in total) with a gap of 1 - 7 items between the first and second presentations, with an average of $\approx$ 15 seconds between presentations. In each trial the written form of the item appeared onscreen and participants had 3500ms to respond. No feedback was given. Responses were recorded with accuracy and response time marked offline.

A repetition priming manipulation was included in the reading task with the goal of exploring whether consolidated words might show a different pattern of priming from unconsolidated words. Words and pseudowords may show contrasting repetition priming effects; Forster and Davis (1984) found faster responses to repeated high- and low-frequency real words (45*ms* and 38*ms* respectively) but slowed responses to repeated nonwords (-8*ms*). Therefore we reason that consolidated words may show a more word-like pattern while unconsolidated words would show a pattern typical of pseudowords. Repetition priming has also been used in a variety of settings to probe the nature of memory for novel stimuli (Henson et al., 2000; Henson, 2001; Scarborough et al., 1977; Stark and McClelland, 2000; Zeelenberg et al., 2004). Here we are interested primarily in repetition priming as an index of overnight changes in the representations of the trained words.

Matching task. On each trial of the matching task participants heard the spoken form of one of the 64 items (trained/untrained, day1/day2), with a visual probe appearing 50*ms* after the end of the spoken item. Participants pressed a button to indicate whether the spoken target matched the written probe. Response times were measured from the onset of the written probe and to ensure speeded responses an upper limit of 3500*ms* was imposed. All 64 pseudowords were used as spoken targets. For each spoken target the written probe could appear in one of six conditions, two of which matched the spoken item and four of which mismatched (see Figure 4.3, page 91). The two matching conditions appeared twice each to ensure that half of the items are matching, half mismatching. In total then there were eight trials for each of the 64 items leading to 512 responses per participant.

We include two conditions in the matching task where the target and probe are congruent, i.e., when the spoken and written forms match. The first of these matching conditions is simply the written form of the spoken word, written using orthography consistent with the training (top row of Figure 4.3, p91). However, it was possible that participants would adopt the strategy of judging written probes by whether they were visually familiar or unfamiliar, rather than translating from print to sound. To avoid this potential situation we included a second matching condition in which the vowel characters were swapped, i.e., the vowel character for ‘A’ learned on day 1 was swapped with the character for ‘A’ learned on day 2 (shown on the second row of Figure 4.3, p91). In this way the probes remained visually unfamiliar at the whole word level and yet required participants to endorse the match. An equivalent substitution took place for the mismatching probes, where the incorrect character could be from the other character set.

We additionally include a comparison to detect changes in reading strategy from serial decoding to holistic word recognition. When reading unfamiliar words decoding of letters takes place in a serial manner (Coltheart et al., 2001; Rastle and Coltheart, 2000; Weekes, 1997). For familiar words however the whole word form can be recognised. One indicator of this difference in reading strategy is the length effect found for familiar and unfamiliar words. For unfamiliar words the time needed to read the word aloud increases with the number of letters in the word (Weekes, 1997). For familiar words by contrast the length effect is less pronounced. The extent of the letter length effect decreases as readers become more familiar

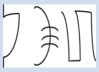
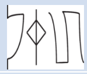
Target	Match or mismatch	Orthographic consistency	Error type	Example
bAv 	Match	Consistent		
		Mixed		
	Mismatch	Consistent	Vowel error	
		Mixed	Vowel error	
		Consistent	Onset error	
		Consistent	Coda error	

Fig. 4.3 Examples of the six possible probe types that could appear for each of the 64 targets. Participants heard the spoken target (left) before seeing the written probes (right). Matching probes appeared twice as frequently as mismatching probes, ensuring an equal probability of that target and probe would match. Each of the 64 spoken words shown in Figure 4.1, p 87 was presented eight times during the matching task accompanied by written probes corresponding to these six probe types.

with words (Zoccolotti et al., 2005) and so the shift from serial to holistic processing may be a useful marker of written lexical knowledge (van den Boer and de Jong, 2015).

Shifts from serial to holistic strategies in reading have recently also been observed in the context of an artificial orthography (Takashima et al., 2016). Here, participants learned to read a novel orthography on days 1 and 5 of the study, along with a further testing session on day 30. To assess whether whole-word recognition emerged participants read trained words, completely novel words, and words made up of recombined syllables. The trained words were recognised both faster and more accurately on day 5, suggesting participants have learned the whole word forms.

In the current experiment we do not manipulate letter length to assess whether participants are adopting serial or holistic reading strategies. Instead we explore the speed at which participants can detect errors in written forms on a spoken-to-written matching task. We include trials in which the spoken and written forms do not match, manipulating whether the mismatching letter occurs at the beginning or the end of the written word. For day 2 words where we expect participants to read in a serial manner, errors in onset should be detected faster than errors in coda position due. If consolidation of trained words from day 1 has led to formation of visual lexical forms then it may be that errors in onset and coda are detected in a similar time.

4.4 Results

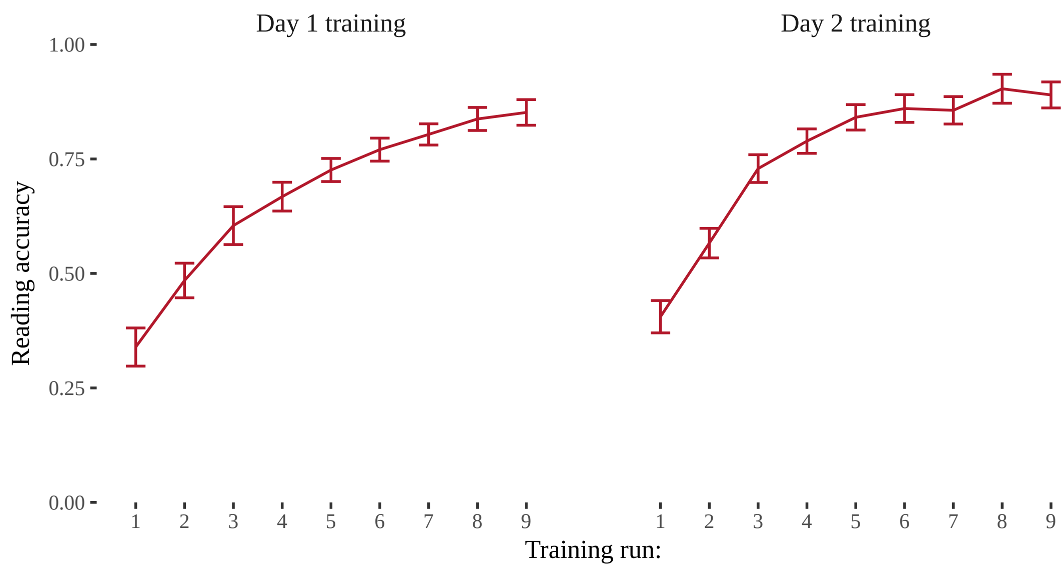


Fig. 4.4 Improvements in reading performance during training.

Analysis approach. Unless otherwise mentioned all behavioural results were analysed using generalised linear mixed effects models (GLMM) using the `lme4` package (version 1.1-12; Bates et al., 2015b) in R (version 3.2.3; R Core Team, 2013). GLMMs have become an important statistical tool as they avoid problems of data-aggregation and allow simultaneous control of the variance associated with both participants and items (Baayen et al., 2008; Judd et al., 2012; Pinheiro and Bates, 2000). In addition these models are extensions of the generalised linear model and so are capable of incorporating link functions to model data that is non-normally distributed. In the behaviour analyses that follow we use a logit link function for accuracy data (Jaeger, 2008) and an inverse gaussian function for response time data (Baayen and Milin, 2010).

Model specification followed the recommendations of Barr et al. (2013). To reduce Type I error we first attempt to fit the maximal random effects structure with random slopes and intercepts for all fixed effects. However, the information contained in the data may not be sufficient to support the large number of components in a maximal model in each case (Bates et al., 2015a). In analyses where this model did not converge we removed from the model in order: (1) correlations between random effects, (2) by-item random effects, (3) by-subject random effects, (4) the random effect that captured the least amount of variance (Matuschek et al., 2017). Once the maximal model that would converge had been identified likelihood ratio tests were used to assess main effects and interactions (Matuschek et al., 2017). In each case a reduced model was constructed that did not contain the effect to be tested and was compared against the full model. Where the likelihood ratio test indicated a significant effect we additionally report the significance of the individual model coefficients using the z- or t-statistic in the model summary.

4.4.1 Training task

Learning to read the orthographies from both days showed similar performance with participants improving from 33.93% ($SD = 47.43$) in the first run of day 1 to 85.15% ($SD = 35.60$) on the final run (Figure 4.4, p92). Accuracy was slightly higher on day 2 both at the beginning of training, ($M = 40.55\%, SD = 49.18$) and on the final run ($M = 88.99\%, SD = 31.35$).

4.4.2 Reading task

Accuracy. Accuracy during the reading task was high across all four conditions (day 1 trained: 86.24%, day 1 generalisation: 84.04%, day 2 trained: 91.49%, day 2 untrained: 88.41%) (Figure 4.5, left panel, p94; Table 4.1, p94). There was no main effect of day, $[\chi^2(1) = 2.40,\ p = 0.12]$, or training, $[\chi^2(1) = 0.01,\ p = 0.91]$, and no interaction of day and training, $[\chi^2(1) = 1.21,\ p = 0.27]$ on accuracy.

Reading latencies. Incorrect responses were discarded and further analyses include only the correct responses. Although figures show untransformed results for ease of interpretation the GLMMs for all response time data included inverse-gaussian link functions to account for non-normal distributions. Trained items from both days (day 1 trained: 2193ms, day 2 trained: 2264ms) were faster than untrained items (day 1 untrained: 2434ms, day 2 untrained: 2489ms) (Figure 4.5, right panel, p94; Table 4.1, p94). Comparing reading latency for the four main conditions revealed a main effect of training, $[\beta = 208.61,\ SE = 26.67,\ t = 7.82,\ p < 0.001,\ \chi^2(1) = 20.00,\ p < 0.001]$. The main effect of day did not reach significance, $[\chi^2(1) = 0.34,\ p = 0.56]$, nor did the interaction of day and training, $[\chi^2(1) = 0.98,\ p = 0.32]$.

Repetition priming. To assess repetition priming effects response times to the second presentation were subtracted from the first (i.e. positive values indicate faster responses to the second presentation). This was done on an item-wise basis, i.e., $RT1 - RT2$ for each item. Items were only analysed if responses to both presentations were correct, 84.14% of trials were retained. The average repetition priming effect across all conditions was

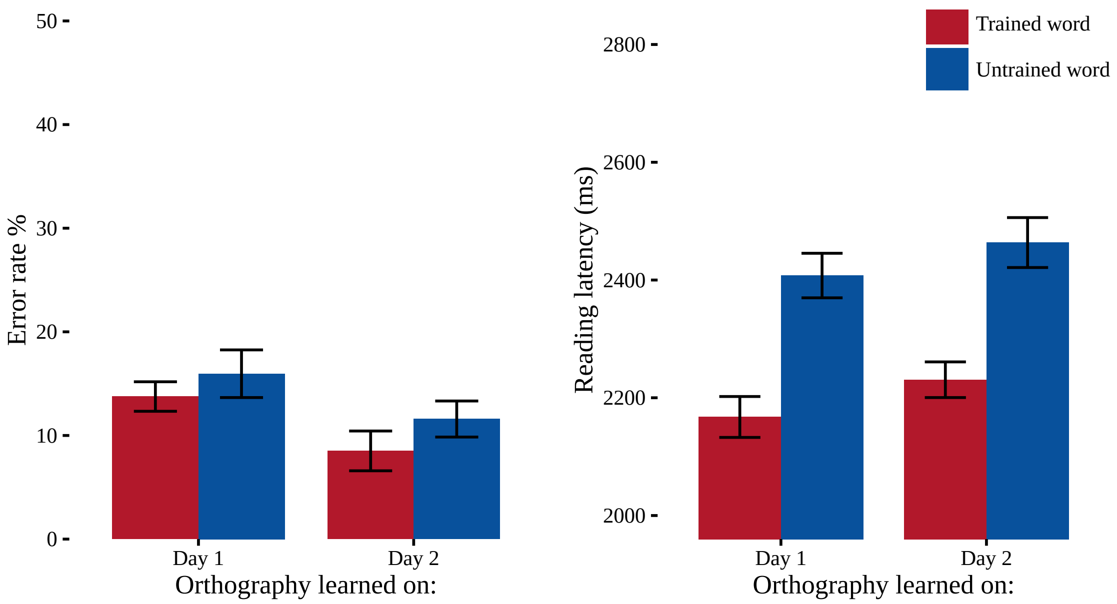


Fig. 4.5 Performance on the reading task. The left panel shows error rates for trained and untrained words from days 1 and 2 on the reading task. The right panel shows reading latency for the same.

Table 4.1 Experiment 2: Performance on reading task

	Accuracy (%)		Reading latency (ms)		Repetition effect (ms)	
	Mean	St. Dev	Mean	St. Dev	Mean	St. Dev
Day 1 trained	86.24	34.47	2167	661	369	623
Day 1 untrained	84.04	36.65	2408	777	454	714
Day 2 trained	91.49	27.93	2231	754	299	693
Day 2 untrained	88.41	32.03	2464	789	375	739

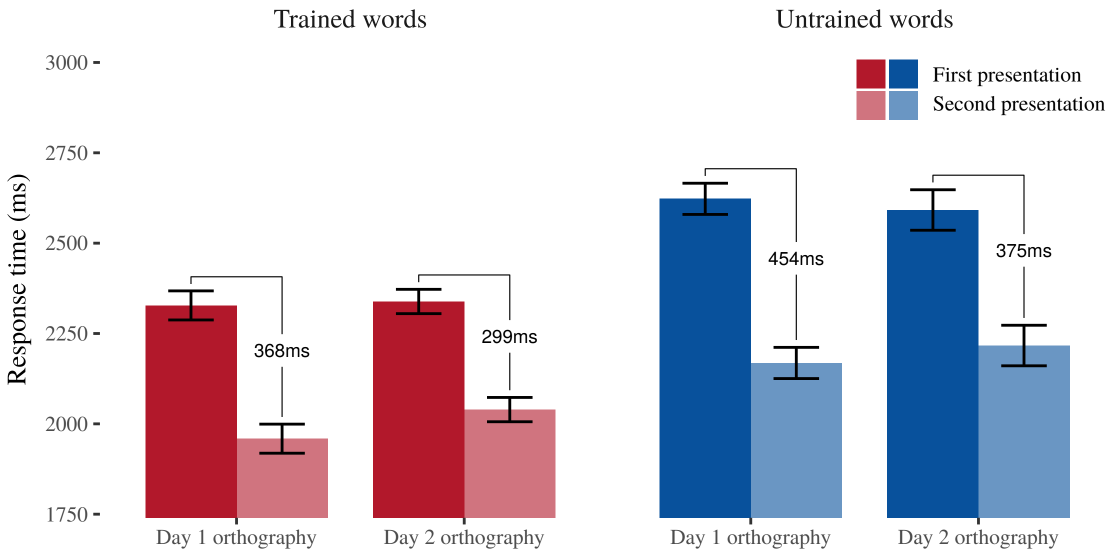


Fig. 4.6 Repetition priming on the reading task: reading times for the first and second presentations of trained and untrained words made up of orthography learned on day 1 or day 2.

372.61*ms*, indicating a benefit of some form to seeing the same item after a short delay. A 2 x 2 repeated measures ANOVA compared the impact of day and training factors on the dependent variable of the magnitude of the priming effect. Repetition priming effects were equally strong in all four conditions, with no main effect of training, $F(1,17) = 3.35,\ ns$, or of day, $F(1,17) = 2.10,\ ns$, as well as no interaction, $F(1,17) = 0.36,\ ns$. (Figure 4.6, p95; Table 4.1, p94). Thus there was no indication of newly learned lexical representations impacting on reading speed.

4.4.3 Matching task

Accuracy. For the primary analysis of matching data the four target conditions (day 1 trained, day 1 untrained, day 2 trained, day 2 untrained) were the primary interest and so we collapsed the six probe conditions. Accuracy was again high in the matching task (day 1 trained: 92.93%, day 1 generalisation: 91.11%, day 2 trained: 91.86%, day 2 untrained: 91.91%) and did not differ across the four conditions. There was no significant difference in accuracy between trained and untrained words on the matching task, $[\chi^2(1) = 1.73,\ p = 0.19]$, nor were participants more accurate for words trained on days 1 or 2, $[\chi^2(1) = 0.31,\ p = 0.58]$. The interaction of day and training was also not significant, $[\chi^2(1) = 0.55,\ p = 0.46]$.

It is possible that correctly rejecting a mismatch involves different cognitive processes than correctly endorsing a match. Moreover, while the experimental conditions 'trained' and 'untrained' accurately describe the spoken target, the status of a mismatching written probe is not clear. Therefore a further analysis excluded mismatching trials and compared accuracy for trained and untrained words from days 1 and 2 on the matching trials. The results largely mirror the outcome of the main analysis. There was no main effect of either training, $[\chi^2(1) = 0.11,\ p = 0.74]$, or day, $[\chi^2(1) = 0.66,\ p = 0.42]$, on accuracy. The interaction approached, but did not reach significance, $[\chi^2(1) = 3.14,\ p = 0.076]$.

Response times. The main analysis of matching trials compared the impact of day of learning (the orthography) and whether the word was trained or untrained on response times. Following the rationale above mismatching trials were excluded from the primary analysis and analysed separately. In addition, mixed orthography trials, and incorrect responses were excluded and response times for the remaining data were analysed (see Figure 4.7, p97). Participants responded significantly faster to words written in the orthography learned on day 1, $[\beta = 90.59,\ SE = 20.35,\ t = 4.45,\ p < 0.001,\ \chi^2(1) = 4.21,\ p < 0.04]$, indicating overnight consolidation at a sublexical level. Participants also responded faster to trained words compared to untrained words, $[\beta = 68.59,\ SE = 15.66,\ t = 4.38,\ p < 0.001,\ \chi^2(1) = 10.06,\ p = 0.002]$. The interaction between day and training that would indicate overnight consolidation of trained words from day 1 was not significant however, $[\chi^2(1) = 1.85,\ p = 0.17]$ (Figure 4.8 p98, left panel).

In order to measure whether participants have learned a visual lexical representation of the trained words we compared responses to words written in consistent and mixed orthographies. As the mixed orthography conditions cannot be assigned to day (i.e., each word contains letters from days 1 and 2) we collapse across the factor of day. The matching probes (i.e. the letters of the written probe matched the phonemes of the spoken target) contained both visually familiar and visually unfamiliar written probes (see Figure 4.7, p97). By alternating the vowel sets used on different days we formed probes in consistent- and mixed-orthography conditions. These mixed-orthography probes while spelled correctly, have never been seen before. Note however that 'visual familiarity' could only apply to the trained items. Thus if participants have learned to recognise the visual forms of the trained words, we should see slower responses to the mixed-orthography probes. By contrast for the untrained words participants have seen neither the consistent- or mixed-orthography probes and so we expected to find no differences between these conditions.

A repeated measures ANOVA of the matching trials compared training (trained or untrained words) and orthographic consistency (consistent or mixed). There were main

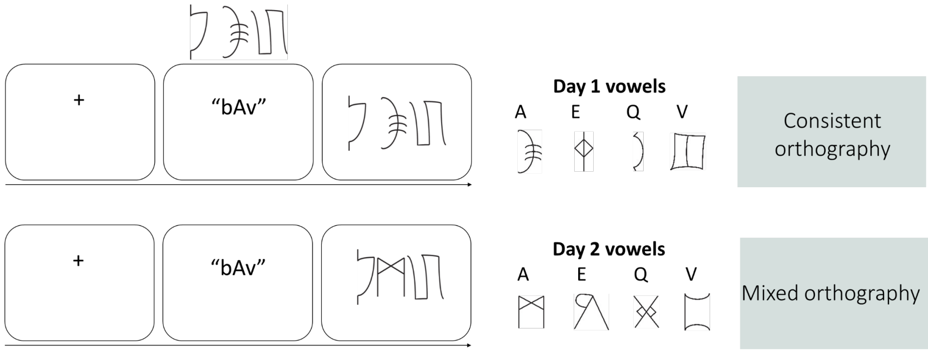


Fig. 4.7 Comparison of 'consistent' and 'mixed' orthography probe items from the matching task. In this example participants have been trained on the word 'bAv' and so are familiar with its phonological form. Both written probes match the spoken form, though visually different. If participants have additionally formed a visual representation of the trained word 'bAv' then they should recognise the written form on the top row. The bottom row shows a probe where a vowel character from the alternate vowel set is used. While the probe on the bottom row correctly spells 'bAv', its visual form is now unfamiliar. Including visually unfamiliar probes that nevertheless matched the spoken target meant that participants were not able to make judgements purely on the basis of visual familiarity with the written probe.

effects of orthographic consistency, $[\beta = 95.29,\ SE = 15.43,\ t = 6.18,\ p < 0.001,\ \chi^2(1) = 12.32,\ p < 0.001]$. The main effect of training did not reach significance, $[\chi^2(1) = 1.34,\ p = 0.25]$, but there was a significant interaction between training and orthographic consistency that would indicate visual familiarity with trained words, $[\beta = -47.43,\ SE = 16.87,\ t = -2.81,\ p < 0.005,\ \chi^2(1) = 4.53,\ p < 0.03]$. Follow-up Tukey's tests were conducted using the `glht` function from the `multcomp` package for R (version1.4-6; Hothorn et al., 2008). Crucially, these comparisons confirmed that the interaction was driven by the difference between trained words written in either consistent or mixed orthographies, $[z = -5.673,\ p < 0.001]$. Meanwhile, there was no difference between trained and untrained words in the mixed orthography condition, $[z = 0.51,\ p = 0.95]$. The slowed responses to visually unfamiliar trained words thus indicates that participants have learned something of the trained words beyond merely decoding letter-by-letter. This knowledge might reflect visual whole-word recognition or familiarity with the bigrams that make up the trained words.

Serial or holistic word recognition. A final subset of probes allowed us an additional test of whether words were still being read in a serial manner, or whether some whole-word recognition was taking place. To this end we analysed response times to a subset of the probes: mismatching probes where the error occurred in either onset or coda position (i.e, the final two probes in Figure 4.3, p91). Note that the assignment of items to 'trained' or 'untrained' conditions refers here to the spoken target. In all cases the written probe is

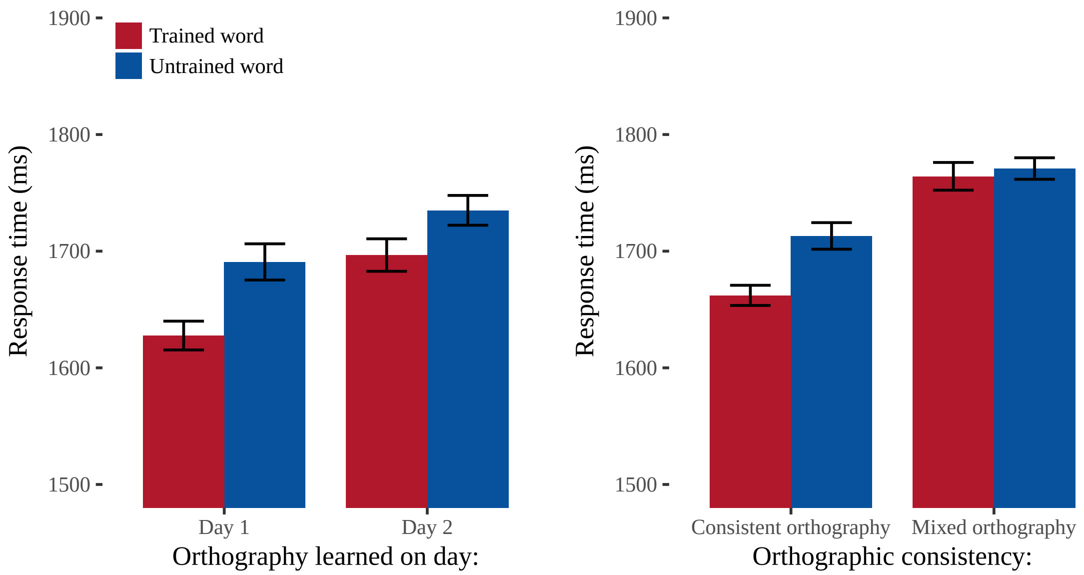


Fig. 4.8 Performance on the matching task. The left panel shows the response times to correctly endorse matches for trained and untrained words from days 1 and 2. Note that mixed orthography probes are excluded from this comparison as they cannot be assigned to either day. The right panel shows response times to trained and untrained words written using consistent or mixed orthographies. In the case of trained words the consistent orthography condition has been seen several times during training, while the mixed orthography condition remains visually unfamiliar.

unfamiliar. The question in this analysis is whether there will be a difference in the length of time required to identify and correctly reject an error at the beginning or end of a word.

Probe conditions were tested in light of the spoken target items to which they were paired, resulting in a GLMM with the factors day of orthography (day 1 or day 2), training (trained or untrained spoken target), and probe error position (onset or coda). As expected, there was a large main effect of error position, [$\beta = -480.81,\ SE = 26.12,\ t = 18.41,\ p < 0.001,\ \chi^2(1) = 455.67,\ p < 0.001$], indicating participants identified errors in onset position faster than errors in coda position. A significant main effect of day, [$\beta = 112.70,\ SE = 30.30,\ t = 3.72,\ p < 0.001,\ \chi^2(1) = 12.67,\ p < 0.001$], showed participants were also faster to identify errors when the written probe was in the day 1 orthography, thus providing additional evidence for overnight consolidation at the sublexical level. There was also a main effect of training, [$\beta = 78.93,\ SE = 25.84,\ t = 3.05,\ p = 0.002,\ \chi^2(1) = 12.00,\ p < 0.001$], indicating participants were faster to identify errors in trained words. The critical interaction between day, training, and error position was not significant however, [$\chi^2(1) = 0.024,\ p = 0.88$]. Thus, we cannot infer that participants have shifted to a whole word reading strategy for trained words after a period of offline consolidation.

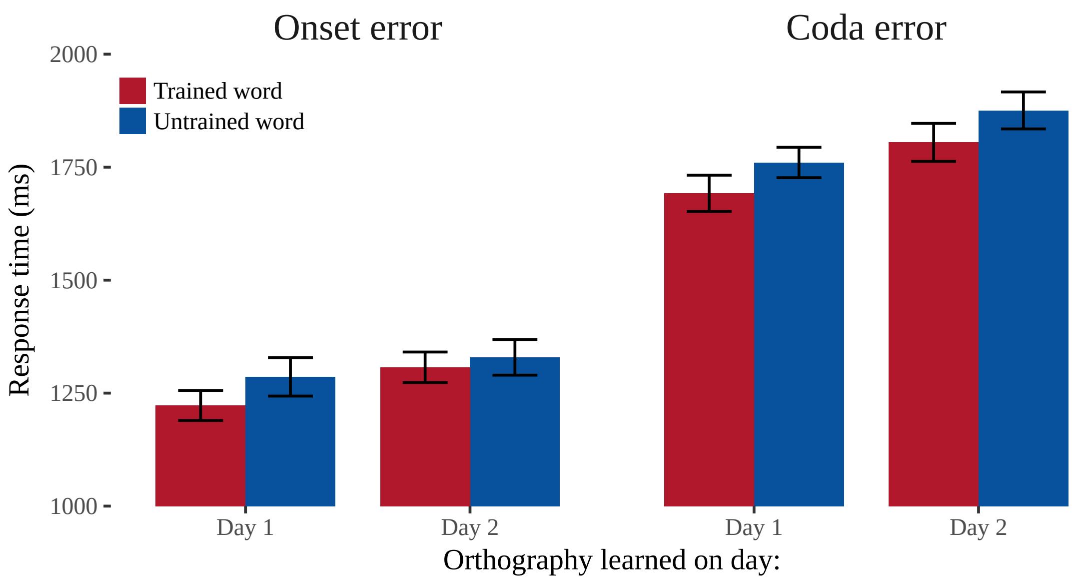


Fig. 4.9 Detecting mismatches according to error position. The figure shows the time to correctly detect an error in onset or coda position according to the factors of day and training conditions. If participants are reading in a serial manner then responses should be faster to onset errors. If participants have learned to recognise trained whole-word forms then the difference in response times between onset and coda errors should be reduced. Untrained words have to be read in a serial manner

4.5 Discussion

The primary goal of this experiment was to assess the impact of overnight consolidation on lexical and sublexical learning. In summary the results show clear effects of sublexical orthographic consolidation but demonstrate little evidence of whole-word consolidation. Accuracy for reading aloud both trained and untrained items was high (89% and 86% respectively), showing that participants could generalise orthographic learning. Reading latencies were faster for trained words versus untrained words, but did not differ for the orthography learned on day 1 as compared to day 2. Meanwhile, a spoken-to-written word matching task showed that probes written in the consolidated day 1 orthography were identified faster than those written in the day 2 orthography. The beneficial impact of consolidation did not interact with whether the items were trained or untrained however; the difference between trained and untrained words from day 1 was not greater than the difference between trained and untrained words from day 2. Nevertheless, improved performance for trained items demonstrates that participants have learned representations for whole items. Overnight consolidation therefore seems to impact at the level of sublexical knowledge rather than the level of lexical knowledge.

4.5.1 Sublexical overnight consolidation

The most significant finding of this experiment was the large effect of day-of-learning on response times in the matching task where participants are faster to read written probes for both trained and untrained words from day 1. The improvement in response times for the day 1 orthography applied both when endorsing matches for trained and untrained items and also when rejecting errors in trained and untrained items. These results suggest overnight consolidation at the sublexical level of orthography.

Given the number of studies showing improvements in visual discrimination with overnight consolidation, one interpretation of these results may be that participants are not demonstrating improvements in visual-verbal mappings, but rather improved visual discrimination skills. These findings are consistent with evidence of overnight improvements in visual discrimination but we know of no other studies showing evidence of overnight consolidation in the context of learning the visual-verbal associations that underlie reading. The inclusion of both matching/mismatching and alternative orthographies suggest these differences cannot be purely visual however as participants must decode the spoken form of the word in order to identify it, rather than relying on visual discrimination. In the context of tracking the time course of emergent literacy this is an important finding; learning these most basic units of literacy must be in place to support later learning.

Task demands and generalisation. There is a striking difference between the impact of consolidation on the reading and matching tasks; participants are much faster for day 1 orthography on the matching task but there are no day related changes on the reading task.

One explanation for this contrast may be that the matching task is speeded and is thus more sensitive to overnight changes. Indeed it was this hypothesis that prompted the inclusion of a matching task. However, an alternative reason for the difference between tasks may be that the tasks differed in their verbal output demands. Decoding and articulating each letter within a word in order to successfully read it aloud may be more burdensome than performing the visual matching task. The very large differences in latency to read trained and untrained words do suggest that articulating an unfamiliar word is difficult. Moreover the reading task itself is challenging and consequently there is a large amount of variance in the data. Trained words from day 1 are read overall 71ms faster than trained words on day 2 but this difference is non-significant. The variability of responses to this new and challenging task may therefore mask an effect.

Another possible interpretation for the differences between tasks has to do with generalisation. The reading task used to test learning was the same as used to train the written words. In a matching task that did not require verbal output and differed from the training task, there were faster responses to letters learned on day 1, whether these letters were part of a trained word or untrained word. This outcome may suggest consolidation is less important if training and testing tasks are the same; participants are familiar with the reading task and so do not need to generalise to a novel situation.

4.5.2 Multiple measures of whole-word learning

One major question of this experiment was whether orthographic training could encourage overnight consolidation of lexical representations, and whether these representations are phonological or visual (orthographic) in nature. In addition to analysing the simple interaction of day and training in the two tasks, the experiment included several nested measures that may be sensitive to lexical consolidation: repetition priming in reading task, onset/coda errors in matching task, and visually familiar or unfamiliar trained words.

No lexical consolidation when reading aloud. Performance on the reading task showed no effect of day or training on accuracy, replicating the pattern found in Experiment 1. Further, neither accuracy nor reading latency indicated improvements due to lexical consolidation, as would be indicated by improved performance for day 1 trained words. An important difference in the current results is the inclusion of response times. These data were not available in Experiment 1 due to the scanner task design. Reading latencies in the current data show very large effects of training, with trained words being read aloud $\approx$ 200ms faster than untrained words. This difference in performance may be attributed either to faster reading times or improved speed of articulation. In either case the outcome suggests something has been learned at the whole-word level, and that response times may be more sensitive to these effects than accuracy.

The reading task also contained an embedded repetition priming measure that was included to explore whether overnight changes in priming might occur. Participants read each word aloud twice, with a short, variable interval between presentations. The repetition effect was reliable in all cases (from 299ms for day 2 trained words to 454ms for day 1 untrained words) which shows that priming of some form is taking place (Figure 4.6, p95). The difference in magnitude of priming effect between the conditions was not significant however.

It may simply be the case that in the current experiment participants have not consolidated the words learned on day 1; the task is focused on orthographic learning and differs from typical lexical consolidation studies in a variety of ways. Another possibility is that the priming measure is not sensitive to changes in the representations of these recently learned words. The priming paradigm is typically used in cases where reading is automatic in nature. It is perhaps unsurprising that the slow and deliberate decoding of recently learned orthography does not show a pattern similar to reading familiar words. Overall this may indicate that reading these artificial items is not sufficiently automatic to support the priming effects seen in real words and pseudowords. Moreover, the priming effects do not take place in the more typical context of a masked priming task (Masson and Isaak, 1999) and thus we cannot claim that the large priming effects seen here are orthographic in nature. The task was included here in an exploratory manner in the hopes of determining the extent to which laboratory-based studies of reading can be used to approximate real reading acquisition. Consistent with the outcomes of Experiment 1 then, the reading aloud task does not suggest lexical consolidation has taken place.

No lexical consolidation on the matching task. The matching task was designed to allow a potentially more sensitive measure of lexical consolidation in the context of reading a novel orthography. The main effect of training demonstrates that participants have indeed learned something about the whole-word form; they are faster to recognise and endorse words they saw during training. However the interaction between day and training that would indicate consolidation of trained words from day 1 is not significant and thus the data do not point to lexical consolidation. The improvement in performance for trained items may be attributed to participants learning the phonological forms of the words, the written forms of the words, or a combination of both. The design used here does not distinguish between these possibilities.

This lack of evidence of whole-word learning is in contrast to the imaging results from Experiment 1, which found reduced hippocampal activation to written words trained on day 1. In that case as the same letters appeared for items from days 1 and 2, the differences could only be explained by whole word learning. Nevertheless the fMRI results in Experiment 1 were at odds with the behavioural outcomes where no whole-word learning was seen. It may be the case that even the matching task is not sensitive to these changes at the behavioural level.

No evidence for holistic reading stategies. Results from the reading task replicate the finding of improved performance reading trained words relative to untrained words seen in Takashima et al. (2016); participants are faster and more accurate for trained words on both the reading and matching tasks. Takashima et al., interpret this outcome as indicating a shift from serial decoding to whole-word recognition. However, the inclusion here of an error detection condition on the matching task suggests a more complex picture. If participants are reading letter-by-letter then they should be faster to identify errors on onset position than errors in coda position. By contrast, if they are reading in a holistic manner then the time needed to recognise onset/coda errors should be more similar. Response times when rejecting probes do not show the pattern expected if participants have shifted to holistic reading strategies since they were faster to reject items with an error in the onset position (Figure 4.9, p99). Faster response times to trained words may simply reflect faster serial processing, rather than a shift to holistic word recognition. Thus, the data shown here argue for a re-interpetation of the Takashima et al. (2016) findings as reflecting improved serial processing.

Across multiple measures therefore we see no evidence for the consolidation of written lexical forms. This outcome is contrast to the findings of Bowers et al. (2005). However, in the case of Bowers et al. the words were learned in a familiar orthography and so lexical learning may benefit from pre-existing knowledge of written forms, i.e., already knowing the word ‘banana’. For example, reading pseudowords that are similar to existing words are more easy to read than pseudowords that are orthographically very different (Grainger et al., 2003). Here by contrast, we aimed to assess whether representation of lexical forms could emerge overnight in a recently learned orthography.

One possibility given the data presented here is that whole word representations of novel words do not form overnight, at least in an artificial orthography. This interpretation is consistent with the outcomes of previous artificial orthography studies running over longer time frames (Mei et al., 2013a; Takashima et al., 2014b). However, in addition to this possibility it may also be the case that experimental factors mask possible lexical learning. In adapting this novel orthography paradigm to explore lexical consolidation there were a number of experimental decisions that may have impacted this outcome.

Another possible reason for the null finding in relation to lexical consolidation is that the items learned here are focused on the orthographic level. Participants are repeatedly exposed to sets of letters which they must read in the context of meaningless words. Although we increased the distinctiveness of the items, the task may have encouraged attention to orthographic more than lexical forms. Thus, by capturing the componential nature of written language in our stimuli and task we may have failed to include the vital idiosyncratic nature of spoken words.

Learning bigrams. The consistent and mixed orthography conditions were introduced in order to prevent participants from responding based on visual familiarity to the written probe, and to separate the impact of knowing the phonological form of a word from knowing its visual form. Because the mixed orthography condition cannot be assigned to day (as the words contain letters from both days) we collapsed across day and instead analysed the impact of encountering letters in unfamiliar bigram contexts.

The analysis compared response times to trained and untrained spoken words according to whether the written, and matching, probes included a consistent or mixed orthography (Figure 4.8, p98). Again, consistent orthography means the consonants from day 1 only appeared with the vowel from day 1, and vice versa for day 2 letters. Mixed orthography means that the day 1 consonants appeared in the context of the day 2 vowels (see Figure 4.3, p91 for examples). As hypothesised, participants responded faster to written probes with consistent orthography; that is, faster responses to probes in which consonants appeared in a familiar vowel context. Participants must therefore have learned something about the visual forms of the written probes; the mixed vowel letters mapped to the same sounds and had equivalent training. The only difference was the bigram context in which vowels and consonants appeared. Furthermore, the follow-up tests suggest that the interaction of training and orthographic consistency may be the result of visual lexical learning. In the case of untrained words whether the orthography was consistent or mixed did not influence response times. However, participants respond faster to trained words that are visually familiar (consistent) than to trained words that are visually unfamiliar (mixed).

One explanation for the slowed responses to the mixed orthography probes is that participants are already sensitive to the visual statistics of letter occurrences: they have learned the bigrams. In the mixed orthography condition the middle letter is alternated, thus removing any familiar bigrams and explaining slowed responses to trained and untrained words. That participants would become sensitive to the bigrams was not anticipated when designing the experiment yet is a finding consistent with other literature. Statistical learning of the transitional probabilities in a non-orthographic setting has been shown in a wide number of settings (e.g., Bertels et al., 2012; Kim et al., 2009). Indeed analogous findings from the visual learning literature are informative. Fiser and Aslin (2005) assessed visual statistical learning through a series of studies manipulating the relationships between parts and wholes. In this paradigm several visual elements presented within a grid of other elements could be combined to form an 'object'. After training participants preferred the structures they had learned to random combinations. Participants, however, did not learn the pairs of elements that could be decomposed from the 'object'. If an object was made up of three elements participants preferred the triple but did not learn the embedded pairs. The authors suggest there is a computational advantage in holistic processing due to the reduced number of features.

Notably, because we cannot separate the orthographic consistency by day we cannot know whether participants' ability to generalise their bigram knowledge when reading unfamiliar words has improved overnight. However there is reason to believe consolidation of sub-lexical structure might take place. Sleep consolidation has been shown to facilitate the acquisition of sub-lexical structure in spoken language. Gaskell et al. (2014) taught novel phonotactic constraints with and without a period of offline consolidation (a 90-minute nap). Participants repeated aloud syllable sequences (e.g., *haf kan sang gam*) that contained embedded structure (e.g., h only appears in onset). After napping a 2-AFC task was used to test participants' ability to generalise these constraints, showing that only the sleep group generalised the phonotactic constraints to novel items. It may be the case that the sub-lexical orthotactic constraints of written language may be similarly influenced by sleep consolidation. We return to this issue in Experiments 3 and 4.

4.6 Experiment 3

Experiment 3 attempts to replicate and extend the main finding of Experiment 2; that overnight consolidation improves response times for words written in a recently learned orthography. In light of the novelty of this finding, replication is vital before strong conclusions can be drawn. To this end, many parts of Experiment 3 remain unchanged. Participants again learn to read two distinct orthographies over the course of two days. Training follows the same approach as Experiment 2, using the same four sets of spoken words counterbalanced so that there is a trained and untrained word set for each day. The only modification to training is an increase in the number of training runs. Participants received an additional block of training (now four blocks rather than three previously) leading to 12 exposures to each of the trained words. In this manner, we hoped to improve learning and potentially maximise any consolidation effects.

Experiment 2 provided incidental evidence for bigram learning, by way of the interaction between training status and orthographic consistency (see Figure 4.8, right panel, p98). The experimental design, however, cannot address whether overnight changes in bigram representations has taken place. The primary change to Experiment 3 follows up on this issue by shifting focus from lexical learning to bigram learning and consolidation. To this end, we omit the repetition priming manipulation during the reading task and instead measure the ability of participants to read words that vary according to the familiarity of their constituent bigrams.

Bigrams may provide an interesting candidate for overnight consolidation as they are an intermediate form between letters and words. Further, they may provide an (imperfect) analogy to consolidation findings in spoken language. Using artificial morphological learning Tamminen et al. (2012), showed participants were faster to generalise affixes to novel stems after a period of sleep. Affixes in spoken language have similarities to the bigrams of interest here in the sense that they form an intermediate level of representation, larger than phonemes but smaller than words. If participants consolidate bigram information then it may be the case that context-independent representations will emerge, 'context' referring here to the word in which the bigram was learned. Notably however, there are substantial differences between the morphemes and the bigrams used here. Morphemes have meaning, while the bigrams used here do not. Consolidation of spoken words is sensitive to the provision of semantic information (Hawkins & Rastle, 2016; Takashima et al., 2014; Dumay et al., 2004) but it is not clear how consolidation of written words might differ. Additionally, the morphemes used by Tamminen et al. (2012) were whole syllables rather than the letter pairs used here.

An alternative way to frame bigram learning would be in terms of orthotactic learning. Gaskell et al. (2014) demonstrated consolidation in spoken language at the level of phonotactics, with participants becoming sensitive to the embedded statistical structure of spoken words after a period of sleep. Consolidation at the level of bigrams would be consistent with these data; sleep might support the learning and generalisation of statistical structure of letter pairs.

To investigate these questions the reading task was adjusted to test how well participants can generalise their knowledge of bigrams. For each set of 16 trained words, four complementary sets of untrained words were constructed. These untrained words were made up of bigrams that varied according to whether they had been seen during training or not. Thus words could be made up of entirely familiar bigrams while remaining unfamiliar as a whole word. We reasoned that if bigram learning preceded lexical learning then untrained words made up of familiar bigrams should show a similar pattern to trained words. Untrained words with unfamiliar bigrams meanwhile would be more difficult. There were thus three ways in which participants could translate from written text to sound on the reading task. Participants could decode letter-by-letter using their knowledge of graphemes. They could also recognise the whole word form in the case of the trained words. Participants could also make use of their knowledge of the bigrams within the words.

The matching task was also adjusted to assess bigram learning and consolidation. In Experiment 2 the matching task included eight possible probe conditions (see Figure 4.3 above, p91). Experiment 3 does not include visual probes written in an alternative orthography (where the letter 'A' from day-2 replaces the letter 'A' from day-1). Instead, the written probes are constructed in such a way as to manipulate the familiarity of the bigrams within each of the written probes, as in the reading task. In this way mismatching probes containing trained or untrained bigrams are presented alongside the spoken words, allowing us to test directly whether participants are learning and consolidating bigrams.

Finally, we include an alternative measure of lexical consolidation in the form of a recognition memory test. Improved recognition memory for consolidated linguistic information has been found across a variety of studies (Davis et al., 2009; Hawkins et al., 2015; Merkx et al., 2011). Here participants made judgements on the training status of both spoken and written words from days 1 and 2. By independently measuring whether written and spoken words are recognised after consolidation we are able to separate the impact of phonological and orthographic learning.

4.7 Methods

4.7.1 Participants

21 native English-speaking adults aged 18 – 40 took part in a study approved by the Cambridge Psychology Research Ethics Committee. No participants reported having dyslexia, speech, or language impairment. Three participants showed poor performance on the reading task and so were excluded from further analysis. One further participant did not complete the recognition memory task and so the remaining 17 participants are included in this analysis.

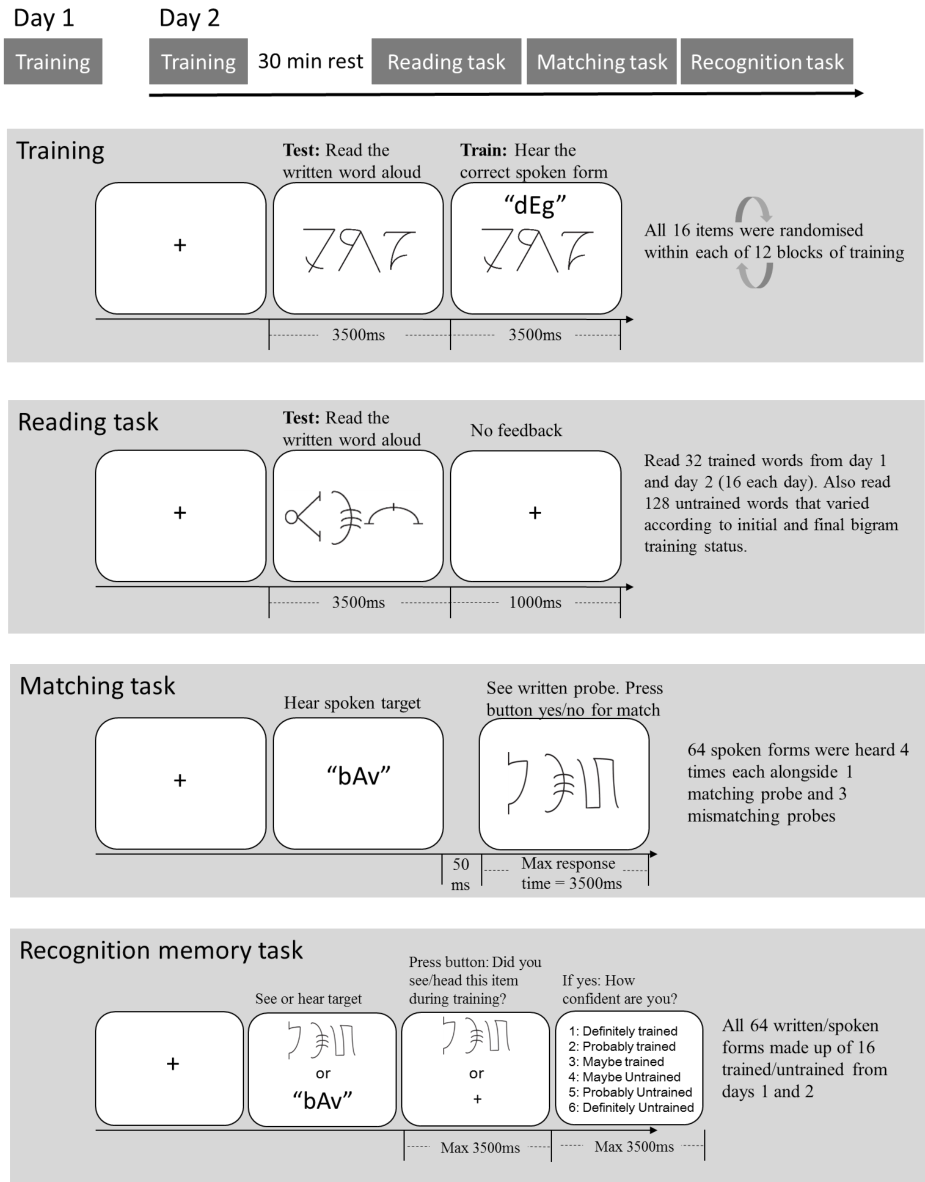


Fig. 4.10 (A) Sequence of training and testing procedures over the two days of the study. (B) Timeline of a training trial. Participants are asked to read the item aloud before hearing the correct spoken form. Participants then have 3500ms to learn before the next trial. The 16 items to be trained were randomised within each run. (C) Timeline of a reading test trial. Participants were tested using a similar procedure to training, but did not hear the correct form of the written word and were given no feedback. (D) Timeline of a matching trial. The spoken target word was rapidly followed by a written probe onscreen. Participants had up to 3500ms to indicate whether the spoken target and written probes matched by pressing a button. (E) Timeline for a recognition memory trial. 16 trained items from days 1 and 2 were presented alongside 32 untrained foils. Participants first completed one run making judgements on the written forms of the items followed by a second run making judgements on the spoken forms of the items. The item was presented and participants made a judgement about its training status. If they indicated the item had been trained then they made a confidence rating about this judgement.

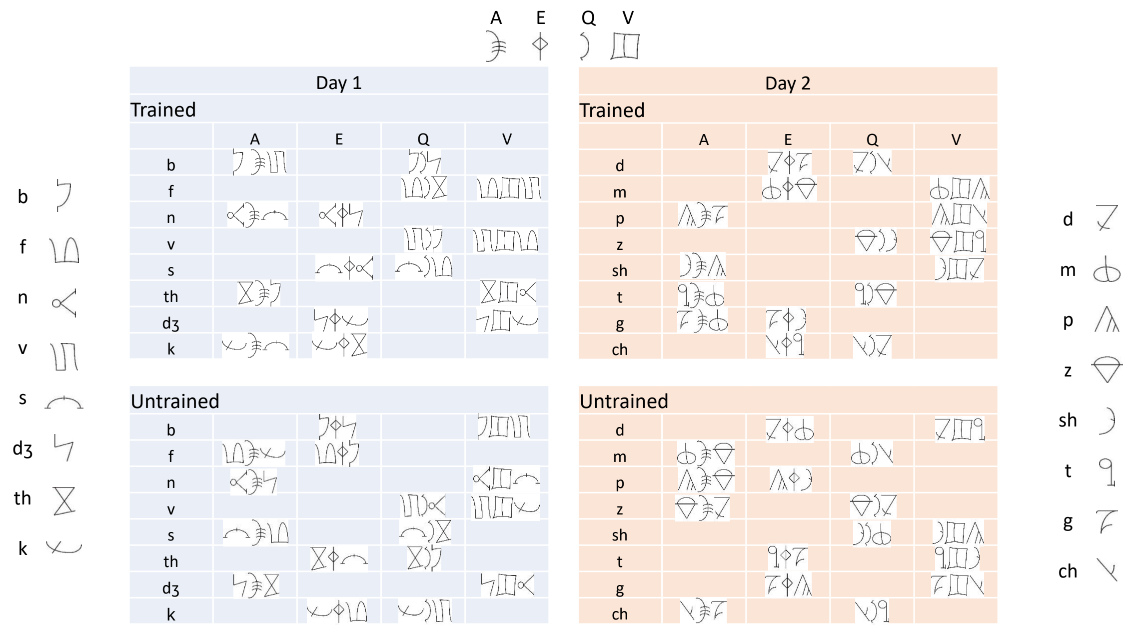


Fig. 4.11 An example assignment of the 64 pseudowords to be trained or untrained on day 1 or 2. 16 items are trained each day, retaining a further 16 untrained. The rows show the initial phoneme while the columns show the vowels. The vowel graphemes learned each day are shown at the top while the 'consonant' graphemes are to the sides. Each 'consonant' appears twice in onset position and twice and coda position, while each vowel appears 4 times in each set.

4.7.2 Materials

The same four sets of pseudoword materials from Experiment 2 are used for training items, along with the corresponding audio recordings and written forms (see Figure 4.11, p109). For each of the four pseudoword sets four probe conditions were also constructed. The words in these probe conditions were made up of onset and coda bigrams that were either (1) both learned during training, (2) trained onset bigram but untrained coda bigram, (3) untrained onset bigram but trained coda bigram, or (4) both bigrams untrained (Figure 4.12). For each of the spoken word sets, two sets of corresponding probes were constructed where both onset and coda bigrams were trained; one set differed from the trained words in the initial letter, the other set in the final letter. Thus for the matching task there were mismatching probes with errors either in onset or coda. Further, for items where both onset and coda bigrams were familiar the bigrams were chosen to ensure the whole word was not familiar. Thus we could achieve a situation where bigram familiarity could be manipulated independently from whole-word familiarity.

Set 1A — trained words: **bAv, nAs, thAb, kAs, nEdʒ, sEn, dʒEk, kEth, bQdʒ, fQth, vQb, sQf, fVv, vVf, thVn, dʒVk**

Initial bigram	Trained final bigram	Untrained final bigram
Trained initial bigram	bAs	bAf
	nAv	nAdʒ
	thAs	thAk
	kAv	kAth
	nEth	nEv
	sEk	sEf
	dʒEth	dʒEb
	kEdʒ	kEs
	bQth	bQn
	fQb	fQv
	vQdʒ	vQs
	sQth	sQn
	fVf	fVb
	vVk	vVs
	thVk	thVdʒ
	dʒVf	dʒVth
Untrained initial bigram	fAv	fAth
	vAb	vAf
	sAv	sAdʒ
	dʒAs	dʒAk
	bEdʒ	bEf
	fEth	fEv
	vEn	vEb
	thEk	thEs
	nQth	nQv
	thQdʒ	thQk
	dʒQf	dʒQs
	kQb	kQn
	bVv	bVth
	nVk	nVs
	sVf	sVdʒ
	kVn	kVth

Set 2A — trained words: **pAg, shAp, tAm, chQd, dEg, mEz, gEsh, chEt, dQch, zQsh, tQz, gAm, mVp, pVch, zVt, shVd**

Initial bigram	Trained final bigram	Untrained final bigram
Trained initial bigram	mAg	mAsh
	pAd	pAsh
	zAg	zAt
	chAz	chAm
	dEsh	dEz
	pEm	pEch
	tEp	tEz
	gEm	gEd
	mQt	mQz
	zQm	zQsh
	shQch	shQg
	chQd	chQg
	dVp	dVz
	shVch	shVm
	tVch	tVd
	gVp	gVd
Untrained initial bigram	dAz	dAp
	shAd	shAp
	tAd	tAch
	gAz	gAt
	mEp	mEz
	zEsh	zEt
	shEg	shEch
	chEm	chEd
	dQch	dQp
	pQm	pQg
	tQch	tQsh
	gQt	gQz
	mVp	mVz
	pVch	pVd
	zVsh	zVg
	chVt	chVd

Fig. 4.12 An example assignment of 16 trained words for days 1 and 2. The 16 bold words in each item set show the trained words that are learned each day. Accompanying each set of trained words are the four untrained item sets that were used on the reading task to assess the impact of bigram familiarity. The items on the top row contain familiar bigrams in onset, while items in the bottom row contain unfamiliar bigrams in onset. Items in the left column contain trained final bigrams, while in the right column the items contain untrained final bigrams. Consequently the item set in the top left contain entirely familiar bigrams but still remain unfamiliar whole words.

4.7.3 Procedure

A full timeline of the experiment can be found in Figure 4.10 (p108). All tasks were presented using custom scripts in Psychopy, a python-based software package for psychophysics (Peirce, 2007, 2009).

Training

As in Experiment 2 training took place over two days with 16 spoken pseudowords being learned in association with written forms on each day. Participants were again familiarised with the written and spoken forms of the 16 items briefly before training began. Training followed the same procedure as Experiment 2 (see Figure 4.10, p108) with participants trying to read aloud the written form before hearing the spoken form and having a chance to learn. Participants saw 12 presentations of each item over the course of 4 blocks of training, increased from 9 presentations in the previous experiment. Responses were recorded and scored offline and were marked correct only if the participant reported all three phonemes in a word correctly.

Testing

As before participants had a 30 minute rest break after training on day 2 before beginning the testing phase. After the break they completed a reminder block in which all of the words from both day 1 and day 2 appeared on screen once each in order to minimise forgetting effects.

Reading task. Participants read aloud 160 written word forms comprising 32 trained items (16 items each from days 1 and 2) and 32 items in each of the four untrained generalisation sets. In each trial the written item appeared onscreen for 4500ms and no feedback was given. Items appeared in a randomised order and responses were recorded and scored offline for accuracy and response time.

Matching task. As before for each trial of the matching task participants first heard one of the 64 spoken pseudowords (trained/untrained, day 1/day 2). A visual probe appeared 50ms after the offset of the spoken word and participants were prompted to respond whether the written and spoken forms matched using a button press. In this version of the experiment there were 4 conditions in which the written probe did not match the spoken form. These mismatching conditions tested the impact of bigram familiarity and error position. Errors in onset position could either occur in context of (1) trained initial and final bigrams or (2) untrained initial but trained final bigrams. Conversely errors in coda position could either occur in context of (3) trained initial and final bigrams, or (4) trained initial but untrained final bigrams. In this manner we could assess the speed with which participants could detect an error in either familiar or unfamiliar bigram contexts. We did not include probes in which both bigrams were unfamiliar as participants could have rejected these on the basis of visual familiarity rather than any necessarily orthographic process.

Recognition memory task. A recognition memory measure was included to assess whether participants were more accurate in identifying lexical items learned on day 1 compared to recently learned items. The recognition memory task was broken into two blocks, one testing

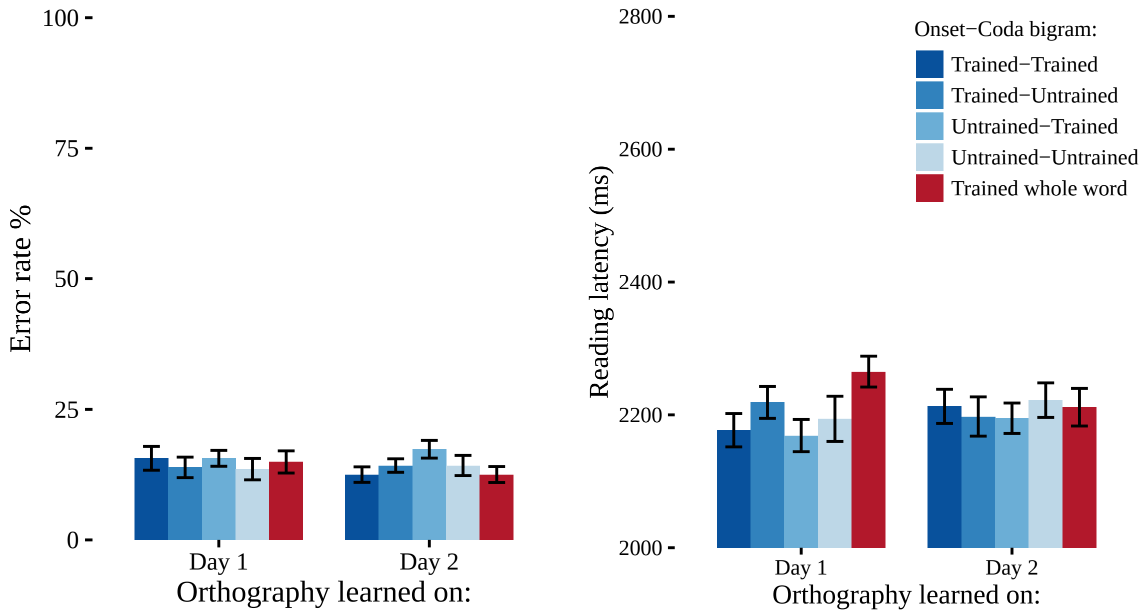


Fig. 4.13 Performance on the reading task. The left panel shows accuracy for trained and untrained words from days 1 and 2. The right panel shows reading latency for the same. Colours in blue represent untrained words according to the training status of the words constituent bigrams.

memory for the written word forms and one testing memory for the spoken word forms. The written block always preceded the spoken block. During the task all 64 target words were presented (thus half untrained), with the untrained words operating as foils to the trained words from days 1 and 2. Participants were first asked whether the item had appeared during training and responded with a button press. A follow-up question asked them how confident they were in that judgement with participants responding along a 6-point scale.

4.8 Results

4.8.1 Reading task

Accuracy. Accuracy on the reading task compared responses according to whether the orthography was learned on day 1 or 2 and according to the familiarity of their constituent bigrams (trained whole word, trained onset-trained coda, trained onset-untrained coda, untrained onset-trained coda, and untrained onset-untrained coda). These factors are not orthogonal - comparing trained whole words against the four untrained word conditions (listed above) would lead to an unbalanced comparison and would require a subsequent analysis to compare differences within the four untrained word conditions. To avoid the

Table 4.2 Experiment 3: Performance on reading task

	Accuracy (%)		Response time (ms)	
Onset-Coda bigram	Mean	St. Dev	Mean	St. Dev
Day 1 trained whole word	88.01	32.54	2265	535
Day 1 trained - trained	89.31	30.95	2177	490
Day 1 trained - untrained	90.57	29.28	2219	532
Day 1 untrained - trained	89.41	30.83	2169	513
Day 1 untrained - untrained	92.49	26.41	2194	524
Day 2 trained whole word	91.22	28.35	2212	542
Day 2 trained - trained	92.04	27.11	2213	551
Day 2 trained - untrained	91.15	28.45	2198	539
Day 2 untrained - trained	86.64	34.09	2194	518
Day 2 untrained - untrained	89.13	31.17	2222	524

inflation of type-1 errors, and because we do not have a strong prediction for the outcome of the analysis, we compare all five conditions together. The analysis is therefore a 2 x 5 level comparison between the day in which the orthography was learned (day 1 or day 2) with the bigram training status of the item (the 5 conditions listed above). There was no main effect of day, [$\chi^2(1) = 0.81,\ p = 0.37$], no effect of bigram familiarity, [$\chi^2(4) = 2.98,\ p = 0.56$], nor any interaction, [$\chi^2(4) = 2.63,\ p = 0.62$] (Figure 4.13, p112, left panel; Table 4.2, p113).

Reading latencies. Incorrect responses were discarded and further analyses include only the correct responses. As before the model used an inverse-gaussian link function for the latency data but figures show untransformed results for ease of interpretation. As with accuracy, reading latencies revealed no main effect of day, [$\chi^2(1) = 1.06,\ p = 0.30$], no effect of bigram familiarity, [$\chi^2(4) = 2.56,\ p = 0.46$], nor any interaction between day and bigram status, [$\chi^2(4) = 0.24,\ p = 0.97$] (Figure 4.13, p112, right panel; Table 4.2, p113).

4.8.2 Matching task

Accuracy. Participants learned to recognise the probes to a high level (performance in all conditions $> 92\%$; Figure 4.14, p114; Table 4.3, p114). However, there was no effect of day of learning on accuracy on the matching task, [$\chi^2(1) = 0.59,\ p = 0.44$], no effect of training, [$\chi^2(1) = 0.26,\ p = 0.61$], nor an interaction between these factors, [$\chi^2(1) = 1.82,\ p = 0.18$].

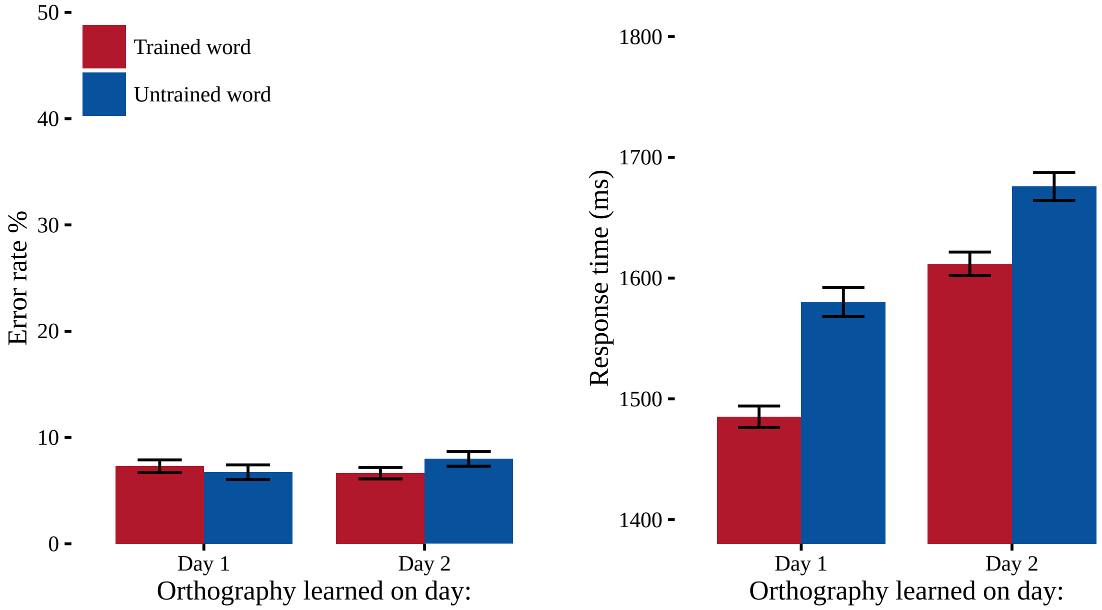


Fig. 4.14 Performance on the matching task. The left panel shows the accuracy for trained and untrained words from days 1 and 2, collapsing matching and mismatching trials. The right panel shows response times to trained and untrained words on matching trials.

Table 4.3 Experiment 3: Performance on matching task

	Accuracy (%)		Response time (ms)	
	Mean	St. Dev	Mean	St. Dev
Day 1 trained	92.71	26.00	1485	469
Day 1 untrained	93.27	25.06	1580	470
Day 2 trained	93.36	24.90	1612	483
Day 2 untrained	92.01	27.11	1676	467

Response times to matching probes. Following the rationale of Experiment 2 matching and mismatching trials were analysed separately. Replicating the findings of Experiment 2 there was a main effect of training showing faster responses for trained items, $[\beta = -41.34,\ SE = 11.08,\ t = -3.74,\ p < 0.001,\ \chi^2(1) = 8.33,\ p = 0.004]$ (Figure 4.14, p114; Table 4.3, p114). Furthermore responses were faster to words written in the orthography learned on day 1, $[\beta = -50.93,\ SE = 11.51,\ t = -4.42,\ p < 0.001,\ \chi^2(1) = 198.10,\ p < 0.001]$. However, even with the extra training in this experiment there was no positive evidence for the day-by-training interaction that would suggest lexical forms from day 1 had been consolidated: $[\chi^2(1) = 0.94,\ p = 0.33]$.

Response times to mismatching probes. The mismatching probes assessed whether participants could detect errors more quickly if the written probe contained familiar bigrams. Unlike in Experiment 2 we do not include error position (onset or coda) as a factor in the analysis but instead analyse onset and coda errors separately. This is because different probes are relevant in each comparison. For errors in onset position we compare the effect of trained or untrained initial bigram, but in both cases the coda bigrams are always trained. In the case of errors in coda position the final bigrams may be trained or untrained but the initial bigrams are always trained. Thus to include by-item random effects the analyses must be run separately. Figure 4.15 (p116) shows response times to detect errors in onset and coda positions. The critical comparison is whether the critical mismatching bigram has been seen during training, and whether this impact varies according to whether consolidation has taken place.

The main effect of bigram familiarity did not influence response times to detect errors in onset position, $[\chi^2(1) = 0.02,\ p = 0.87]$, nor was there a main effect of day, $[\chi^2(1) = 0.06,\ p = 0.81]$. The interaction between bigram familiarity and day was likewise not significant, $[\chi^2(1) = 0.001,\ p = 0.97]$ (Figure 4.15, p116, left panel). For errors in coda position however there was a main effect of probe condition showing slowed responses to probes containing trained bigrams, $[\beta = 42.94,\ SE = 16.65,\ t = 2.58,\ p = 0.01,\ \chi^2(1) = 5.66,\ p = 0.017]$. The main effect of day was not significant, $[\chi^2(1) = 0.53,\ p = 0.47]$, and there was no interaction between day and bigram familiarity, $[\chi^2(1) = 1.24,\ p = 0.27]$ (Figure 4.15, p116, right panel).

4.8.3 Recognition memory

The recognition memory task compared accuracy and response times for the question ‘Did you see this item during training’ for trained words from days 1 and 2. Untrained words were included as foils. D' scores were calculated to account for potential response bias between

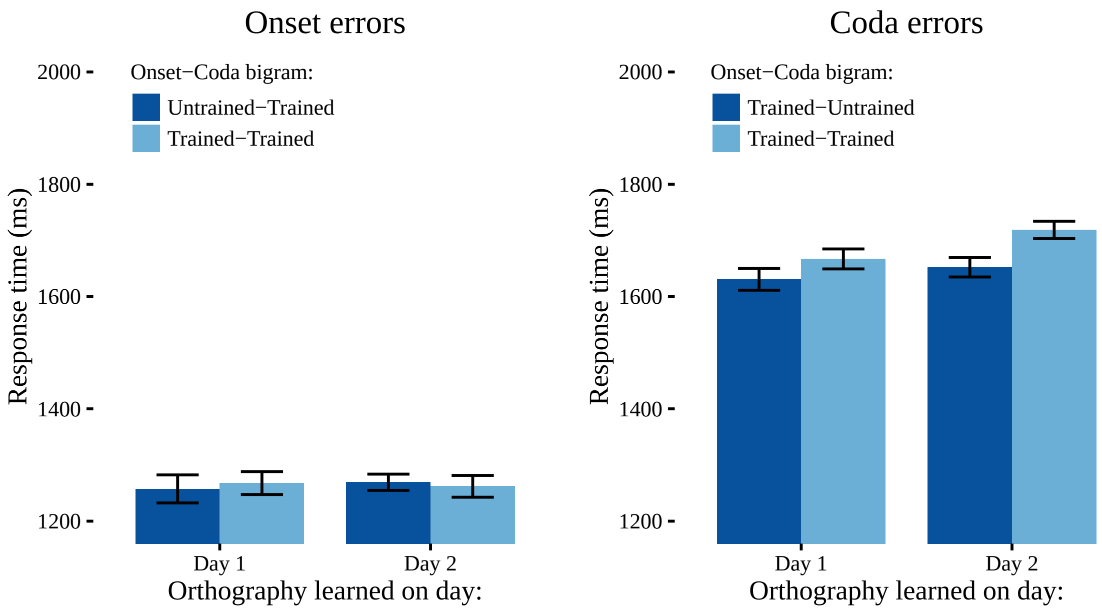


Fig. 4.15 Detecting mismatches according to error position. Errors in onset position are compared according to whether the initial bigram was trained or untrained. Errors in coda position meanwhile are compared according to whether the final bigram was trained or untrained. Note that as these are mismatching probes all items are unfamiliar as whole word forms.

different participants. In cases where the proportion of hits for a given participant and condition was equal to 1 (i.e., $N_{hits} = N_{trials}$) we instead used the formula $\frac{N_{hits}+0.5}{N_{trials}+1}$ (Snodgrass and Corwin, 1988). If participants did not respond within the 3500ms trial time the data was not included. This led to one participant being excluded from this analysis as they only responded to two trials within the time allowed. In addition, response times were analysed after discarding incorrect trials. A follow-up question compared participants' confidence ratings when judging whether written words were trained but due to a software error the confidence ratings for the auditory items were not recorded.

Recognition memory for written forms. Accuracy to recognise trained words from both days 1 and 2 was high, $[M = 81.16\%,\ SD = 39.20]$ and $[M = 70.16\%,\ SD = 45.88]$, respectively (Figure 4.16, p117, top left panel). D' scores were calculated for each participant for items from each day and a paired t-test compared sensitivity to words from days 1 and 2. This comparison showed that sensitivity did not differ for items from days 1 and 2, $[t(16) = 1.52,\ p = 0.15]$ (Figure 4.16, p117, top middle panel). Response times below 1500ms were excluded from analysis based on visual inspection of a histogram of response times. Response times were faster for trained words than for untrained words

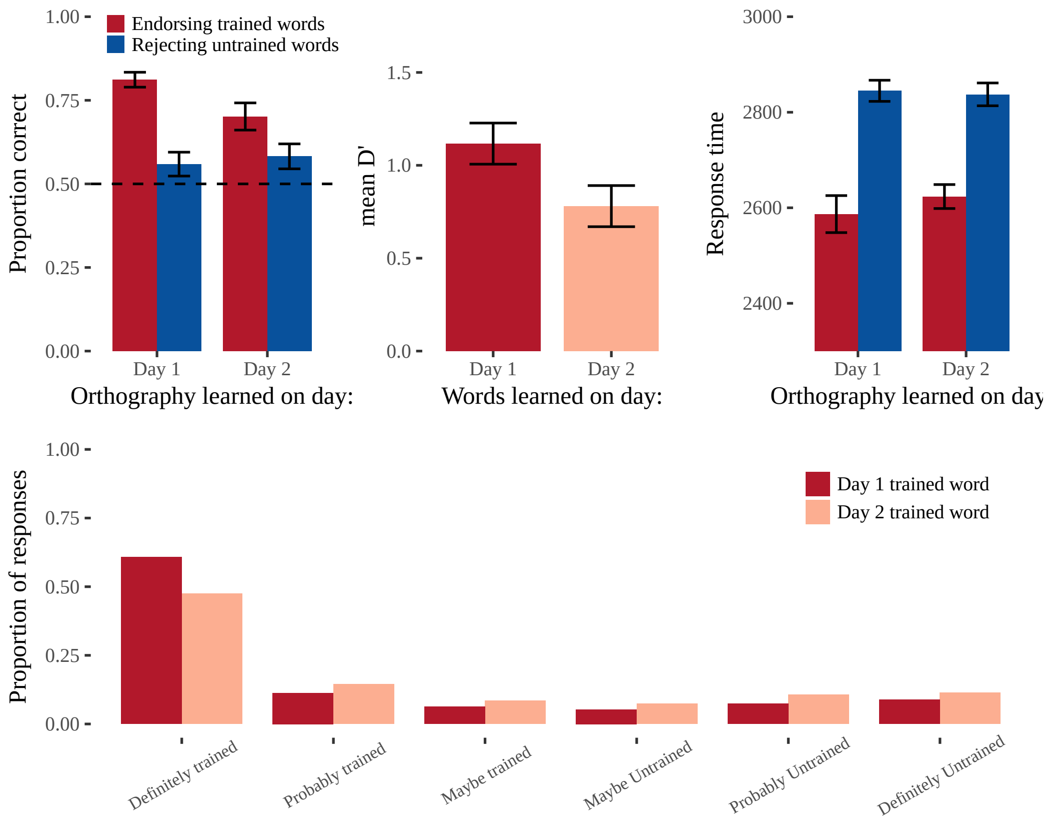


Fig. 4.16 Accuracy, d' scores and response times for written items on the recognition memory task. The dotted line in the top left figure represents chance performance on the recognition memory task. D' scores indicate sensitivity to distinguish trained and untrained words from days 1 and 2. The top right figure shows response times to make judgements about trained and untrained words from days 1 and 2. The bottom row shows participants' confidence ratings for judgements about trained words from days 1 and 2

$[\beta = 312.98,\ SE = 46.80,\ t = 6.69,\ p < 0.001,\ \chi^2(1) = 61.75,\ p < 0.001]$. There was no main effect of day $[\chi^2(1) = 0.14,\ p = 0.71]$, nor was there an interaction between day and training on response time, $[\chi^2(1) = 0.45,\ p = 0.50]$, (Figure 4.16, p117, top right panel). A Kruskall-Wallis test on the ordinal confidence ratings revealed that there was a significant effect of day on participant's confidence in their recognition memory judgements, $[H(1) = 8.77,\ p = 0.003]$ (Figure 4.16, p117, bottom panel). Inspection of the group means suggests that participants are more confident of their judgements on trained words from day 1. We do not analyse confidence rating for correct rejections of untrained words as it is not clear what better performance would indicate.

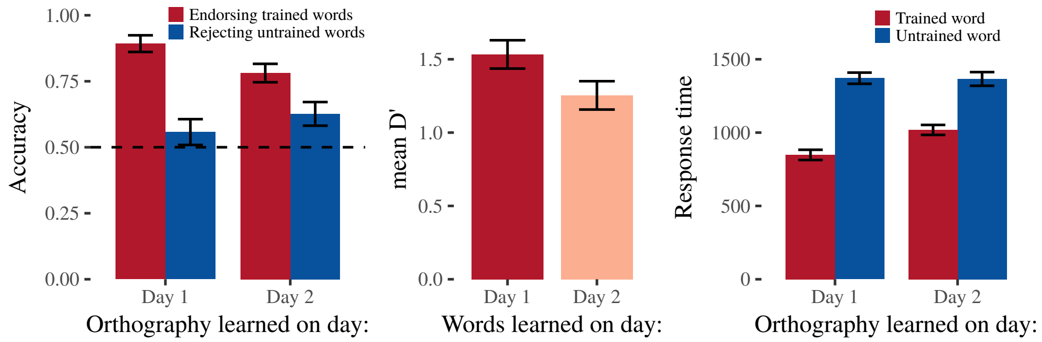


Fig. 4.17 Accuracy, d' scores, response times, and confidence judgements for spoken words on auditory recognition memory task. The dotted line in the top left figure represents chance performance on the recognition task. D' scores indicate sensitivity to distinguish trained and untrained words from days 1 and 2. The top right panel shows response times to make judgements for trained and untrained words from days 1 and 2. Confidence judgements were not recorded for the spoken word recognition task due to a software error.

Recognition memory for spoken forms. Recognition memory for spoken words was also high, $[M = 89.26\%,\ SD = 31.02]$ and $[M = 78.11,\ SD = 41.42]$, respectively (Figure 4.17, p118, left panel). D' scores showed that in the case of spoken words there was no difference in sensitivity to words from days 1 or 2, $[t(16) = 1.45,\ p = 0.17]$. Response times to spoken words were faster than for written words; in this case response times to spoken words that were greater than 3000*ms* were discarded based on visual inspection of the response time histogram. Responses were faster to trained words than untrained words, $[\beta = 516.46,\ SE = 36.96,\ t = 13.97,\ p < 0.001,\ \chi^2(1) = 114.06,\ p < 0.001]$. Responses were also significantly faster to words learned on day 1, $[\beta = 134.11,\ SE = 27.85,\ t = 4.86,\ p < 0.001,\ \chi^2(1) = 15.13,\ p < 0.001]$. For spoken words there was additionally an interaction between day and training indicating faster responses to trained words from day 1, $[\beta = -165.77,\ SE = 51.40,\ t = -3.22,\ p = 0.001,\ \chi^2(1) = 3.57,\ p < 0.059]$. (Figure 4.17, p118, right panel). Follow-up Tukey's tests (using the `glht` function from the `multcomp` package for R; version 1.4-6; Hothorn, Bretz, & Westfall, 2008) confirmed the interaction was driven by faster responses to trained words from day 1 compared to trained words from day 2, $[z = 4.51, p < 0.001]$, and that there was no difference between untrained words from days 1 and 2, $[z = 0.38, p = 0.97]$.

Finally, to directly test whether the pattern of response times differed between judgements for written and spoken forms a GLMM compared the effects of day of learning (day 1 or 2), training (trained or untrained), and modality (written or spoken words). The maximal model would not converge with both by-participant and by-item intercepts and so the by-item intercept was dropped. The critical three-way interaction between day, training, and modality was not significant, $[\chi^2(1) = 0.45,\ p = 0.50]$. Thus, we do not have evidence that consolidation effects differ for auditory and written forms.

4.9 Discussion

4.9.1 Replication of sublexical consolidation

This follow-up experiment achieved its primary goal of replicating the novel finding of overnight consolidation of sublexical knowledge, as shown by improved response times to the day 1 orthography on a spoken-to-written matching task. The critical comparison in the matching task therefore had minimal modifications from Experiment 2; in the condition where the spoken target and written probes matched exactly the same stimuli were used. The results therefore provide strong replication that a period of offline consolidation impacts on the ability to recognise novel written characters. Indeed participants were $111ms$ faster to correctly identify matching probes written using the day 1 orthography, a larger effect than the $57ms$ increase found in Experiment 2. One possible reason for this difference is that the increased training led to stronger memory traces prior to sleep, and thus stronger consolidation effects.

Again, neither the reading nor matching tasks showed evidence that the whole word forms learned on day 1 were consolidated, further replicating the findings of Experiment 2. Large improvements in response times for trained words over untrained words on the matching task indicate whole-word learning was taking place in some form. This learning did not lead to measurable consolidation effects however; the benefits to day 1 orthography applied equally to trained or untrained words. In contrast, there is no evidence that this whole-word learning helped participants in any way on the reading task. Unlike Experiment 2 where there were very large differences in reading latency between trained and untrained words, in the Experiment 3 reading task there were no differences. One interpretation for this discrepancy is that the inclusion of a greater number of untrained words on the reading task (1 : 4) has prompted participants to switch to a more serial reading strategy, thus relying less on their lexical knowledge. A recent experiment using different training methods in an artificial orthography setting has demonstrated that different reading strategies can be used to decode novel words (Taylor, Davis, & Rastle, 2017). Here the large number of untrained words may have encouraged participants to use the one factor familiar for both trained and untrained words; the letters.

No evidence for bigram consolidation. If learning and consolidation are taking place at the level of letters then what of bigram learning? Participants were equally fast to read trained and untrained words regardless of whether the constituent bigrams were familiar or not (see Figure 4.13, p112). This pattern suggests participants have adopted a completely serial reading strategy, concentrating on letters rather than any larger grain size. The matching task also failed to show evidence of bigram learning. In the mismatching condition (where target and probe differed) errors were detected in the same amount of time regardless of whether the critical bigrams were trained or untrained (Figure 4.15, p116).

On the whole, then Experiment 3 suggests not only is there not consolidation of bigram knowledge but that participants may not even have acquired bigram knowledge. The finding that there is no difference according to training status or bigram familiarity is in contrast to the results of Takashima et al. (2016) who showed both whole-word and bigram knowledge supported reading in a novel orthography. Moreover, this outcome seems to go against the outcome of Experiment 2 which suggested participants had learned bigram-level representations for trained words; responses to both trained and untrained words were slowed when written in a mixed orthography. The data from Experiment 3 show no indication that bigram learning aids in either the reading or recognition of recently learned written forms. We address this issue further in Experiment 4.

4.9.2 Consolidation of spoken word representations

One of the goals throughout these experiments has been to tease apart the interrelated contributions of spoken language and visual word knowledge. The recognition memory measure provides contrasting perspective on the lack of whole-word consolidation effects in other measures, showing evidence for consolidation of both spoken and written word forms.

The most notable outcome of the recognition memory task is that participants are significantly faster to make judgements about trained spoken words from day 1, thus indicating they have consolidated something of the spoken word form. The complementary measure of written word recognition did not show this pattern however; there was no statistically significant difference between response times to written and spoken forms. We cannot therefore conclude that different processes of consolidation impact on written and spoken word learning (Figure 4.16, p117, and Figure 4.17, p118).

Overnight changes in the ability to make visual discrimination has been shown elsewhere (Atienza et al., 2004). The experimental design here, however, does not require discrimination of visual patterns. Participants must instead make a judgement about the configuration of the whole word. The foils on this task were made up of the balanced untrained item set, and so there are no sub-lexical cues that could aid recognition. Consequently, the finding cannot be attributed to low-level perceptual changes, or to changes at the level of orthography. This interpretation also fits with existing research; learning complex visual objects improves with a period of sleep (Baeck et al., 2014).

Nevertheless, the effect of these changes are not apparent in the reading aloud or matching tasks; participants are neither quicker nor more accurate for trained words from day 1. What then to make of these improvements on the recognition memory measure? One possibility is that participants have consolidated both written and spoken representations of the novel word but have not consolidated the mapping from written to spoken form. Unlike the matching task, the recognition memory task does not require participants to make a judgement about both letters and sounds. Another study of cross-modal consolidation showed that overnight changes could occur within modalities, but that consolidation across modalities did not emerge until one week later (Bakker et al., 2014). It may be the case that participants have consolidated both spoken and written forms following training but the mapping between them remains weak. Using first and second languages Van Assche et al. (2016) explored the cross-modal transfer effects in reading (lexical decision) and object naming, again finding the reduced transfer of learning across modalities.

4.10 Experiment 4

Experiment 4 extends the findings of Experiments 2 and 3 to explore bigram learning and consolidation by skewing the input distribution of the bigrams. This manipulation is achieved while maintaining equivalence of training at the level of letters and words. During training participants see all letters and all words an equal number of times, but within those words some bigrams appear more frequently than others.

This approach may be valuable in linking bigram learning in the context of an artificial orthography to analogous results in children (e.g., Cassar and Treiman, 1997; O'Brien, 2014; Pacton et al., 2005). One way to frame the impact of learning these bigram pairs is in terms of consonantal context; when reading unfamiliar words the choice of vowel pronunciation may be influenced by both the preceding and subsequent consonants. This effect is found in adult as well as developing readers and so learning the pairings between consonants and vowels may play an important role in the early stages of reading acquisition (Treiman et al., 2006). Indeed, segmentation of words into intermediate forms may precede the ability of children to break syllables into sounds (Bernstein and Treiman, 2001). This perspective is consistent with proposals by Goswami and Bryant that orthographic learning takes place at different grain sizes from the beginning of reading (1990). Thus, in addition to the primarily letter-level learning of phonics, learning bigrams may offer an alternative route into early literacy learning.

A related perspective on these findings has to do with learning orthotactic constraints. Grainger et al. (2003) show that children are more accurate to recognise words (e.g., TABLE) and pseudowords (e.g., TOBLE) compared to illegal nonwords (e.g., TPBFE) using a Reicher-Wheeler paradigm, but children's accuracy for words and pseudowords did not differ. These data suggest children have already formed a sublexical representation of the orthographic constraints in English. Adults meanwhile do show significant differences between words and pseudowords. Thus, bigram learning might play an intermediary role in the shift from letter decoding to word recognition. Chetail (2017) shows evidence consistent with this interpretation using an artificial orthography paradigm. Here participants were presented with the visual forms artificial words written in an artificial orthography. Within these words some bigrams appeared more frequently. Just a few minutes of exposure to the written forms was enough for participants to show sensitivity to bigrams in a wordlikeness measure.

Faster responses to probes written in a consistent orthography suggested bigram learning had taken place in Experiment 2. Experiment 3 tested this question by having participants read untrained words that varied according to whether the bigrams that made up the words had been trained and consolidated. While the results of Experiment 3 showed no evidence of bigram learning, the design did not manipulate the frequency of bigrams during training. In other words, the experimental design carefully controlled the frequency of letters and words but not of bigrams. Consequently, the experiment only assessed incidental learning of bigrams. In the current experiment we go a step further by simultaneously controlling frequency throughout the various levels of the written form. Each letter appears an equal number of times during training, within words that also appear an equal number of times during training. On top of this we embed bigram structure such that some bigrams appear twice as frequently as others during training. Over the course of 12 blocks of training bigrams pairs could thus appear with a frequency of 12 or 24. A critical feature of this experiment was the construction of paired trained and untrained word sets that allowed full counterbalancing of the design. The word sets were constructed so that each could be rotated to appear as trained or untrained word sets. Thus, each untrained item set contained bigrams had appeared once or twice in the other trained word set, as well as bigrams that remain novel. Consequently each of the untrained words sets contained bigrams with a frequency of 0, 12, and 24 during training. In this way we can measure the impact of bigram learning and consolidation in the context of an unfamiliar word.

In addition to the reading and matching tasks in the previous experiments we also include a recognition memory measure again. The results of Experiment 3 show a pattern of results that suggested overnight changes in recognition memory may have taken place. Technical problems and consequent low participant numbers meant that these results are underpowered. By replicating the previous analysis to compare recognition memory for written and spoken words we attempt to address this problem. A finding of overnight improvements in recognition memory for spoken words would link the current work with other findings of lexical consolidation.

Finally this experiment allows yet another opportunity to replicate the novel finding of overnight sublexical consolidation. Moreover, the design of this experiment required creation of a new set of items, thus assessing consolidation using items that differ from the original finding. Where Experiment 3 deliberately held constant as many experimental factors as possible from Experiment 2, here we test the robustness of the effect by using a different item set.

4.11 Methods

4.11.1 Participants

22 native English-speaking adults aged 18 – 40 took part in a study approved by the Cambridge Psychology Research Ethics Committee. No participants reported having dyslexia, speech, or language impairment. Additionally two participants did not complete the recognition memory task and so data from the remaining 20 participants are reported for this analysis.

4.11.2 Materials

Experiment 4 directly manipulated participants' exposure to bigram frequency during training. To this end four novel sets of spoken pseudowords were constructed (see Figure 4.19, p125 for sample assignment). The item sets were assigned to appear either as trained or untrained words on day 1 or 2 in a counterbalanced manner across participants. As before these sets were made up of two non-overlapping phoneme sets, so that letter-sound mappings learned on day 1 would be distinct from those learned on day 2. The same phoneme and vowel combinations from Experiment 3 were used: b, f, n, v, s, θ, ʤ, and k made up one set of consonants. The consonants d, m, p, z, t, ʃ, h, tʃ made up the second set. Both sets of consonants appeared with the same four vowels: æ, ɛ, ɒ, ʌ.

The items were constructed to ensure that the 16 written words in each set contained each consonant twice in onset position and twice in coda position. In contrast to the previous experiments however the items were constructed such that within each set of 16 words the

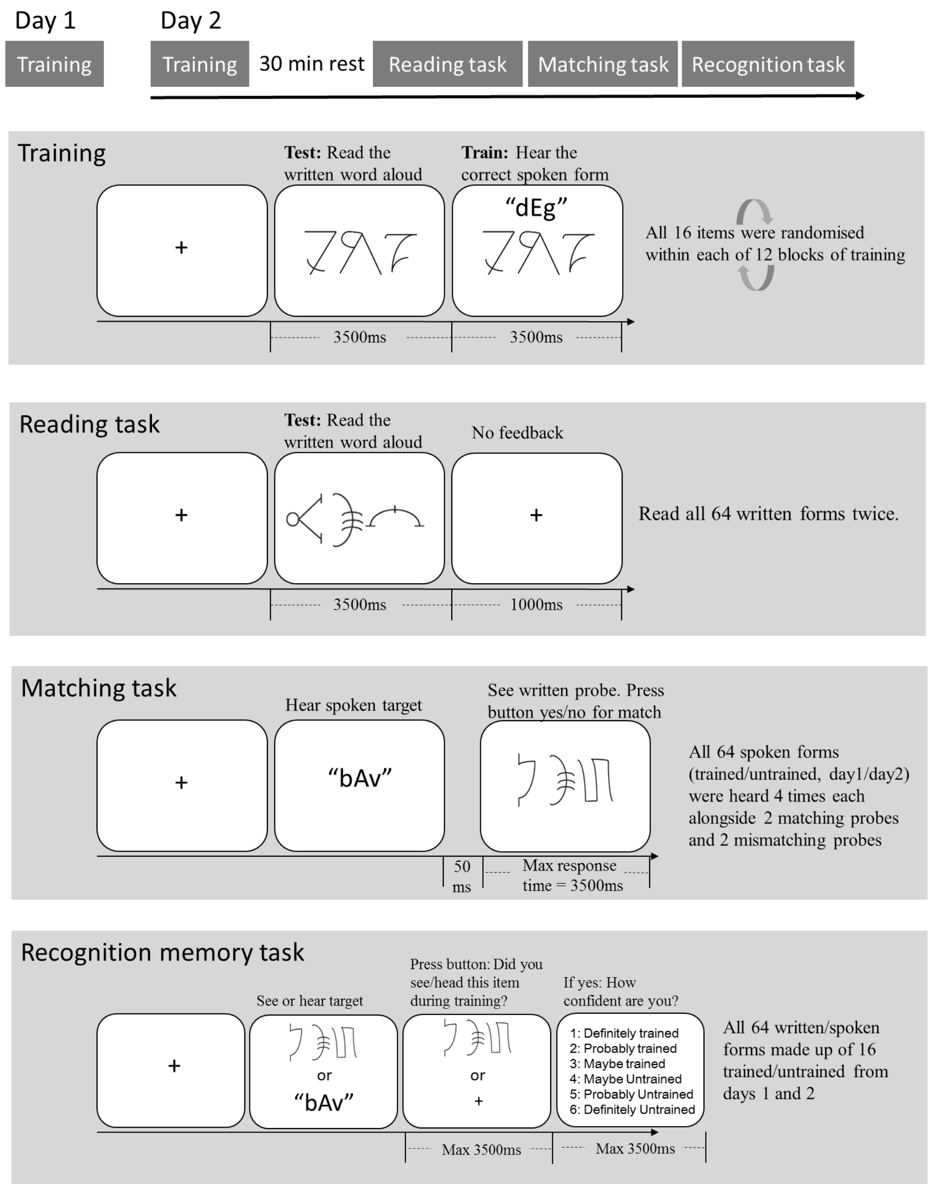


Fig. 4.18 (A) Sequence of training and testing procedures over the two days of the study. (B) Timeline of a training trial. Participants are asked to read the item aloud before hearing the correct spoken form. Participants then have 3500ms to learn before the next trial. The 16 items to be trained were randomised within each run. (C) Timeline of a reading test trial. Participants were tested using a similar procedure to training, but did not hear the correct form of the written word and were given no feedback. (D) Timeline of a matching trial. The spoken target word was rapidly followed by a written probe onscreen. Participants had up to 3500ms to indicate whether the spoken target and written probes matched by pressing a button. (E) Timeline for a recognition memory trial. 16 trained items from days 1 and 2 were presented alongside 32 untrained foils. Participants first completed one run making judgements on the written forms of the items followed by a second run making judgements on the spoken forms of the items. The item was presented and participants made a judgement about its training status. If they indicated the item had been trained then they made a confidence rating about this judgement.

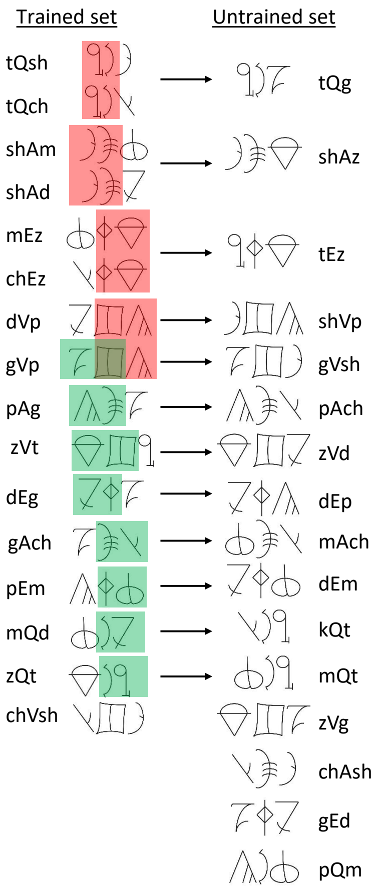


Fig. 4.19 Sample stimulus assignment containing examples of frequently trained bigrams. This figure shows a set of 16 words to be trained and the 16 corresponding untrained set used to assess bigram generalisation. Bigrams highlighted in red appear twice in the training set and are tested in an untrained word in the corresponding untrained word set. Bigrams highlighted in green appear once in this training set and are tested in an untrained word in the corresponding untrained word set. Bigrams not highlighted do not appear in the untrained word set. The item sets were counterbalanced for different participants.

frequency of bigrams varied. In item set on the left of Figure 4.19 (p125) for example the bigrams *'tQ'* and *'Ez'* each appear in two different words. Moreover, the item sets were constructed in a reciprocal manner so that the complementary untrained word set would allow us to test generalisation. For example, in the untrained word set on the right of the figure, the bigrams *'tQ'*, *'pA'*, and *'gE'* all appear in different words. Although now embedded within an untrained word, the bigram *'tQ'* appeared in two words of the trained word set. *'pA'* by contrast appeared once in the trained word set, while *'gE'* is a completely novel combination of letters. In this way we can ask whether participants will respond faster to untrained words containing familiar or frequent bigrams.

4.11.3 Procedure

A full timeline of the experiment can be found in Figure 4.18 (p124). All tasks were presented using custom scripts in Psychopy, a python-based software package for psychophysics (Peirce, 2007, 2009).

Training

The training procedure was kept consistent with Experiment 3 (Figure 4.18, p124). Participants learned to read 16 written words each day with a different orthography learned each day. As before participants completed a brief familiarisation block where the written and spoken forms of the 16 items were presented briefly. During training participants were asked to read aloud the written form onscreen. After 3500*ms* the correct spoken form of the word was presented and participants had a further 3500*ms* to learn the association between letters and sounds. The 16 items were randomised within each block of training and participants saw 12 presentations of each item over the course of four training runs. Responses were recorded and scored offline and were marked correct only if the participant reported all three phonemes in a word correctly. The manipulation of bigram frequency within the words meant that each bigram had a frequency of 1, or 2 in the training set. Thus during the 12 blocks of training a bigram could appear 12, or 24 times.

Testing

Participants had a 30 minute rest break after training on day 2. Following the break they had a single reminder block where the written and spoken forms of the trained words from both days were presented once.

Reading task. Reading trials followed the form used in Experiments 2 and 3 (Figure 4.18, p124). A written word appeared onscreen and participants had 3500*ms* to read the item aloud and no feedback was given. Items appeared in a randomised order and responses were recorded and scored offline for accuracy and response time. Participants read aloud the 32 trained items from days 1 and 2 (16 each day) as well as 64 untrained items. The untrained items varied according to the frequency with which their constituent bigrams appeared during training. Untrained words could contain bigrams that were either entirely novel to the reader, or which had been learned 12 or 24 times during training.

Matching task. As before for each trial of the matching task participants first heard one of the 64 spoken pseudowords (trained/untrained, day 1/day 2). A visual probe appeared

50*ms* after the offset of the spoken word and participants were prompted to respond whether the written and spoken forms matched using a button press (Figure 4.18, p124). Half of the written probes (256) matched the spoken target while half mismatched. As noted above the untrained test items were designed in such as way that each bigram had a training frequency of 0, 12, or 24. In this version of the matching task the primary question was whether participants would be faster to recognise matches that contained frequently trained bigrams. The mismatching written probes could have an error in either onset or coda positions, as before.

Recognition memory task. The same procedure as experiment 2 was used for the recognition memory task (Figure 4.18, p124). The first block tested memory for the written word forms while the second test memory for the spoken word forms and the written block always preceded the spoken block. The 16 trained words from days 1 and 2 (thus 32) were included alongside 32 untrained foils. As before participants were first asked whether the item had appeared during training and responded with a button press. A follow-up question asked them how confident they were in that judgement with participants responding along a 6-point scale.

4.12 Results

4.12.1 Reading task

Accuracy. The primary analysis of the reading data compared accuracy according to the day of learning and training status. Accuracy was high across all conditions: [day 1 trained words 84.18%, day 1 untrained words 80.64%, day 2 trained words 88.94%, and day 2 untrained words 85.31%] (Figure 4.20, left panel, p128; Table 4.4, p128). The main effect of training was significant according to the likelihood ratio test but this effect did not reach significance within the maximal model, [$\beta = -0.42$, $SE = 0.30$, $z = -1.37$, $p = 0.17$, $\chi^2(1) = 4.52$, $p = 0.033$]. In addition, neither the main effect of day, [$\chi^2(1) = 0.25$, $p = 0.61$], nor the interaction of day and training, [$\chi^2(1) = 0.40$, $p = 0.53$], reached significance.

Reading latency. As before incorrect responses were discarded from the analysis of reading latencies. Fastest responses were for trained words from day 1 ($M = 1946$*ms* and trained words from day 2 ($M = 2051$*ms*), followed by untrained words from day 1 ($M = 2085$*ms*) and day 2 ($M = 2190$*ms*; Figure 4.20, right panel, p128; Table 4.4, p128). There was a main effect of day on reading latencies, [$\beta = 55.93$, $SE = 21.35$, $t = 2.62$, $p = 0.009$, $\chi^2(1) =$

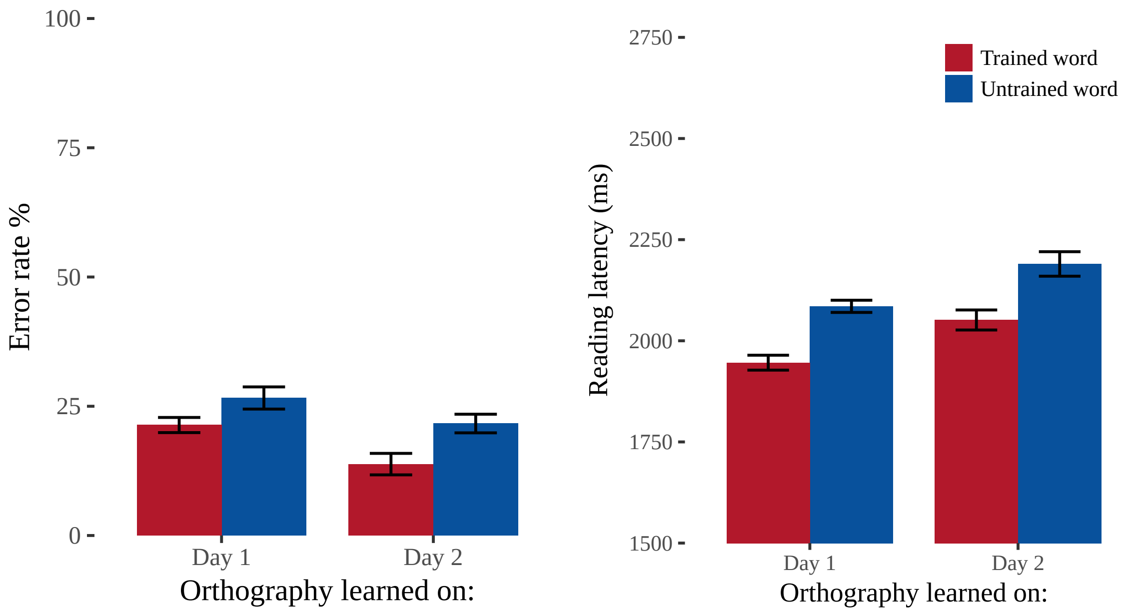


Fig. 4.20 Accuracy and response times on the reading task

Table 4.4 Performance on reading task

	Accuracy (%)		Reading latency (ms)	
	Mean	St. Dev	Mean	St. Dev
Day 1 trained	78.63	41.02	1946	546
Day 1 untrained	73.40	44.22	2085	536
Day 2 trained	86.19	34.52	2051	537
Day 2 untrained	78.34	41.22	2190	577

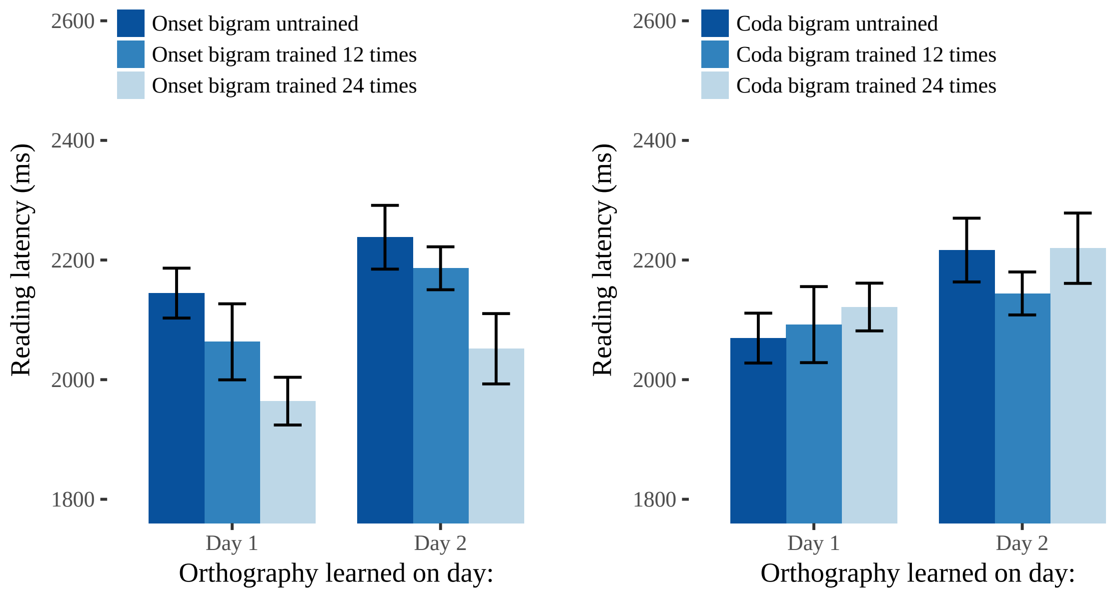


Fig. 4.21 Reading latency for untrained words according to bigram training status.

6.33, $p = 0.012$], indicating faster reading latencies for words written in the day 1 orthography. Participants were also significantly faster to read trained than untrained words, [$\beta = 188.90$, $SE = 28.98$, $t = 6.52$, $p < 0.001$, $\chi^2(1) = 24.67$, $p < 0.001$]. The interaction between day and training of learning and training that would indicate additional improvements for trained words from day 1 did not reach significance, [$\chi^2(1) = 0.58$, $p = 0.45$].

Effect of bigram frequency on reading latency. In order to assess the impact of bigram frequency during training we further compare reading latencies for untrained words according to day of learning and bigram frequency. Figure 4.21 (p129) shows the impact of bigram frequency during training on reading untrained words. Models that attempted to include the effect of both onset and coda bigram frequencies did not converge due to the large number of parameters. Instead the impact of onset and coda bigrams are analysed separately. Bigram frequency in onset position during training did not impact on reading speed, [$\chi^2(2) = 4.04$, $p = 0.13$]. In addition there was no main effect of day, [$\chi^2(1) = 2.88$, $p = 0.09$], nor any interaction between day of learning and frequency of bigram, [$\chi^2(2) = 0.56$, $p = 0.75$]. A similar pattern of results was found for bigrams learned in coda position during training. There was no effect of either bigram frequency, [$\chi^2(2) = 0.04$, $p = 0.98$], or day, [$\chi^2(1) = 3.71$, $p = 0.05$], on reading latency. The interaction between bigram frequency and day of training was also not significant, [$\chi^2(2) = 0.16$, $p = 0.92$].

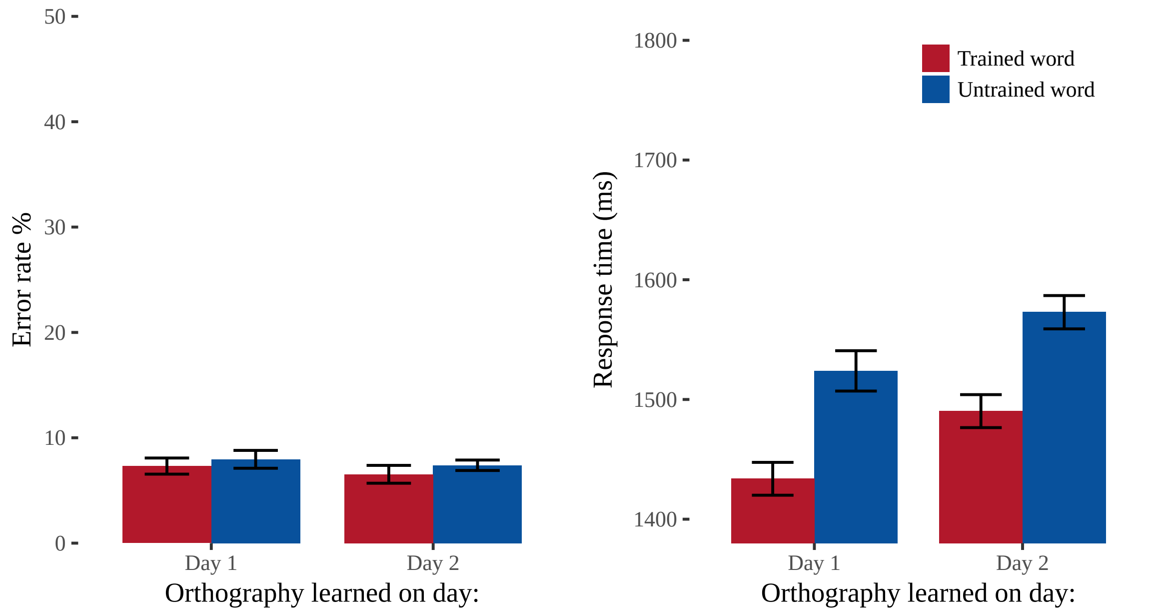


Fig. 4.22 Accuracy and response times on the matching task

Table 4.5 Performance on matching task

	Accuracy (%)		Reading latency (ms)	
	Mean	St. Dev	Mean	St. Dev
Day 1 trained	92.68	26.04	1398	541
Day 1 untrained	92.05	27.06	1466	538
Day 2 trained	93.47	24.72	1455	534
Day 2 untrained	92.61	26.17	1493	533

4.12.2 Matching task

Accuracy. Accuracy on the matching task was high, with performance in all conditions $> 90\%$ (see Figure 4.22, left panel, p130; Table 4.5, p130). As in the experiments 2 and 3 there was no effect of training $[\chi^2(1) = 1.31,\ p = 0.25]$ or of day $[\chi^2(1) = 0.96,\ p = 0.33]$ on accuracy. Moreover there was no interaction between day and training, $[\chi^2(1) = 0.02,\ p = 0.90]$

Response times. The overall pattern of responses shown here offer a second replication of the finding of overnight consolidation at the sublexical level (Figure 4.22, right panel, p130; Table 4.5, p130). Participants responded faster to words written in the day 1 orthography, $[\beta = -21.63,\ SE = 9.22,\ t = -2.35,\ p = 0.02,\ \chi^2(1) = 2.92,\ p = 0.08]$. In this case the likelihood ratio test comparing the maximal model to a model that did not include a

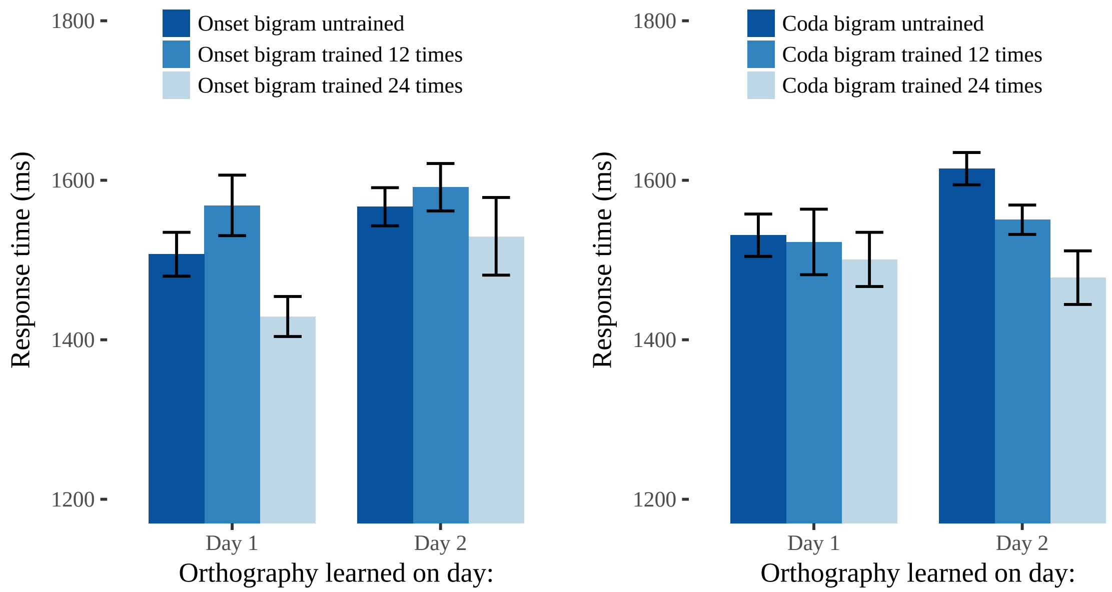


Fig. 4.23 Response times to endorse matches for untrained words according to bigram training status and thus indicates how well participants can generalise bigram learning. These whole words have never been seen and each letter has been trained equally. Therefore only the frequency with which bigrams appeared during training varies across conditions.

main effect of day is marginally significant but the coefficient estimate from the maximal model model is significant. Responses to trained words were faster than to untrained words, [$\beta = -24.38,\ SE = 7.40,\ t = -3.39,\ p < 0.001,\ \chi^2(1) = 8.31,\ p = 0.004$], but there was no interaction between day and training, [$\chi^2(1) = 0.94,\ p = 0.33$].

Effect of bigram frequency on response time. Here we assess how well participants can generalise their knowledge of bigrams according to the frequency of the bigram during training, and whether the bigrams have had an opportunity for consolidation (Figure 4.23, p131). In order to assess generalisation only trials involving untrained spoken targets are included. The effect of bigram frequency followed the same rationale used elsewhere in the thesis, using model comparison to select the best fitting mixed effects model. However, in this analysis attempts to simultaneously model the impact of onset bigram frequency and coda bigram frequency did not converge. Instead these two effects are modelled separately. Thus, we compare first the impact of onset bigram frequency and day of training, and then the effect of coda bigram frequency and day of training.

The frequency of bigrams in onset did not affect response times on the matching task, [$\chi^2(2) = 0.2.77,\ p = 0.25$], nor was there a main effect of day [$\chi^2(1) = 0.96,\ p = 0.33$]. Additionally the interaction between onset bigram frequency and day was not significant,

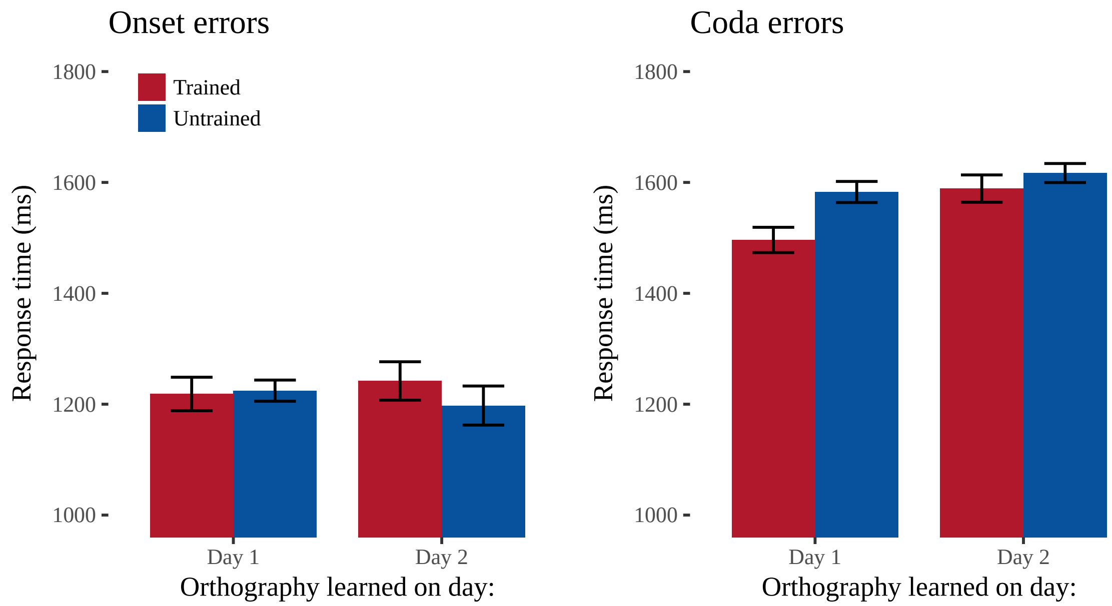


Fig. 4.24 Response times to detect mismatching probes according to error position.

$[\chi^2(2) = 0.71,\ p = 0.69]$. Analysis of the impact of coda bigram frequency during training showed no main effect of day, $[\chi^2(1) = 0.13,\ p = 0.14]$. There was also no main effect of coda bigram frequency, $[\chi^2(2) = 0.89, p = 0.64]$. There was no interaction between day of training and coda bigram frequency, $[\chi^2(2) = 1.33, p = 0.51]$.

Detecting mismatches in onset and coda position. A subset of the data, those trials in which the target and probe did not match, were analysed separately. This analysis allowed us to ask whether participants were faster to detect mismatches when the error occurred in onset or coda position, and whether this pattern changed after a period of consolidation. There was a significant three-way interaction between day of training, error position, and training status: $[\chi^2(1) = 12.50,\ p < 0.014]$ (Figure 4.24, p132). The two-way interactions between day of training and training status were analysed at both levels of error position separately. For errors in onset position there was no significant effect of day, $[\chi^2(1) = 0.62,\ p = 0.43]$, or of training, $[\chi^2(1) = 0.05,\ p = 0.82]$, nor any interaction, $[\chi^2(1) = 0.39,\ p = 0.53]$. By contrast for errors that occurred in the coda position participants were significantly faster to detect errors in the day 1 orthography, $[\beta = -35.61,\ SE = 12.29,\ t = -2.90,\ p = 0.004,\ \chi^2(1) = 5.78,\ p < 0.016]$. Responses were also faster to trained words compared to untrained words, $[\beta = -30.38,\ SE = 10.46,\ t = -2.90,\ p = 0.004,\ \chi^2(1) = 8.51,\ p = 0.004]$. There was no significant interaction between day and training however, $[\chi^2(1) = 0.15,\ p = 0.70]$. Notably

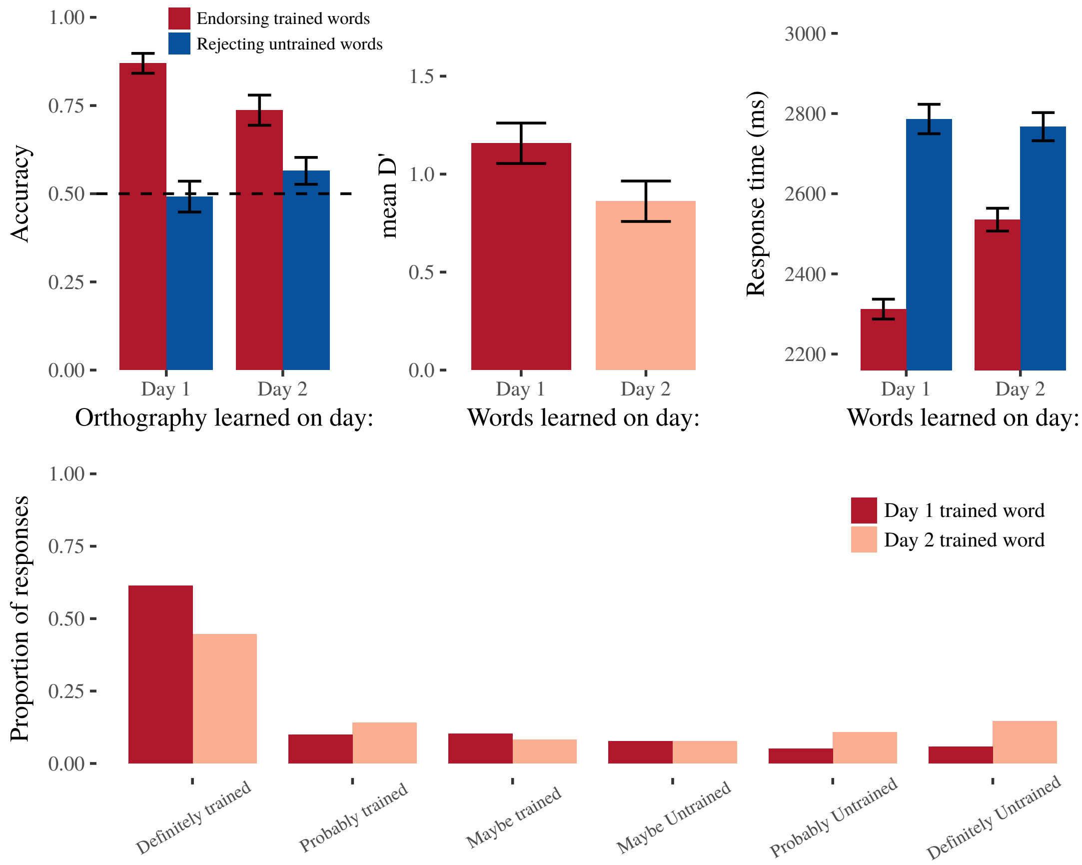


Fig. 4.25 Accuracy, d' scores and response times for written words on the recognition memory task. The dotted line in the top left figure represents chance performance on the recognition memory task. D' scores indicate sensitivity to distinguish trained and untrained words from days 1 and 2. The top right figure shows response times to make judgements about trained and untrained words from days 1 and 2. The bottom row shows participants' confidence ratings for judgements about trained words from days 1 and 2

this pattern of results differs from the equivalent analysis in Experiments 2 and 3 in failing to show faster responses for the day 1 orthography.

4.12.3 Recognition memory

Recognition memory for written forms. D' scores were calculated to assess accuracy of recognition memory for written words learned on days 1 and 2 using the same approach as Experiment 3. Although d' scores were numerically greater for words learned on day 1, $[M = 1.16,\ SD = 0.10]$, than words learned on day 2, $[M = 0.86,\ SD = 0.10]$, a paired t-test revealed no significant difference in participants' sensitivity to detect trained words from days 1 and 2, $[t(18) = 1.43,\ p = 0.17]$, (Figure 4.25, panel A, p133). Response times to correctly identify trained and untrained words from days 1 and 2 were also analysed. There were main effects of both day, [$\beta = 270.17,\ SE = 57.05,\ t = 6.21,\ p < 0.001,\ \chi^2(1) = 22.51, p < 0.001$], and training, [$\beta = 467.42,\ SE = 63.35,\ t = 7.38,\ p < 0.001,\ \chi^2(1) = 21.47, p < 0.001$]. In addition there was an interaction between day and training, [$\beta = -277.18,\ SE = 59.39,\ t = -4.67,\ p < 0.001,\ \chi^2(1) = 10.29,\ p = 0.001$]. Follow-up Tukey's tests were conducted using the `glht` function from the `multcomp` package for R (version 1.4-6; Hothorn, Bretz, & Westfall, 2008). This confirmed that the interaction was driven by faster responses to trained words from day 1 compared to trained words from day 2, [$z = 6.59, p < 0.001$], and that there was no difference between untrained words from days 1 and 2, [$z = 0.11, p = 0.99$], (Figure 4.25, panel B, p133). This pattern suggests overnight changes have taken place in participants' representations of trained words from day 1. As in the previous experiment confidence ratings were compared using a Kruskall-Wallis test, again showing participants were more confident in their recognition memory judgements for day 1 items, [$H(1) = 22.57,\ p < 0.001$].

Recognition memory of spoken forms. Analysis of recognition memory for spoken forms from days 1 and 2 followed a similar rationale to analysis of the written forms. As in Experiment 3 response times were very fast and so responses greater than 2000*ms* were discarded from visual inspection of the response time histogram. D' scores for trained words from days 1 and 2 approached significance, [$t(18) = 1.94,\ p = 0.068$] with the trend indicating better sensitivity to words from day 1 (Figure 4.26, panel A, p135). The maximal model for response time would not converge with inclusion of both by-participant and by-item random intercepts and so the by-item intercept was dropped from the model. Response times were faster to words from day 1, [$\beta = 126.61,\ SE = 30.31,\ t = 4.18,\ p < 0.001,\ \chi^2(1) = 14.00,\ p < 0.001$]. Trained words were also identified more quickly than untrained words, [$\beta = 397.12,\ SE = 60.64,\ t = 7.55,\ p < 0.001,\ \chi^2(1) = 95.33,\ p < 0.001$]. However, in contrast to the pattern for recognition of written words above, for spoken words there was no interaction between day and training, [$\chi^2(1) = 2.15,\ p = 0.14$], (Figure 4.26, panel B, p135). A Kruskall-Wallis test revealed that there was a significant effect of day on participant's confidence in their recognition memory judgements, [$H(1) = 24.40,\ p < 0.001$], with increased confidence for judgements about day 1 words.

Finally, to directly test whether the pattern of response times differed between judgements for written and spoken forms a GLMM compared the effects of day of learning (day 1 or 2), training (trained or untrained), and modality (written or spoken words). The maximal model would not converge with both by-participant and by-item intercepts and so the by-item intercept was dropped. The critical three-way interaction between day, training, and

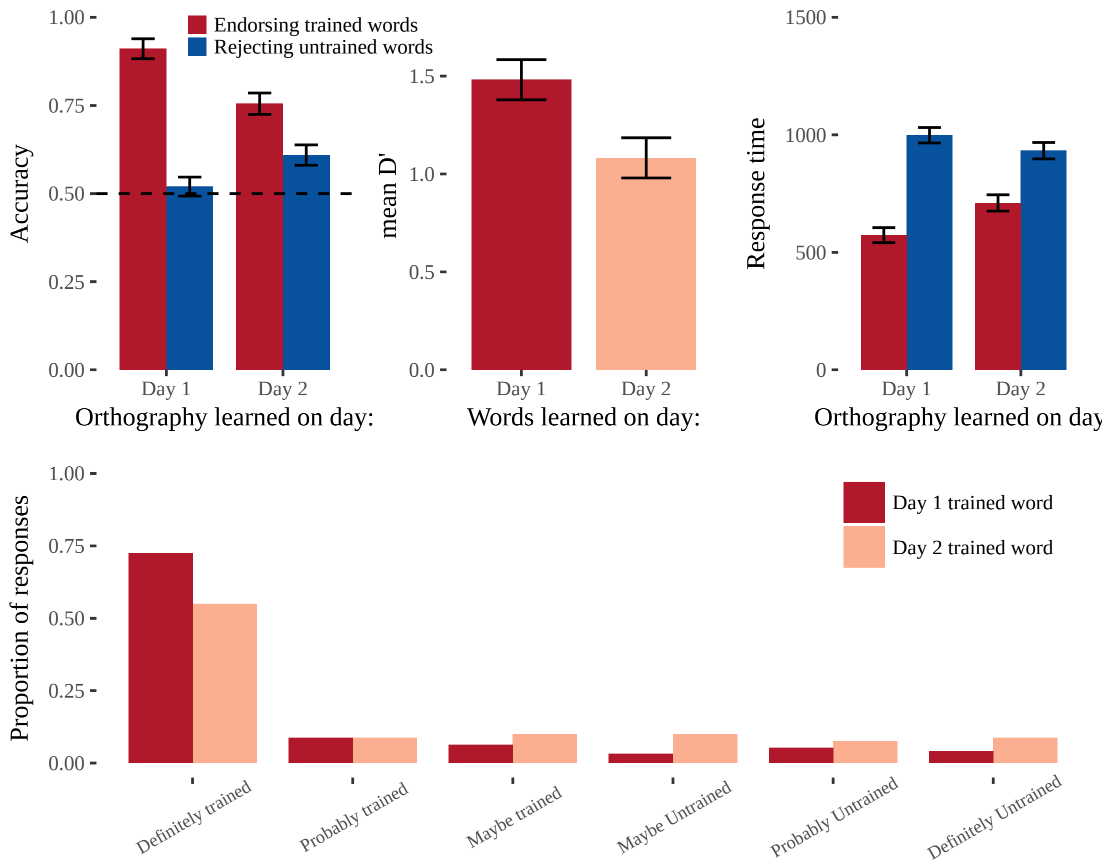


Fig. 4.26 Accuracy, d' scores and response times for spoken words on the recognition memory task. The dotted line in the top left figure represents chance performance on the recognition memory task. D' scores indicate sensitivity to distinguish trained and untrained words from days 1 and 2. The top right figure shows response times to make judgements about trained and untrained words from days 1 and 2. The bottom row shows participants' confidence ratings for judgements about trained words from days 1 and 2

modality was not significant, $[\chi^2(1) = 0.03,\ p = 0.86]$. Thus we do not have evidence that consolidation effects differ for auditory and written forms.

4.13 Discussion

The primary goal of this experiment was to assess whether frequent bigrams embedded within words would impact on learning to read a novel orthography, and to assess whether these effects might be impacted by overnight consolidation. In summary, this experiment does not show direct evidence for improvements in reading due to bigram frequency; untrained words containing frequent bigrams did not differ on either the reading or matching tasks.

Importantly, this experiment replicates yet again the finding of faster recognition of the orthography learned on day 1. Finally, response time differences on the recognition memory task provide strong evidence for overnight consolidation of whole words written in a novel orthography.

4.13.1 Replication and extension of sublexical consolidation

Here for the third time there is evidence that overnight consolidation of orthographic knowledge impacts on the sublexical level. Response times in the matching task were faster for trained and untrained words written in the day 1 orthography. As before, faster responses to trained words indicate participants have learned the whole word representations from days 1 and 2, but there was no additional benefit of overnight consolidation for the trained words from day 1. A more complex pattern of results emerges from the error detection comparison of the matching task. These data showed participants were significantly faster to detect errors in coda position for words written in the day 1 orthography (Figure 4.24, p132). The faster responses to words in the day 1 orthography were not found in onset position however. That this outcome only appears for errors in coda position supports the idea that participants are still reading these words in a serial manner; hence reading three letters (to detect an error in the final position) is more sensitive to the speed of processing than reading the first letter (to detect an error in the initial position).

An important novel finding in this experiment was that the benefits of sublexical consolidation extended to reading the words aloud. Mirroring the results of the matching task, participants were faster to read aloud trained and untrained words written in the day 1 orthography. Here again participants had learned the whole words during training on day 1, shown by faster responses to trained words. Despite learning the whole word form prior to a period of offline consolidation however there were no additional benefits to reading aloud the trained and consolidated words from day 1.

Impact of bigram frequency during training. The primary manipulation of interest in this experiment was the inclusion of frequently occurring bigrams during training. We hypothesised that participants might be sensitive to these bigrams in the early stages of orthographic learning, and that overnight consolidation might offer additional benefits in learning and generalising these bigrams.

On the reading task we asked whether participants were faster to read untrained words when those words contained bigrams that were novel, or had a training frequency of 12 or 24. Despite a trend towards faster responses for frequent bigrams in onset position, no differences were found in reading latencies (Figure 4.21, p129). Moreover, we did not find evidence for our hypothesis that the statistical structure provided by bigrams would be particularly amenable to overnight consolidation; participants' responses to frequent bigrams did not change after a period of sleep. The matching task mirrored the lack of bigram-related changes, showing neither main effects of bigram frequency nor any overnight changes to indicate consolidation at the level of bigrams. Untrained words meanwhile did not differ when the orthographies were mixed. This lack of direct evidence for improved performance on frequent bigrams seems somewhat surprising given the improvements in reading latencies for day 1 words. One explanation for this pattern is that the frequent bigrams supported learning by simplifying the task, but that knowledge of the bigrams themselves is not generalised.

4.13.2 Consolidation of written word representations

The recognition memory data from Experiment 4 offer an important complement to the previous findings. Participants made judgements about the spoken and written forms of trained words using unfamiliar words as foils. We hypothesised that overnight consolidation might support improvements in recognition memory for day 1 items. However, an outstanding question was whether any overnight improvements would impact on the spoken or written forms. Previous findings of lexical consolidation (e.g., Davis et al., 2009) have focused on learning spoken words and so it is possible that improvements may be found for spoken words. Unlike previous studies however, the focus of learning in this experiment was on the written forms (in contrast to phoneme monitoring in the case of Davis et al., 2009).

In light of this question the most notable outcome was that large differences in response time emerged for the written forms of words learned on day 1. Trained words were identified fully 223*ms* faster than trained words from day 2 (Figure 4.25, panel B, p133). This effect cannot be attributed to the faster processing of the consolidated orthography as the day 1 trained words were identified 474*ms* faster than untrained words written in the same orthography. Moreover, there was no difference in the speed to make judgements about untrained words from days 1 and 2 (2786*ms* versus 2767*ms*). Instead this finding suggests participants have learned and consolidated whole word forms. This finding is the first clear evidence for overnight lexical consolidation of written word forms across these experiments.

On the recognition task, where no phonological judgement is required the impact of whole word consolidation is apparent. The lack of apparent benefits for trained words from day 1 on the reading and matching tasks may then reflect the strategy adopted by participants on the task. The changes seen here are compatible with suggestions by Coltheart (2006) that whole-word representations are learned from the earliest stages of reading, and that the shift from serial to holistic word recognition is merely a reflection of a larger orthographic lexicons coming to the fore. Moreover the large changes in processing speed here may be a better indicator of the automatic processing of sight words suggested by Ehri (2005).

The recognition memory task did not clearly differentiate between consolidation of written and spoken representations. Overall recognition memory (as measured by d') did not differ for day 1 words when making judgements of either written or spoken forms. Nevertheless, participants were more confident in their judgements of words learned on day 1 for both written and spoken forms. Critically, although response times on the recognition memory task indicated overnight consolidation of visual representations for trained words from day 1 we do not show evidence that consolidation of auditory representations has not also taken place. Comparison between visual and auditory response times did not reveal the interaction that would support different processes of consolidation across modalities. Thus, the lack of significant differences for day 1 words in the auditory recognition memory task may reflect a lack of power rather than the absence of an effect.

4.14 Overall discussion

The key question of this chapter was whether overnight consolidation might support the learning of novel orthographic representations, and if so, whether consolidation would take place at a lexical or sublexical level. Across three experiments we tested this question by examining learning and consolidation at the level of letters, bigrams, and whole words.

Overnight consolidation of letters. All three experiments showed evidence for consolidation at the level of letters. This effect was most clearly seen when endorsing matching probes in the matching task where all three experiments showed faster responses to trained and untrained words written in the day 1 orthography. Moreover, whereas the stimuli in Experiments 2 and 3 were kept identical by design in this case a new set of spoken items was used, offering compelling evidence that the effect generalises to a novel set of stimuli. The magnitude of decrease in response time for day 1 orthography was similar for Experiments 2 and 4 (57*ms* and 53*ms* respectively). In Experiment 3 by contrast the reduction was 111*ms* faster for consolidated letters. Experiment 3 used the experimental stimuli as Experiment 2, while maintaining the same training regime as Experiment 4: 12 presentations of 16 words on each day. One explanation for the difference in magnitude of outcomes is that Experiment 3 has a greater proportion of untrained words: 1 : 4 trained to untrained in Experiment 3 compared to 1 : 1 in the other experiments. Faced with a situation where whole word knowledge was of little value in identifying the written word participants may have shifted towards a serial decoding strategy. Hence, the faster processing of day 1 orthography is more prominent. Nevertheless the consistency of this finding across three experiments provides strong evidence for memory-related changes playing a key role in the early stages of learning to read.

One interpretation of these improvements might be that participants are better able to identify and distinguish the novel letters after a night of sleep. This interpretation is consistent with accounts of visual consolidation indicating both visual discrimination (Karni and Sagi, 1993; Stickgold et al., 2000) and an improved ability to segment complex visual objects (McDevitt et al., 2014) after a night of sleep. The overnight improvements are not purely visual however, the matching task requires participants to make judgements about both written and spoken forms of the words. Alternatively the improvements seen here might be a reflection of improved ability to generalise knowledge after sleep (Davis and Gaskell, 2009; Tamminen et al., 2012, 2015). Across these studies participants were trained using a reading task and testing with a combination of the reading and matching tasks. The fact that the reading task is a familiar context may explain why the effects of consolidation are only apparent on the matching task; a novel task on which the participants have not had practice. These two alternative accounts could be resolved by including a matching task during training on both days as well a during testing. If letter-level improvements were due to generalising to a novel task then the consolidation effects should disappear.

Framed in terms of the triangle model of reading the rapid (overnight) changes seen here seem consistent with the finding that the systematic ($O \mapsto P$) connections can be learned more quickly than the arbitrary mappings from ($O \mapsto S$) mappings (Harm and Seidenberg, 2004). The fact that overnight changes take place at the level of letters rather than words may therefore be unsurprising. The letter-sound mappings of the stimuli used in these three experiments involved consistent letter-sound mappings and so may have promoted letter-level consolidation. It is important to note however that we have not demonstrated overnight improvements in ($O \mapsto P$). These data can be interpreted purely as reflecting improvements in visual processing, thus, consolidation within the orthographic (O) element of the model. Nevertheless, the data shown here provide a novel demonstration that overnight consolidation plays an important role in learning a novel orthography.

Little evidence for consolidation of bigrams. The question of whether learning takes place at the bigram level is less clear cut. Experiment 2 showed evidence for bigram learning in the form of faster responses to probes written in a consistent orthography. There, trained words that were visually familiar drew faster responses than trained words that were visually unfamiliar. This effect was confounded with day however and so it was not clear whether overnight consolidation was impacting on learning. Experiment 3 tested whether overnight consolidation would support learning and generalisation of bigrams learned incidentally during training on day 1. Results from the reading task showed no effect of bigram knowledge, nor was there an effect when endorsing matching probes. Participants were slower however to reject mismatching probes when the probe contained a familiar bigram. Bigram learning may therefore have taken place but again there was no indication that overnight consolidation supported learning at this level. This outcome may not be surprising given the training situation here. There is a consistent mapping between graphemes and phonemes, and during testing the majority of words are unfamiliar. Participants may therefore have discovered that the most effective ‘grain size’ of learning is between letters and sounds (Ziegler and Goswami, 2005). This concern was addressed in Experiment 4 where a skewed frequency distribution of bigram mappings was included to encourage learning at this level. While there was no significant effect of bigram frequency on either the reading task or the matching task, the inclusion of frequent bigrams during training may explain the apparent discrepancy between the large differences in reading latency shown here and the lack of differences in Experiments 2 and 3. It may be that the inclusion of frequent bigrams during training has prompted learning at larger grain sizes and thus simplified the process of learning the letters. Participants would then have had more opportunity to form representations of the letters. Improved letter learning alone is not sufficient to explain the faster reading latencies for day 1 words however, as training-related improvements would apply to letters learned on both day 1 and day 2. Instead, it may be that the improved learning led to letter representations that were strong enough to encourage overnight memory consolidation, and thus faster reading latencies. As noted already the consistent mappings between letters and sounds throughout these experiments may naturally discourage bigram level consolidation. Artificial orthography studies have demonstrated that learners are sensitive to intermediate orthographic representations such as bigrams in situations where they support accurate reading (Taylor et al., 2011). Including inconsistent mappings (i.e., irregular spellings) in the overnight consolidation framework used here may thus provide an additional method of testing whether overnight consolidation supports bigram learning.

Consolidation of whole words? Evidence for whole word consolidation was generally absent across the three experiments despite the inclusion of a variety of measures. The absence of this predicted outcome is striking in light of the extensive evidence that participants are learning the whole words. The matching tasks across all three experiments, as well as the reading tasks in Experiments 2 and 4, showed participants were faster to identify/read aloud trained words than untrained words. Nevertheless, the learned lexical items from day 1 showed no benefits of consolidation. One the proposed reasons for the lack of lexical consolidation was that participants' learning of the whole words was not strong enough, as shown by low d' scores for the trained words. Despite increased training and more distinctive stimuli in Experiments 2, 3, and 4, recognition memory remained low. This explanation seems to fly in the face of the improvements for trained words over untrained words however. It may be that the strength of memory representation required for overnight consolidation is greater that that required to see improvements in the tasks used here. As noted above, future work could include recognition memory tests for the letters as well as words. This would allow us to assess the importance of strength of memory in determining the locus of consolidation effect.

An alternative reason for the lack of whole word consolidation is the lack of memory schema in this orthographic setting. Learning and consolidation are influenced by memory schema; pre-existing knowledge. Learning is easier and quicker when the new information is consistent with pre-existing knowledge (Lindsay and Gaskell, 2010; Tse et al., 2007). Notably the impact of pre-existing knowledge on consolidation of novel linguistic information has been found in spoken language (Havas et al., 2017). Havas et al. taught participants words that resembled familiar words in either their L1 or L2 and in the context of either familiar or unfamiliar objects. When tested after a period of sleep participants were more accurate to recognise novel words that had been learned in association with a familiar object. In the case of Bowers et al. (2005) then, the impacts of memory consolidation may be more readily detectable due to the similarity between the novel and existing words (e.g., 'banara' and 'banana'). In the case of the novel written words used here the orthography was also novel and so orthographic neighbours could not have supported learning and consolidation. Thus, the lack of supporting schema might prevent lexical consolidation.

Evidence for overnight consolidation of lexical forms appears in two places in these experiments: faster recognition memory for day 1 spoken forms in Experiment 3, and faster recognition memory for the day 1 written forms in Experiment 4. In both cases however, the difference in response times between written and spoken judgements ('the difference in differences') was not significant itself. Consequently, it may be that both experiments are sensitive to consolidation of both written and spoken forms. Higher confidence ratings in the two experiments for both written and spoken words from day 1 supports this interpretation.

Despite the evidence for whole word consolidation on the recognition task, performance on the reading and matching tasks showed no benefits for trained and consolidated whole words. Experiment 1 also showed apparent evidence for whole word consolidation in the form of reduced hippocampal responses but no improvements in the reading task. It may be that although representations of whole words are forming, these representations are not yet strong enough to impact the as yet serial decoding process. In comparing DRC and

Triangle accounts of reading Coltheart (2006) suggests lexical and sublexical orthographic representations may develop in parallel, rather than whole word representation emerging some time after readers have learned letter representations. While the data here do suggest lexical and sublexical representation both develop early in the reading process, the fact that only sublexical knowledge influences performance on reading tasks argues for orthographic knowledge forming in stages. Another possibility is that the reading and matching tasks here naturally encourage participants towards componential processing. A recent study by Taylor et al. (2017) used reading tasks that focussed either on mapping from print to sound, or from print to meaning. The results showed that task demands play an important role in determining reading strategy. It may be the case that by including a semantic element to the training might encourage lexical rather than sublexical learning, and thus overnight lexical consolidation.

Alternatively consolidation may not have impacted on the mappings between orthographic and phonological lexical knowledge. The recognition memory task differs from both the reading and matching tasks in that it does not require participants to use orthographic and phonological knowledge to make judgements. In the reading task participants must use orthographic knowledge to produce the spoken form of the word. In the matching task the spoken target must be recoded to orthographic form. The pattern seen here may therefore reflect a lack of consolidation in the mappings between orthography and phonology. This interpretation with findings of Bakker et al. (2014) that lexical consolidation across modalities requires longer periods of consolidation than lexical consolidation within modality.

Future directions. Together these three experiments demonstrate the value of combining artificial orthography learning with an experimental approach that highlights the processes of memory formation. With this framework in place a number of additional studies are possible.

First among these is whether overnight consolidation supports the learning of letter-sound mappings. The three studies presented in this chapter do not answer the question of whether the consolidation effects found here relate purely to visual processing speed, or consolidation of the mappings from letters to sounds. Some evidence, in particular the recognition memory results from Experiment 4, indicate whole word recognition may also be forming overnight. Nevertheless it may be the case that skill in visual recognition of both letters and words develop in parallel, while fluency in letter-sound processing takes longer to come online. The experimental approach used here offers a clear method by which to test this question. Adapting the design would allow letter-sound mappings to be trained and tested with or without a period of sleep. This approach could also investigate the suggestion by Coltheart

(2006) that whole-word representations develop in parallel to knowledge of letter-sound mappings.

The question of whether overnight consolidation aids learners to identify the most consistent grain sizes in the orthography is also relevant both to cognitive and educational questions in reading. Ziegler & Goswami (2005) argue that learners are sensitive to orthographic consistency when learning to read, and that the inconsistencies of English grapheme-phoneme correspondences leads to reliance on larger grain sizes. Indeed, Taylor et al. (2011) demonstrate that learners are particularly sensitive to orthographic consistency when learning to read a novel orthography. Moreover, both the consistent/mixed orthography condition of Experiment 2, and the results of Experiment 4 suggest learners are sensitive to orthographic consistency early in the learning process. An important difference between the triangle and DRC models is in how they deal with regularity/consistency (Zevin & Seidenberg, 2006; Taylor, Rastle, Davis, 2014). The triangle model succeeds in accounting for the spelling-sound consistency effects. Adapting the framework of overnight consolidation to expand on the inconsistent mappings used by Taylor et al. (2011) may therefore offer useful insights into how readers learn which orthographic grain size is most useful.

Finally, despite the evidence here for consolidation of written word forms, d' scores remain relatively low for trained words from days 1 and 2, and the difference between d' scores between trained items from days 1 and 2 was not significant. It may be that the lexical forms learned here are still not distinctive enough, even after the changes made to the stimuli and to the training regime. One way in which to address the question of whether distinctiveness plays a role in encouraging overnight consolidation would be to shift the ratio between words and letters, or to give the words meanings. In the current studies we carefully balanced the items to ensure there was a variety of both letters and words. Adapting the stimuli to contain a smaller number of trained words each day may influence whether lexical-level consolidation occurs.

4.15 Conclusion

The goal of this chapter was to track the initial acquisition of lexical and sublexical orthographic knowledge, and to assess whether overnight consolidation may support either of these processes. While phonics-based teaching methods emphasise letter-level decoding as the primary process of early reading, it is unclear when and how whole-word recognition begins to emerge. Evidence for overnight consolidation at the level of letters across three behavioural experiments stands as additional evidence for the primacy of letter-level learning in reading acquisition. Together these data show for the first time effects of overnight consolidation in the context of learning a novel orthography. The ability to generalise componential letter-sound mappings is the critical element that makes reading so powerful. Consequently, the demonstration that overnight consolidation of sublexical knowledge leads to improved ability to use and generalise orthographic knowledge lies at the heart of learning to read. Moreover, these data extend the findings of the fMRI experiment by again highlighting that holistic (lexical) and componential (sublexical) learning may involve difference memory systems, and hence, have different learning trajectories. These data demonstrate the value of laboratory-based methods in capturing these processes in fine detail and motivate the further use of consolidation-related paradigms to investigate literacy acquisition.

Chapter 5

The impact of spoken word consolidation on learning a novel orthography.

1. This chapter explores the influence of consolidated and unconsolidated spoken language on learning to read.
2. In Experiment 5 recently learned words supported the acquisition of a novel orthography but this effect did not appear in Experiment 6.
3. However, Experiments 5 and 6 show two failed attempts to observe spoken word consolidation. Potential reasons for this outcome are discussed.

5.1 Background

Children typically have a well-developed spoken lexicon before they start the task of learning to read. Consequently, the goal of decoding a written word is to access a familiar spoken word. Existing lexical knowledge thus supports literacy acquisition (Nation and Cocksey, 2009). This situation contrasts with the experimental approach used in Experiments 1 - 4 where participants learn to decode unfamiliar spoken words in a novel orthography. Even when participants successfully translate from artificial letters to sound in these experiments, they are not accessing a meaningful word. The finding of orthographic consolidation in Experiments 2, 3, and 4 may thus reflect the focus of the experiment on written, rather than spoken, word learning.

This chapter differs from the previous experiments by measuring the impact of spoken language consolidation on learning a novel orthography. In tackling this question we aim to engineer a learning situation that is more similar to that experienced by children when learning to read. Here participants learn novel spoken words over two days before learning to read a novel orthography on the second day. We hypothesise that trained and consolidated words may be easier to recognise during the process of learning to read, and thus to play a role in supporting early decoding attempts.

Lexical knowledge and learning to read. While the core skill of reading is the ability to map from letters to sounds (Byrne, 1998; Ehri, 1995; Share, 1995), lexical knowledge also plays an important role in supporting reading (Nation, 2017; Nation and Cocksey, 2009; Nation and Snowling, 2004; Stuart et al., 1999). Indeed, Gough and Tunmer (1986) characterise reading as a reciprocal process between oral vocabulary and knowledge of letter-sound mappings. Perfetti meanwhile argues "[i]n reading, the singular recurring cognitive activity is the identification of words" (2007, p357). Indeed, oral vocabulary skills contribute to word reading and comprehension independently of decoding skills (Ricketts et al., 2007) and support the development of word recognition skills (Nation and Snowling, 2004).

Participants' lack of spoken lexical knowledge in the previous experiments may thus have influenced the pattern of results. For example, Ziegler and Goswami (2005) point to learners becoming sensitive to the grain size that provides the closest correspondence between print and spoken language. In Experiments 2 - 4 the most consistent mapping is at the level of graphemes and phonemes; the pseudowords were not taught in a meaningful way. It may not be surprising then that the consolidation effects noted thus far seem to take place at the orthographic level. Notably, Experiment 1 taught participants both words and objects on day 1, showing reduced hippocampal (HC) responses on day 2. One possible explanation of these reduced HC responses is that participants have learned and consolidated the spoken forms of both words and objects. Trained and consolidated spoken words might, therefore, support early decoding attempts.

Lexical knowledge in the triangle model. The triangle model of reading offers a useful computational framework for understanding these behavioural findings. While the original Seidenberg and McClelland (1989) model implemented only the orthographic (O) and phonological (P) elements of the model, later implementations have emphasised semantic (S) contributions to learning in the system (Harm and Seidenberg, 2004; Plaut et al., 1996). The emergence of direct mappings in the model from ($O \mapsto S$; identifying a written word without having to decode its phonology) is supported by the pre-existing mappings between sound and meaning ($P \mapsto S$). Put another way, the lexical knowledge captured in the ($P \mapsto S$) mappings support ($O \mapsto S$) learning. It may thus be the case that decoding the phonological form of a familiar spoken word may be a first step to whole word recognition. This view is borne out in a recent connectionist model that asked whether training spoken words prior to training the ($O \mapsto P$) mappings would support the model in learning to read (Chang et al., 2017). Only when pre-trained with spoken language the model was able to extend its knowledge of letter-sound mappings to perform well on semantics reading tasks.

In the context of the current study, learning whole-word forms may, therefore, be sensitive to whether participants have a representation of the spoken word form and meaning. The previous experiments explored the impact of overnight consolidation in the formation of novel orthographic representations; Experiments 2, 3, and 4 showed evidence for consolidation at the level of letters. None of these experiments isolated the impact of prior knowledge of either phonology or semantics, however. While written forms were trained in these experiments, they corresponded to no known spoken form. The consolidation effects seen in these experiments could be interpreted purely in terms of learning in the orthographic elements of the model. Here we attempt to investigate consolidation in the phonological and semantic sides of the triangle and ask whether this knowledge might in turn influence the orthographic side.

Building on literature showing consolidation of spoken words. Another motivation for this approach in the thesis is to link more directly to previous experiments on spoken word consolidation. A large body of literature on linguistic consolidation suggests that, after a period of overnight consolidation, recently learned spoken words may become lexicalised and integrated with wider lexical knowledge (Davis et al., 2009; Dumay and Gaskell, 2007; Gaskell and Dumay, 2003; Merkx et al., 2011; Tamminen et al., 2012). Incorporating this approach to lexical consolidation may allow the inclusion of lexicalised novel spoken words in an experiment about orthographic learning, while still maintaining tight experimental control over all stimuli. Notably, an artificial orthography study by Taylor et al. (2017) adopted a comparable approach by pre-training novel words (i.e., $P \mapsto S$ information) one day prior to the initial orthographic training session. Overnight consolidation was not the focus of this study however and so the design does not indicate whether consolidated novel spoken words provide an additional support to orthographic learning.

Adapting these consolidation paradigms to the experimental context used here makes clear that some measures of consolidation would not be viable. For example, the most common index of lexicalisation is the competition effect shown in Gaskell and Dumay (2003). This effect is reliant on finding a sparse lexical neighbourhood and measuring the impact of a recently learned word on that neighbourhood. After a new neighbour becomes lexicalised the time required to distinguish between these two neighbours increases, as predicted by the distributed cohort model (Gaskell and Marslen-Wilson, 1997). The stimuli used here attempt to maintain a statistical balance of component parts (phonemes or graphemes depending on modality). Consequently, they are incompatible with trying to find the lexical isolates required for competition tasks and so we adapt the approach used by (amongst others) Davis et al. (2009): phoneme monitoring for training and delayed repetition to test consolidation effects. The first experiment of Davis et al. (2009) presented novel spoken words while participants listened for various target phonemes. Participants thus heard each novel word 12 times while listening for 6 phonemes twice each. To measure changes in the representations of words participants completed a spoken repetition test in which they heard and then repeated novel words. Three conditions of words were used, trained and consolidated, trained and unconsolidated, and untrained. After hearing the spoken word a tone appearing at variable intervals prompted participants to respond and the latency of responses was used as the dependent measure. Critically the results showed significant improvements in repetition latency for novel words that had been trained and consolidated compared to the other two conditions.

In summary, the combination of an artificial reading task with a spoken word consolidation paradigm may allow us to engineer a situation akin to that faced by children. Children when learning to read have to decode an unfamiliar orthography in order to access familiar spoken words. By training participants to recognise the phonological forms of spoken words over two days, we can ask whether learning a novel orthography is easier when supported by knowledge of consolidated spoken words. In other words, does knowing a word help you to read it? Thus, the artificial reading paradigm allows us to go a step further in investigating how orthography, phonology, and semantics interact when learning to read.

5.2 Experiment 5

The primary goal of this experiment was to assess whether it would be easier to learn a novel orthography in the context of consolidated or unconsolidated spoken words. A secondary goal of the experiment was to directly tie our research on orthographic consolidation to more established findings in the domain of spoken word consolidation. To this end, we teach participants to recognise novel spoken words using a phoneme monitoring task that has been used in studies of spoken word consolidation (e.g., Davis et al., 2009; Gaskell and Dumay, 2003). Following the train-twice, test-once approach used thus far, participants were exposed to two sets of spoken words over the course of two days using the phoneme monitoring task. Critically, the spoken words learned on day 1 and day 2 did not share any phonemes; if the phoneme /b/ appeared in several words learned on day 1, it would not appear in any words heard on day 2. Following training on day 2 spoken word consolidation was assessed using a repetition task (Davis et al., 2009).

Participants then learned to read a novel orthography by reading trained and untrained words. Because words from days 1 and 2 shared no phonemes, the set of letters that were mapped to these words also shared no letters. Consequently, participants in effect learned two novel orthographies, one orthography mapping to the day 1 words, and the other orthography mapping to the day 2 words. We refer thus refer to day 1 and day 2 orthography, even though participants learn both orthographies on the second day of the experiment.

As before, the two independent variables used here are day-of-learning (the spoken words) and the training status of the words seen during training (trained or not on the phoneme monitoring task). Unlike Experiments 1 - 4, the critical measure here is the trajectory of learning. If knowledge of spoken words supports learning of the new orthography then participants should be better at reading the trained words that they heard on the phoneme monitoring task on days 1 and 2. If participants have consolidated the spoken words trained on day 1 then we may see improved performance when learning to read the novel orthography that maps to the day 1 words. We additionally include the spoken-to-written matching task used in Experiments 2, 3, and 4. This measure allows both an additional chance to measure any differences that may occur due to learning and consolidation of spoken words and makes the outcome of this experiment comparable to the previous findings.

5.3 Methods

5.3.1 Participants

18 native English-speaking adults aged 18 – 40 took part in a study approved by the Cambridge Psychology Research Ethics Committee. No participants reported having dyslexia, speech, or language impairment. One participant was excluded from the analysis for failing to learn (below 30% correct on the reading task) leaving 17 participants in the analysis.

5.3.2 Materials

The same four sets of 16 monosyllabic CVC pseudowords were used in Experiments 2 and 3 were used for consistency (see Figure 5.1, p151 for sample stimulus assignment). Thus, two of the pseudoword sets contained 8 phonemes (b, f, n, v, s, θ, ʤ, k) and 4 vowels (æ, ɛ, ɒ, ʌ). The two remaining stimulus sets contained 8 further phonemes (d, m, p, z, t, ʃ, h, tʃ) but shared the same 4 vowels (æ, ɛ, ɒ, ʌ). As before, three of these phonemes mapped to digraphs in English orthography (/tʃ/ $\mapsto$ ⟨*ch*⟩, /ʃ/ $\mapsto$ ⟨*sh*⟩, and /θ/ $\mapsto$ ⟨*th*⟩) but were included to allow for a wider item set containing more distinctive lexical items. Each of the four sets was assigned to trained or untrained conditions on either day 1 or 2 in a counterbalanced manner across participants. As with the previous experiments the consonant frequency was matched across stimulus groups with each consonant appearing twice in onset and twice in coda. Unfamiliar visual symbols were mapped to each of the phonemes in a one-to-one manner. Note that as the orthographic learning in this experiment took place on day 2 of the experiment we do not include 2 sets of vowel characters as in Experiments 2, 3, and 4.

5.3.3 Procedure

A full timeline of the experiment can be found in Figure 5.2 (p152). All tasks were presented using custom scripts in Psychopy, a python-based software package for psychophysics (Peirce, 2007, 2009).

Phoneme monitoring. Adopting the phoneme monitoring procedures used elsewhere (Davis et al., 2009; Gaskell and Dumay, 2003; Hawkins et al., 2015; Snoeren et al., 2009) participants heard 16 spoken pseudowords on each day. Participants were asked to press a button when they heard a target phoneme, with each of the 8 ‘consonants’ appearing as the target in different blocks, one phoneme per block. Before each block, a written cue indicated which phoneme participants should listen for (see Figure 5.2, p152). The written cue also

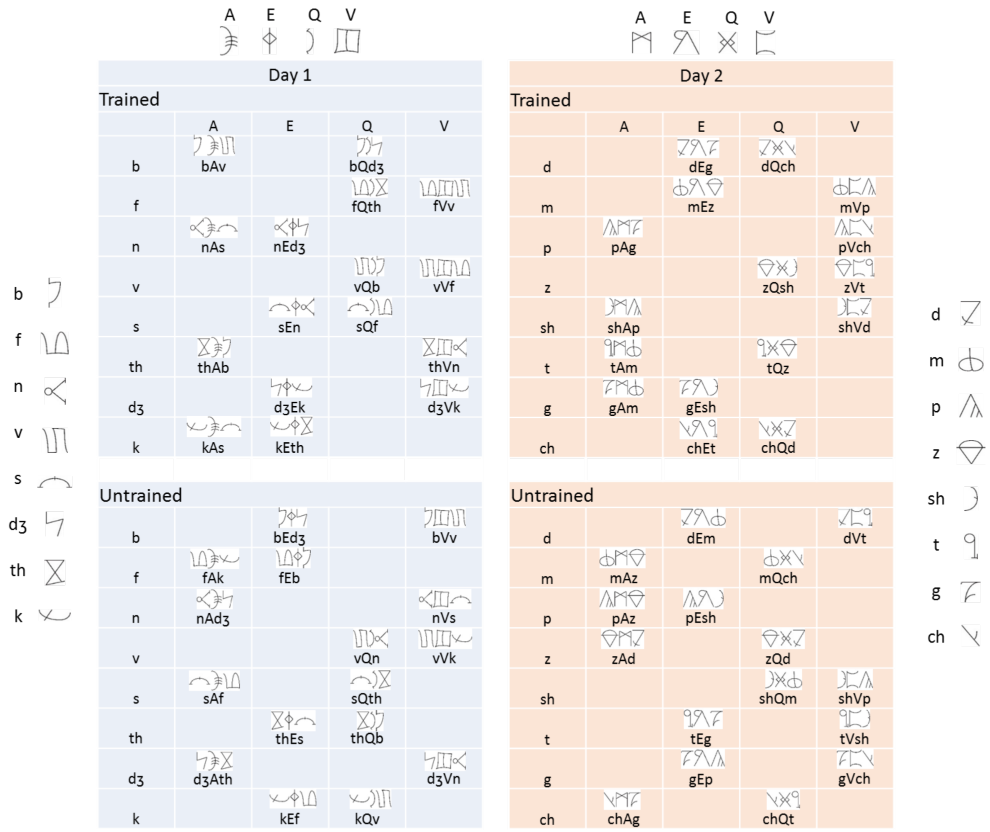


Fig. 5.1 An example assignment of the 64 pseudowords to be trained or untrained on day 1 or 2. 16 items are trained each day, retaining a further 16 untrained. The rows show the initial phoneme while the columns show the vowels. The spoken forms are learned on day 1 and day 2. The phonemes that make up the items for each day are distinct; a phoneme that appears in a day 1 word will not appear in a day 2 word. Thus when participants learn to read these words they in effect learn two orthographies that share vowels.

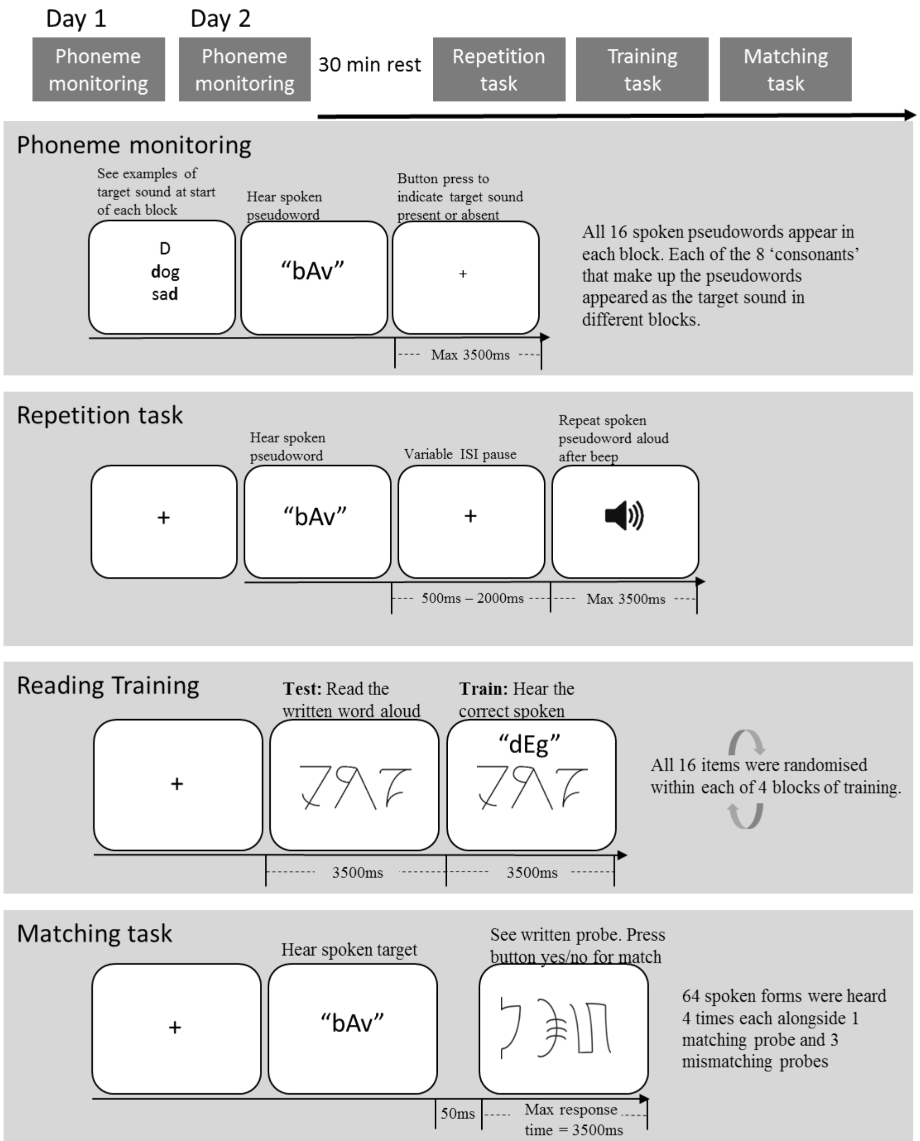


Fig. 5.2 Experiment 5: Timeline and procedures

showed two example words containing the phoneme in onset and coda positions. In each block, participants heard all 16 items in a randomised manner. There were 4 runs with each phoneme appeared as a target once per run, thus 512 trials per participant per day (16 words for each of 8 target phonemes over 4 runs). Each spoken word, therefore, appeared 32 times during training. Due to the statistically matched nature of the artificial items, each phoneme was present twice in onset and twice in coda on each run, leading to four 'hits' in each run of 16 items. Participants had up to 3500ms to respond on each trial and received no feedback on accuracy.

Spoken word repetition. Following training on day 2 participants had a 30-minute rest break followed by a spoken word repetition test to assess consolidation of the trained spoken words. In order to assess whether consolidation of day 1 trained words had taken place participants completed a spoken repetition task to assess consolidation effects (Davis et al., 2009). On each trial, participants heard one of the 64 spoken items (trained/untrained, day 1/day 2) followed by a beep that occurred with a variable ISI between 500 and 2000ms. Participants were instructed to wait for the beep and then to repeat the word as quickly as possible. Responses were recorded and speech onsets were marked by hand offline using CheckVocal (Protopapas, 2007).

Orthographic training task. The critical question in this experiment was whether hearing and consolidating a spoken word would help to read and recognise that word when learning to read a novel orthography. Participants were trained to read the novel orthography using the same training paradigm used in Experiment 2. On each trial, participants saw the written word onscreen and tried to read it aloud. After 3500ms they heard the correct spoken form of the word and then had a further 3500ms in order to learn. Unlike Experiments 2 and 3 where participants learned 16 words each day (made up of 12 graphemes) at a time, participants here learned to read all 64 items (made up of 20 graphemes: 16 consonants and 4 vowels). All 64 items (heard and unheard spoken words from days 1 and 2) were presented in a fully randomised manner in each of 4 training runs.

Matching task. As in the previous experiments, the matching task allowed a measure of lexical and orthographic learning that was not dependent on a verbal response. Participants heard the spoken form of one of the 64 items (heard/unheard, day1/day2), with a visual probe appearing 50ms after the end of the spoken item. Participants pressed a button to indicate whether the spoken target matched the written probe. Response times were measured from the onset of the written probe and to ensure speeded responses an upper limit of 3500ms was imposed. All 64 pseudowords were used as spoken targets.

Note that the matching task used a different probe structure to the previous experiment. In order to avoid the potential confounds caused by the mixed orthography participants did not see mixed orthography probes. There were three probe types: matching, mismatching in onset, and mismatching in coda. We further introduced the manipulation that roughly half (44.53%) of the mismatching probes had been seen by the participants. This was to avoid participants rejecting mismatching items purely on the basis of visual familiarity.

5.4 Results

5.4.1 Phoneme monitoring

Accuracy on the phoneme monitoring task was very high across both days [$Day\ 1: M = 95\%,\ SD = 20.88\%;\ Day\ 2: M = 95\%,\ SD = 20.78$]. In order to account for potential response bias d' scores were calculated for each participant and day. Across the group d' was very high [$Day\ 1: M = 3.38;\ Day\ 2: M = 3.48$] but more importantly no participant had a d' of less than 3 for both days [$Min = 3.01,\ Max = 3.76$]. A paired t-test on the d' scores showed participants were equally attentive on days 1 and 2, [$t_p(15) = 0.50,\ p = 0.62$].

5.4.2 Spoken word repetition

In order to test whether overnight consolidation has taken place participants completed a spoken word repetition task. Participants repeated aloud both trained and untrained words from days 1 and 2 (Figure 5.3, p155, Table 5.1, p155). As expected accuracy was extremely high: [$M = 99.63\%$]. Response times were compared using a 2 x 2 repeated measures ANOVA with the factors of day-of-learning and training (trained or not). There was no significant effect of day, [$F(1,16) = 0.71,\ p = 0.16$], or training, [$F(1,16) = 2.14,\ p = 0.41$], and no interaction between the two factors: [$F(1,16) = 0.01,\ p = 0.91$].

5.4.3 Orthographic training

The critical measure in this experiment was whether knowledge (consolidated or otherwise) of spoken language influences the ease with which participants can acquire a novel orthography. Figure 5.4 (p156) shows accuracy across the four runs of training (Table 5.2, p156). A repeated measures ANOVA used the proportion of correct responses as the dependent variable, with a 2 x 2 x 4 design including the factors of day (heard on day 1 or day 2),

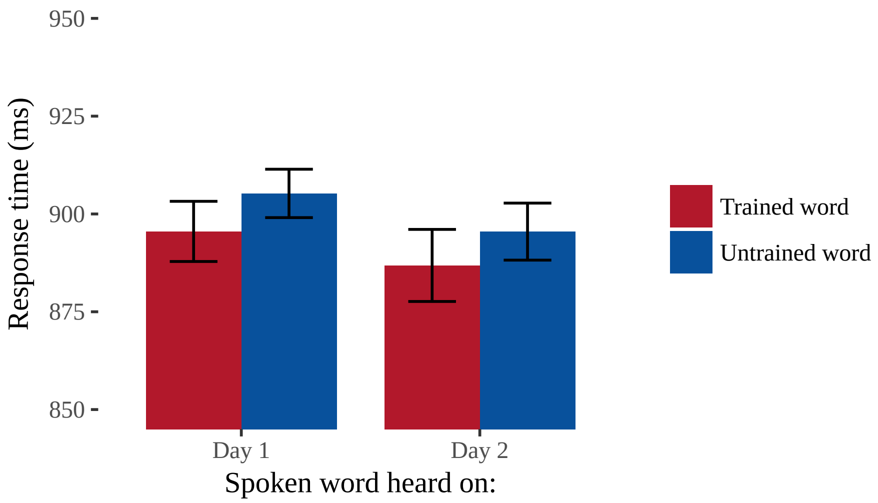


Fig. 5.3 Response time to repeat aloud trained and untrained words from days 1 and 2.

Table 5.1 Experiment 5: Performance on spoken word repetition task

	Response time (ms)	
	Mean	St. Dev
Day 1 trained	896	197
Day 1 untrained	905	186
Day 2 trained	887	199
Day 2 untrained	895	190

training (trained or untrained), and run (runs 1 - 4). There were main effects of both training, $[F(1,16) = 12.23,\ \eta_p^2 = .43,\ p = 0.002]$, and run, $[F(3,48) = 146.31,\ \eta_p^2 = .90,\ p < 0.001]$, but no main effect of day, $[F(1,16) = 0.003,\ p = 0.96]$. There were no significant two way interactions, $[day{*}training : F(1,16) = 1.44,\ p = 0.25; day{*}run : F(3,48) = 2.16,\ p = 0.11; training{*}run : F(3,48) = 0.27,\ p = 0.85]$. Finally, the three way interaction between day, training, and run that would suggest improved performance due to consolidation of trained words from day 1 was not significant, $[F(3,48) = 1.26,\ p = 0.30]$.

5.4.4 Matching task

Accuracy. For the primary analysis of matching data, the four target conditions were the primary interest. Data were entered into a 2 × 2 repeated measures ANOVA with the factors day of orthography (day 1 or day 2), and training (trained or untrained). The left panel of Figure 5.5 shows accuracy was again high in the matching task $[day\ 1\ trained : 92.93\%$,

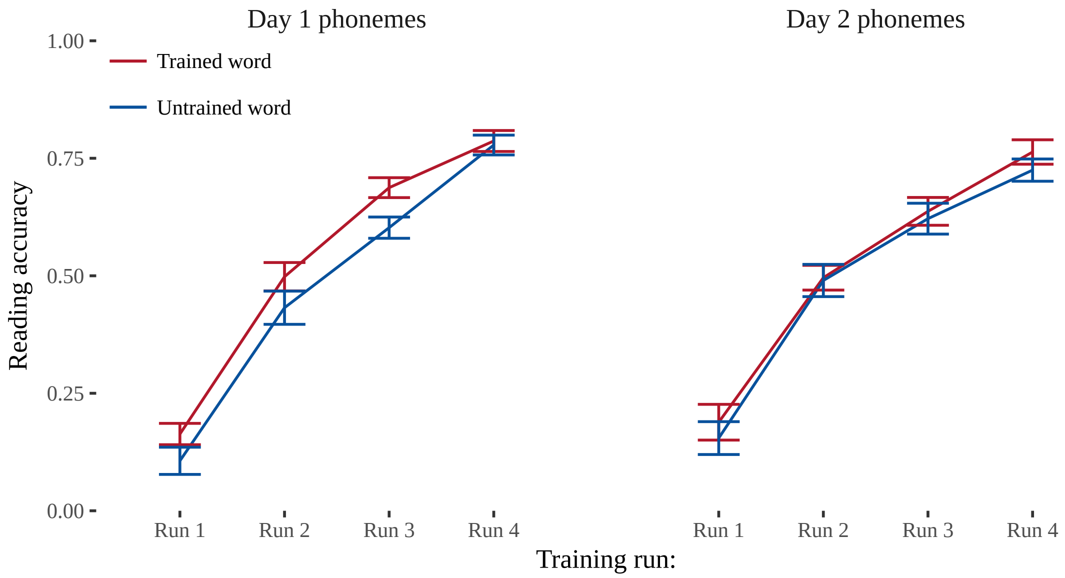


Fig. 5.4 Performance when learning to read the two orthographies on day 2. Note that day 1/day 2 refers to when the participants heard the spoken forms of the words. All learning took place during the same training session, with day 1 and 2 phonemes separated here for clarity.

Table 5.2 Accuracy during training

	Mean (*SD*)			
	Run 1	Run 2	Run 3	Run 4
Day 1 trained	16.32% (37.05%)	49.79% (50.10%)	68.75% (46.44%)	78.68% (41.03%)
Day 1 untrained	10.63% (30.89%)	43.20% (49.63%)	60.23% (49.04%)	77.82% (41.62%)
Day 2 trained	18.84% (39.20%)	49.59% (50.10%)	63.71% (48.18%)	76.34% (42.58%)
Day 2 untrained	15.46% (36.25%)	49.00% (50.10%)	62.15% (48.60%)	72.48% (44.75%)

$day\ 1\ untrained : 91.11\%$, $day\ 2\ trained : 91.86\%$, $day\ 2\ untrained : 91.91\%]$. Incorrect responses were discarded and response times for the remaining data were analysed. Accuracy did not differ across the four conditions: no main effect of day, $[F(1,16) = 1.11,\ p = 0.31]$, no main effect of training, $[F(1,16) = 2.29,\ p = 0.15]$, and no interaction, $[F(1,16) = 0.27,\ p = 0.61]$.

Response times to matching probes. As with the previous experiment, we collapse matching and mismatching for the primary analysis of response times. Figure 5.5 (p157, right panel) shows the effect of day and training on response times. A repeated measures ANOVA showed there was no main effect of day, $[F(1,16) = 0.66,\ p = 0.43]$, no main effect of training,

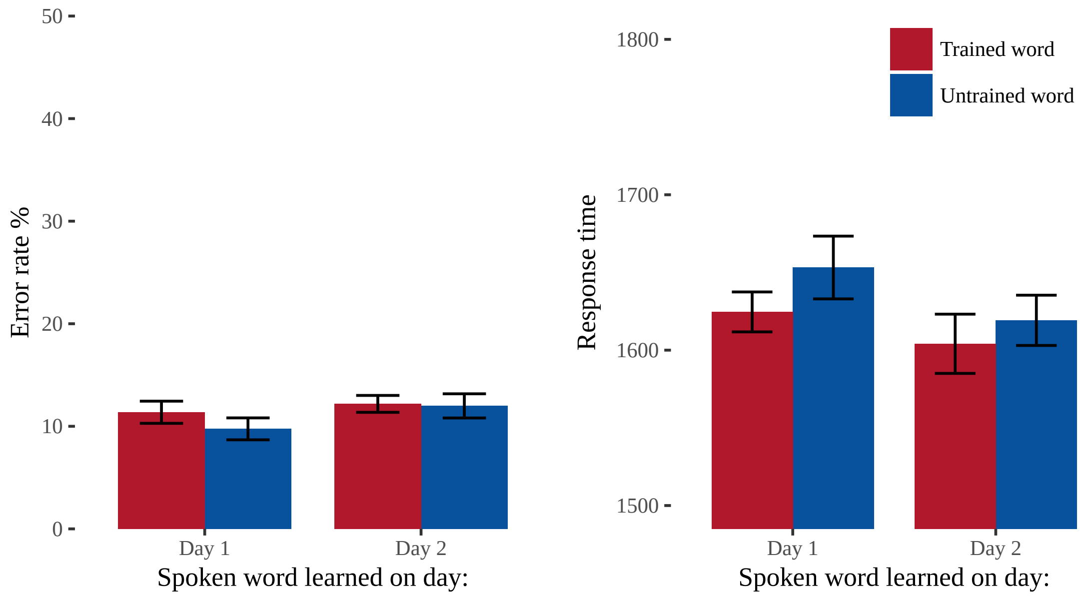


Fig. 5.5 The left panel shows errors rates on the matching task. The right panel shows response times for trained and untrained spoken words from days 1 and 2. Note: response time figure includes correct responses only.

$[F(1,16) = 1.58,\ p = 0.23]$, and no interaction, $[F(1,16) = 0.79,\ p = 0.79]$. As matching and mismatching conditions likely tap into different cognitive processes, with matching trials giving a better assessment of word learning, we analysed the data including only the matching trials. As before there were neither main effect of day, $[F(1,16) = 0.06,\ p = 0.81]$; main effect of training, $[F(1,16) = 0.23,\ p = 0.64]$; nor interaction, $F(1,16) = 0.05,\ p = 0.37$.

Response times to mismatching probes. Following the rationale used in the previous experiments, the mismatching probes assessed whether knowing a spoken word influenced reading strategy (Figure 5.6, p158) Response times to detect errors were entered into a 2 x 2 x 2 ANOVA with the factors day (day 1 or day 2), training (trained or untrained), and error position (onset or coda error). As was the case in the previous experiments errors in onset were detected faster that errors in coda, $[F(1,16) = 52.47,\ \eta_p^2 = .77,\ p < 0.001]$. However there was no effect of day, $[F(1,16) = 1.88,\ p = 0.19]$, or training status, $[F(1,16) = 1.36,\ p = 0.26]$. There were no significant interactions.

5.5 Discussion

The primary goal of this experiment was to replicate findings of overnight consolidation of spoken words in a novel setting, and using these changes as an experimental factor to assess

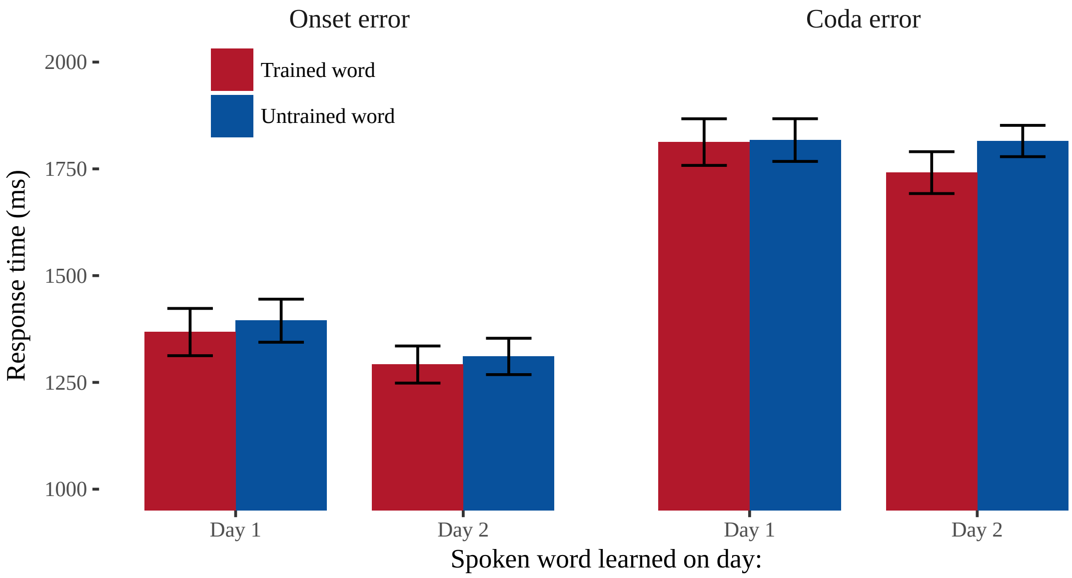


Fig. 5.6 Response times to identify errors in onset and coda positions.

the role of spoken language when learning to read a novel orthography. In summary, the results show that knowledge of recently learned spoken words supports orthographic learning. The strength of these findings however is weakened by the lack of replication of spoken word consolidation.

5.5.1 Failure to observe consolidation effects in positive control.

Previous studies have demonstrated overnight consolidation of novel spoken words using a combination of phoneme monitoring for training and a spoken word repetition task for testing (Davis et al., 2009; Gaskell and Dumay, 2003; Hawkins et al., 2015; Leach and Samuel, 2007; Snoeren et al., 2009). We adopted the same approach in order to replicate previous findings of overnight lexical consolidation. In this sense, the task functions as a positive control in the current experiment; a manipulation where we expect a specific outcome and so a failure to find that outcome might indicate problems in the experimental design. The data here not only fail to replicate but actually trend in the opposite direction, with (statistically non-significant) faster responses to recently learned words on the delayed repetition task (Figure 5.3, p155).

There are a number of reasons why a failure to observe overnight consolidation here may not reflect lack of a real consolidation effect but rather conditions specific to this experiment. As noted previously the pseudowords used here differ from the highly distinctive items used elsewhere, e.g., 'cathedruke' in Gaskell and Dumay (2003). Instead, all trained items are very similar. The significance of this is demonstrated by the isolation paradigm. Here, an item to be remembered, e.g. 'QXR', appears either in a homogenous or heterogenous list. If the item appears in a list of numbers (a heterogenous list) then it is likely to be remembered as a consequence of its distinctiveness. If it appears amongst a list of similar three-letter strings (a homogenous list) then it is less likely to be recalled (Hunt, 1995). Hunt's (1995) proposal is congruent with the levels of processing account of memory formation (Craik and Lockhart, 1972). In this case the perceptual distinctiveness of the novel word would induce a greater depth of processing and consequently greater recall. The lack of distinctiveness in the stimuli used here may have impeded the formation of memory traces, and so prevented consolidation of those memory traces. In addition, the pseudowords, however, occupy an extremely dense phonological space. This density may have interfered with the formation of representations of individual words within that phonological space. Moreover, while Experiments 2 - 4 taught pseudowords with similar phonological forms, those words always appeared alongside visual cues. Here there was no visual cue to distinguish items.

Alternatively, the lack of overnight changes could be due to the items used here being too different from existing words. Consolidation effects have most frequently been observed in cases where the novel word is similar to an existing word (e.g., 'cathedral' and 'cathedruke' in Gaskell & Dumay, 2003). Learning novel linguistic information is easier when the novel item is consistent with an existing 'schema' (Havas et al., 2017; Lindsay and Gaskell, 2010; Tse et al., 2007). For example, Havas et al. (2017) taught Spanish-speakers novel spoken words that either resembled words in their native language, or that were unfamiliar. These novel spoken words were also paired with pictures of either familiar or unfamiliar objects. Critically, familiarity modulated changes in representations when tested 12 hours later; both similarity to familiar spoken words and association with a familiar object aided learning. The items used in the current experiment do not by design resemble existing words. A major consequence of the failed positive control is that the central manipulation upon which all other manipulations rest is not reliable. In light of this outcome, we should not expect to see any improvements in performance for the day 1 items.

5.5.2 Effects of lexical knowledge on learning to read.

The lack of consolidation effects on the repetition task makes it unlikely that we would find improvements for day 1 words on the training task, and indeed that is the case. Participants were equally good at learning orthography associated with day 1 and day 2. Nevertheless, familiar words were significantly easier to read during training than unfamiliar words. When learning to read the novel letters and written words, participants were more accurate at reading aloud the words they had heard during the phoneme monitoring task. This finding validates the core hypothesis of this experiment: that knowledge of spoken words may support the initial stages of learning to read. These data are in line with data showing children are more accurate to read unfamiliar words if they have been familiarised with the phonological forms prior to reading the written forms (McKague et al., 2001). Here, however, the benefit of spoken word knowledge extends to decoding an unfamiliar orthography.

5.5.3 No differences on the matching task.

Confirming the absence of lexical learning effects, the matching task showed no differences in either accuracy or response time for trained or untrained words from days 1 and 2. This outcome is in contrast to the previous experiments where there were large effects of both day and training on the matching task. Indeed, the lack of differences according to day or training here provide additional support that the consolidation-related findings of Experiments 2, 3, and 4 are reliable; without the benefit of orthographic consolidation or evidence that participants have learned spoken words we see no differences between conditions. Moreover, when rejecting errors in the matching task we replicate the pattern that would indicate serial processing: faster responses to errors in onset. The data, however, do not show the overnight changes apparent in the previous experiments. The replication of the difference in response time between onset and coda errors indicates the task is measuring similar processes to before, but in this experimental context there is no lexical knowledge to influence performance.

5.6 Experiment 6

Experiment 6 attempts to address some of the potential issues raised above by attempting to make the spoken items more distinctive. In addition to hearing the trained pseudowords in a phoneme monitoring task, participants completed an object naming task in which novel objects were associated to the trained pseudowords. There are two ways in which provision of semantic information (in the form of associated objects) might influence the results: by supporting learning of spoken words, or by supporting learning of written words.

Semantics may support consolidation of spoken words. Training spoken words in association with visual objects may support learning and consolidation in several ways. First, learning the association between a spoken pseudoword and a novel object may increase the distinctiveness of the spoken word. As discussed elsewhere in this thesis the mapping between visual object and spoken word is holistic; you cannot name an object by looking at its component parts as in reading. Moreover, learning a word in a meaningful context may lead to more robust memory traces. For instance, Henderson et al. (2013) found the children had better memory for spoken words learned in a meaningful context one week after training. Associating objects with spoken words may thus encourage formation of distinctive memory traces for those spoken words, which may, in turn, lead to more pronounced consolidation effects.

Whether provision of semantic information directly supports lexical consolidation remains an open question. Hawkins et al. (2015) directly investigated how the provision of semantic information influences overnight consolidation of spoken words, training some words in context of a single picture referent. Other spoken words appeared alongside multiple pictures and so no mapping between spoken word and picture could be learned. For words trained with a single picture elicited mismatch negativity (MMN) responses, an ERP measure of lexical learning that can discriminate words from pseudowords. For unconsolidated words there was a correlation between the MMN and semantic learning during training. Following a period of consolidation this correlation was not related to semantic learning, suggesting different neural processes may have been involved for consolidated and unconsolidated words. Moreover, the impact of learning to name a novel object may be influenced by whether or not the associated objects are familiar Havas et al. (2017). Against this perspective, Leach and Samuel (2007) found that when novel words were learned alongside a picture of a novel object, lexicalisation was enhanced compared to when novel words were trained using only a phoneme monitoring task.

Takashima et al. (2014a) suggest visual object learning actually delay lexicalisation by focusing attention away from the phonological form of the word. Participants learned novel words using a phoneme monitoring task (as in Experiment 5 here) in which half of the spoken words were presented in association with unfamiliar visual objects. When tested after 24 hours lexical competition was delayed in words learned with an associated visual object. Thus, in the phonological-only condition overnight changes only occur when there is opportunity for a mapping between the novel word and an existing word. Extending these findings, Takashima et al. (2017) trained words in association with a novel visual object, a novel definition, or only trained the phonological form of the word. Behavioural testing occurred after a scanning session 8 days later and revealed improved memory for meaningful words; words learned either with a definition or an associated visual object. Nevertheless, while Takashima et al. did not observe overnight consolidation effects in the object-learning condition, memory for these items was stronger. The strength of memory for trained words was one of the potential reasons for the lack of consolidation effect. Therefore, stronger memory traces for words learned in association with novel objects may provide a sufficiently rich representation to support orthographic learning in the current experiment.

Semantics may support learning to read. In addition to potentially supporting consolidation of spoken words, there is reason to expect that including a semantic element to training may also support learning to read and recognise written words. Knowledge of word meanings is also an important contributor to reading skills (Ricketts et al., 2007).

Children's semantic skills in language support reading both familiar words and exception words with knowledge of lexical semantics contributing to children's reading skill independently of phonological skill (Nation and Snowling, 2004). Castles and Nation (2006) also emphasise the importance of lexical knowledge in supporting reading, highlighting that poor comprehenders of spoken language tend to be less successful readers. Moreover, poor oral vocabulary (a proxy for semantics) is associated with poor reading of inconsistent words (Nation, 2017; Ricketts et al., 2007; Stuart et al., 1999). Likewise, adults are more accurate to read inconsistent words aloud if those words are associated with meanings (McKay et al., 2008).

The role of semantics in reading is not clear-cut, however. Nation and Cocksey (2009) provide contrasting evidence for semantic involvement in reading, testing the impact of word-level semantic knowledge on reading. Testing 7-year-old's knowledge of phonology and semantics showed no benefit of semantics when reading individual words. Similar findings come from McKague et al. (2001) who set out to compare the predictions of the DRC and Triangle models by testing the effect of oral vocabulary when reading novel words. Here children learned oral vocabulary over several days before being tested for non-word naming performance. Children demonstrated improved performance when reading nonwords that they had heard during training. In this case, the addition of semantic training alongside the phonological training did not impact on reading performance. One important consideration when interpreting these data is that both the phonological and semantic training benefitted from several days of consolidation prior to testing reading performance. In addition participants in this study were reading nonwords in a familiar orthography unlike the totally novel orthography used here.

Nevertheless, experimental evidence from tasks similar to the current experiment seem to point to a beneficial role for semantics when learning to read. For example, Experiment 2 of Laing and Hulme (1999) demonstrated that the semantic richness of spoken words influenced how well young children could learn associations with written abbreviations. Meanwhile, Experiment 2 of Taylor et al. (2011) exposed participants to either the phonology or meaning (by way of a definition) of novel words before participants learned to read a novel orthography similar to that used here. Knowledge of both phonology and semantics supported learning of the writing system. Crucially, participants pre-exposed to phonology or semantics performed less accurately when generalising their orthographic knowledge to read untrained words. This outcome again suggests a complex trajectory of learning in which orthographic knowledge interacts with lexical knowledge during learning; there seems to be a trade-off between holistic and componential learning.

Together this literature highlights the important role played by meaning when learning both spoken and written forms. We therefore adapted the procedure of Experiment 5 to additionally train novel spoken words in context of a novel object. After completion of the phoneme monitoring task participants learned to name unfamiliar objects using the spoken words they had encountered in the phoneme monitoring task. We hypothesise that the addition of a semantic element may support learning and consolidation of spoken words, and thus support learning to decode the novel orthography.

5.7 Methods

5.7.1 Participants

24 native English-speaking adults aged 18 – 40 took part in a study approved by the Cambridge Psychology Research Ethics Committee. No participants reported having dyslexia, speech, or language impairment. Due to experimenter error, two participants did not complete the phoneme monitoring task. In addition, one participant did not complete the matching task. In both cases, data for the remaining participants are reported.

5.7.2 Materials

The same materials as above were held constant: four sets of 16 monosyllabic CVC pseudowords. Each of the four sets was assigned to trained or untrained conditions on either day 1 or 2 in a counterbalanced manner across participants. Consonant frequency was matched across stimulus groups with each consonant appearing twice in onset and twice in coda and unfamiliar visual symbols were mapped to each of the phonemes in a one-to-one manner.

5.7.3 Procedure

The training and testing procedure used here was the same as Experiment 5 but with the addition of an object training session prior to the phoneme monitoring session each day. A full timeline of the experiment can be found in Figure 5.7 (p164).

Phoneme monitoring. The phoneme monitoring adopted the same procedure as the previous experiment. 16 spoken pseudowords were trained on each day of the experiment.

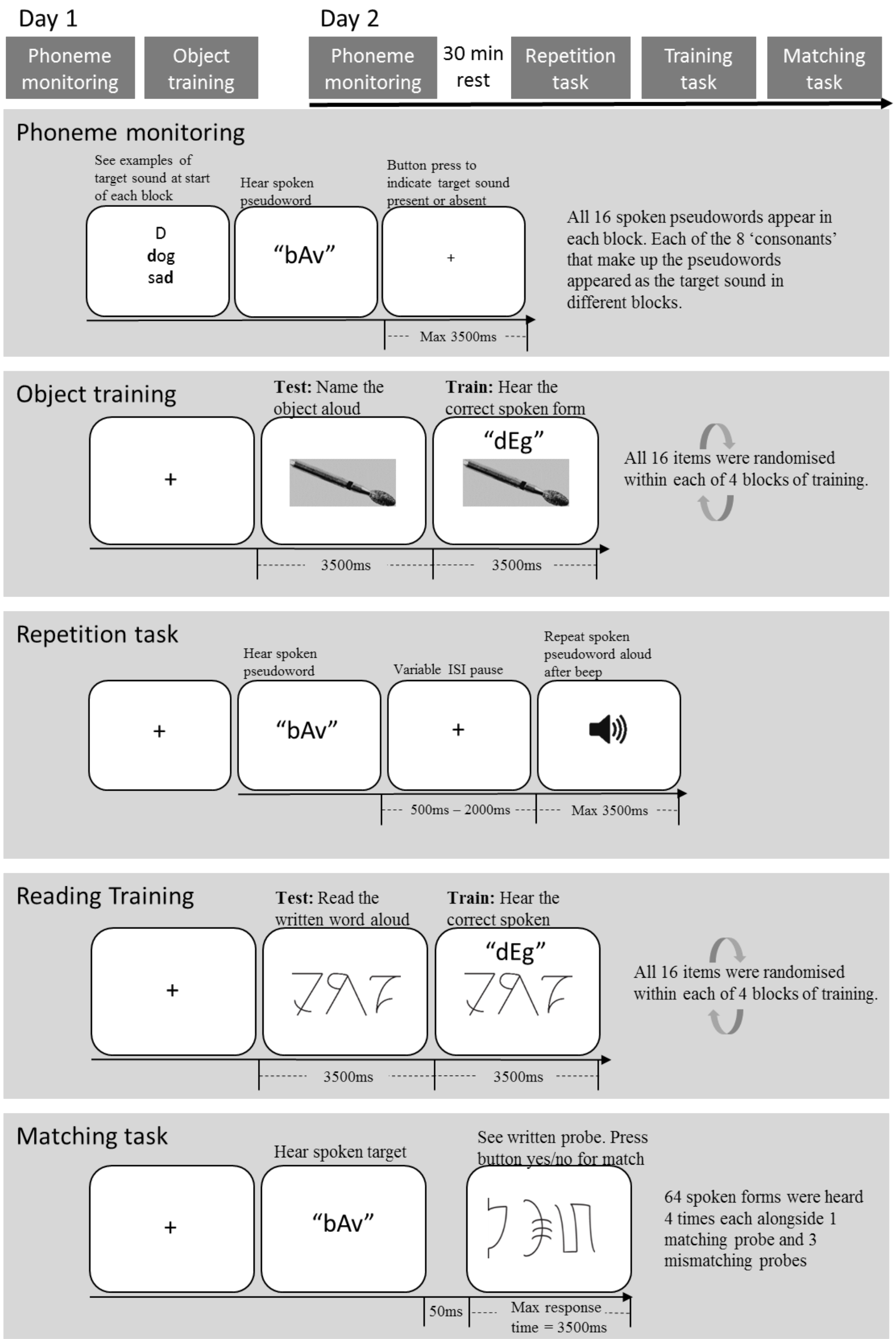


Fig. 5.7 Experiment 6: Timeline and procedures

Participants heard the spoken forms of the words and pressed a button when they heard a target phoneme. Each of the 8 'consonants' appearing as the target phoneme in different blocks, one phoneme as target per block. Before each block, a written cue indicated which phoneme participants should listen for. The written cue also showed two example words containing the phoneme in onset and coda positions. The 16 spoken words were randomised within each block. There were 4 runs (of 8 blocks) with each phoneme appeared as a target once per run, thus 512 trials per participant per day (16 words presented for each of 8 target phonemes over 4 runs). Each spoken word, therefore, appeared 32 times during training. Due to the statistically matched nature of the artificial items, each phoneme was present twice in onset and twice in coda on each run, leading to four 'hits' in each run of 16 items. Participants had up to 3500ms to respond on each trial and received no feedback on accuracy.

Object naming training. The 16 spoken pseudowords were assigned in a randomised manner to one of the 16 novel objects. The participants learned to name the novel objects using the spoken pseudowords. The object naming task followed a similar procedure to the orthographic training task. Participants were familiarised with the task and items in a block where they saw and heard all 16 items once in a randomised manner. Testing and training were combined in the same trials (Figure 5.7, p164). In each trial, the visual object was presented onscreen for 3500ms and participants were instructed to name it aloud. Participants then heard the correct spoken form of that object and had a further 3500ms in which to learn while the object remained onscreen. There was a gap of 1000ms between trials in which a fixation cross appeared onscreen. All 16 items appeared in a randomised manner within each block and 4 blocks were included in total.

Spoken word repetition. The spoken word repetition task followed a 30-minute rest break after the completion of training. On each trial, participants heard one of the 64 spoken items (trained/untrained, day 1/day 2) followed by a beep that occurred with a variable ISI between 500 and 2000ms. Participants were instructed to wait for the beep and then to repeat the word as quickly as possible. Responses were recorded and speech onsets were marked by hand offline using CheckVocal (Protopapas, 2007).

Orthographic training task. As in Experiment 5, the critical measure in this experiment was whether learning and consolidating a spoken word would help learning to read a novel orthography. Participants were trained to read the novel orthography using the same training paradigm used previously. On each trial, participants saw the written word onscreen and tried to read it aloud. After 3500ms they heard the correct spoken form of the word and then had a further 3500ms in order to learn. Unlike Experiments 2 and 3 where participants learned 16 words each day (made up of 12 graphemes) at a time participants here learned to read all 64 items (made up of 20 graphemes: 16 consonants and 4 vowels). All 64 items (heard and unheard spoken words from days 1 and 2) were presented in a fully randomised manner in each of 4 training runs.

Matching task. As in the previous experiment, the matching task allowed a measure of lexical and orthographic learning that was not dependent on a verbal response. Participants heard the spoken form of one of the 64 pseudowords (heard/unheard, day1/day2), with a visual probe appearing 50ms after the end of the spoken item. Participants pressed a button to indicate whether the spoken target matched the written probe. Response times were measured from the onset of the written probe and to ensure speeded responses an upper limit of 3500ms was imposed. All 64 pseudowords were used as spoken targets. There were three probe types: matching, mismatching in onset, and mismatching in coda. As before roughly half (44.53%) of the mismatching probes had been seen by the participants in order to avoid participants rejecting mismatching items purely on the basis of visual familiarity.

5.8 Results

5.8.1 Phoneme monitoring

Again, accuracy on the phoneme monitoring task was very high across both days $[Day\ 1: M = 95.19\%,\ SD = 21.40\%; Day\ 2: M = 95.50\%,\ SD = 20.74\%]$. D' scores were again very high $[Day\ 1: M = (3.32); Day\ 2: M = (3.42)]$, indicating participants were attending to the task. A paired t-test on the d' scores showed participants were equally attentive on days 1 and 2, $[t_p(21) = 0.58,\ p = 0.57]$.

5.8.2 Object learning

Accuracy on the final run of training was high on both days, indicating participants had learned to associate the spoken words with the visual objects: $[Day\ 1:\ M = 80.73\%,\ SD = 39.49;\ Day\ 2:\ M = 87.76\%,\ SD = 32.82]$. Accuracy was higher for the items on day 2, $[t_p(23) = -2.28,\ p = 0.03]$.

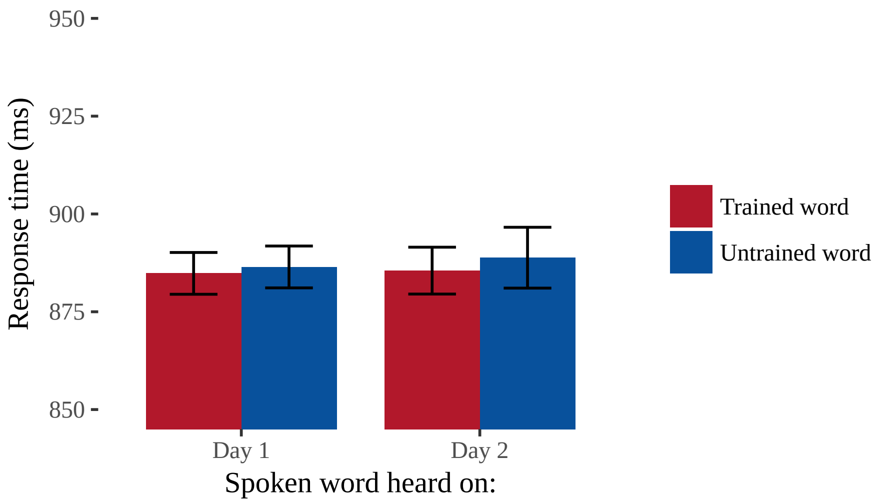


Fig. 5.8 Response time to repeat aloud trained and untrained words from days 1 and 2.

Table 5.3 Experiment 6: Performance on spoken word repetition task

	Response time (ms)	
	Mean	St. Dev
Day 1 trained	885	196
Day 1 untrained	886	200
Day 2 trained	886	207
Day 2 untrained	889	203

5.8.3 Spoken word repetition

In order to test whether overnight consolidation has taken place participants completed a spoken word repetition task. Participants repeated aloud both trained and untrained words from days 1 and 2 (Figure 5.8, p167; Table 5.3, p167). As before participants were highly accurate in repeating the spoken words: $[Day\ 1: M = 97.03\%,\ SD = 16.99\%;$ $Day\ 2: M = 98.08\%,\ SD = 13.74\%]$. Log-transformed response times were compared using a 2 x 2 repeated measures ANOVA with the factors of day-of-learning and training (trained or not). Even after the addition of an extra object-learning training session there was no significant effect of day, $[F(1,23) = 0.01,\ p = 0.92]$; or training, $[F(1,23) = 0.68,\ p = 0.10]$; and no interaction between the two factors: $[F(1,23) = 0.07,\ p = 0.79]$.

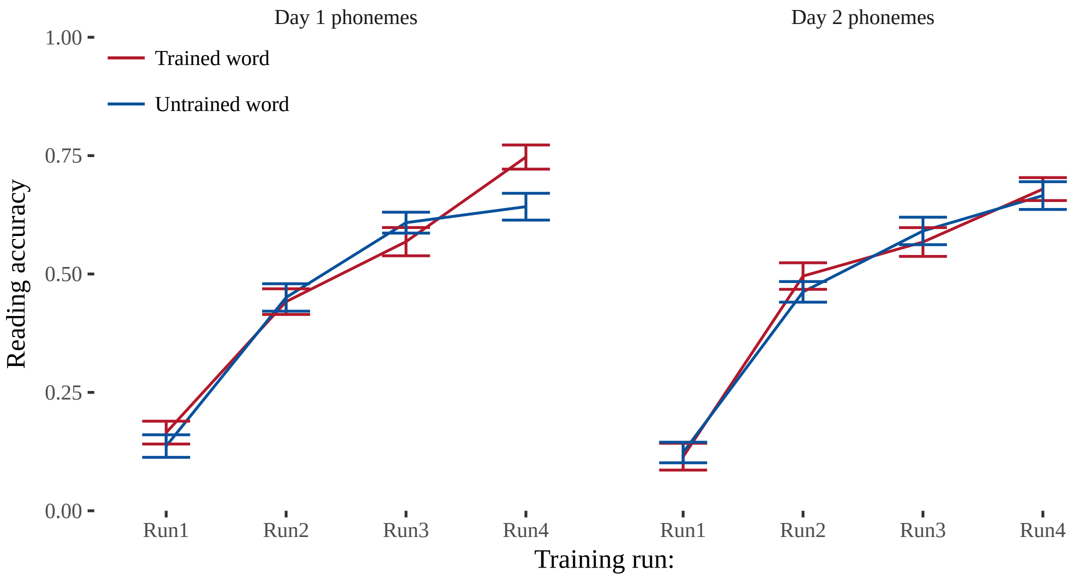


Fig. 5.9 Performance when learning to read the two orthographies on day 2. Note that day 1/day 2 refers to when the participants heard the spoken forms of the words. All learning took place during the same training session, with day 1 and 2 phonemes separated here for clarity.

5.8.4 Orthographic training

As in Experiment 5, the critical measure in this experiment was whether novel learned (and consolidated) spoken words help to learn a novel orthography. Figure 5.9 (p168) shows the learning trajectory across the four runs of training (and Table 5.4, p169). Four participants did not successfully learn to read the novel orthography ($< 20\%$ accuracy in the final run of training) and so are excluded from further analysis. Overall accuracy was lower than in previous studies. In the final run of training participants read words in the day 1 orthography correctly in 69.61% of trials ($SD = 46.02\%$), and words in the day 2 orthography in 67.24% of trials ($SD = 46.97\%$). There was a large effect of run, with accuracy improving over each run of training: $[F(3,57) = 131.38,\ p < 0.001,\ \eta_p^2 = .87]$. However, there was no effect of training on the accuracy with which participants learn the novel orthography $[F(1,19) = 1.11,\ p = 0.31]$. Additionally, there were no effects of day $[F(1,19) = 2.15,\ p = 0.16]$ and no interactions between day, training, and run.

5.8.5 Matching task

Accuracy. Accuracy was high in all conditions, $[day\ 1\ trained : M = 88.41\%,\ SD = 32.03\%;\ day\ 1\ untrained : M = 89.51\%,\ SD = 30.65\%;\ day\ 2\ trained : M = 88.15\%,\ SD =$

32.33%; $day\ 2\ untrained: M = 88.02\%,\ SD = 32.49\%$] (Figure 5.10, left panel, p170). A 2 × 2 repeated measures ANOVA compared the factors day of orthography (day 1 or day 2), and training (trained or untrained). There were no significant differences in accuracy across the four conditions: no main effect of day, $[F(1,19) = 0.14,\ p = 0.71]$, no main effect of training, $[F(1,19) = 0.98,\ p = 0.51]$, and no interaction, $[F(1,19) = 1.71,\ p = 0.50]$.

Response times to matching probes. Incorrect responses were discarded and response times for the remaining data were analysed. Matching and mismatching trials were collapsed for the primary analysis of response times. A repeated measures ANOVA showed no main effect of day, $[F(1,19) = 0.13,\ p = 0.72]$; no main effect of training, $[F(1,19) = 2.44,\ p = 0.10]$; and no interaction, $[F(1,19) = 1.06,\ p = 0.30]$ (Figure 5.10, right panel, p170). As matching and mismatching conditions likely tap into different cognitive processes, with matching trials giving a better assessment of word learning, we analysed the data including only the matching trials. As before the ANOVA showed neither main effect of day, $[F(1,19) = 0.12,\ p = 0.73]$, main effect of training, $[F(1,19) = 1.72,\ p = 0.20]$, nor interaction, $[F(1,19) = 0.13,\ p = 0.73]$.

Response times to mismatching probes. Response times to detect errors were entered into a 2 x 2 x 2 ANOVA with the factors day (day 1 or day 2), training (trained or untrained), and error position (onset or coda error), (Figure 5.11, p170). Again participants were much faster to detect errors in onset position, $[F(1,19) = 72.15,\ \eta_p^2 = .79,\ p < 0.001]$. The effect of day was not significant, $[F(1,19) = 0.08,\ p = 0.78]$, nor was training status, $[F(1,19) = 1.53,\ p = 0.23]$. There were no significant two- or three-way interactions.

5.8.6 Cross-experiment comparison.

One potential reason for the failure to observe spoken word consolidation effects was the relatively small sample sizes used here. In modifying the procedure for Experiment 6 we

Table 5.4 Accuracy during training

	Mean (SD)			
	Run 1	Run 2	Run 3	Run 4
Day 1 trained	16.50% (37.19%)	44.16% (49.74%)	56.83% (49.61%)	74.70% (43.54%)
Day 1 untrained	13.66% (34.40%)	45.05% (49.83%)	60.84% (48.88%)	64.22% (48.88%)
Day 2 trained	11.42% (31.85%)	49.57% (50.07%)	56.76% (49.63%)	67.94% (46.75%)
Day 2 untrained	12.31% (32.90%)	46.23% (49.94%)	59.09% (49.24%)	66.56% (47.25%)

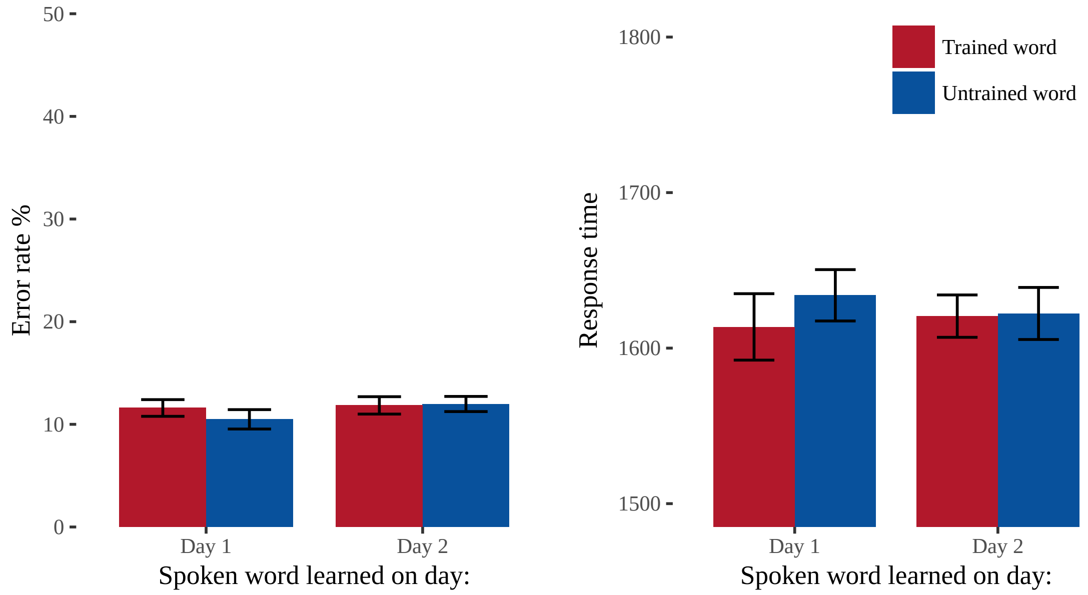


Fig. 5.10 The left panel shows errors rates on the matching task. The right panel shows response times for trained and untrained spoken words from days 1 and 2. Note: response time figure includes correct responses only.

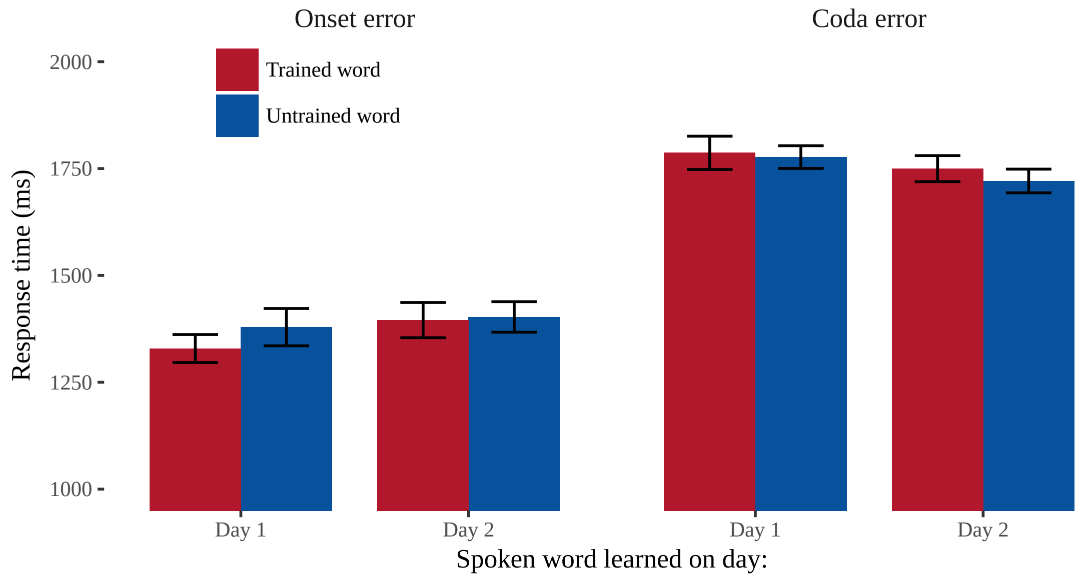


Fig. 5.11 Detecting mismatches according to error position.

were careful to use the same stimuli to allow the data from both experiments to be analysed together, thus increasing the power to detect potentially small effect sizes. To assess the extent to which these data stand as evidence for the null hypothesis we additionally report Bayes Factors for each of the combined comparisons using the `BayesFactor` package for `R` (Morey et al., 2015; Rouder et al., 2017). Bayes factors larger than 3 or less than 1/3 are viewed as strong evidence, while Bayes factors between 1/3 and 3 is weak evidence (Dienes, 2014).

Spoken word repetition. Combining delayed repetition data from the two samples confirmed there was no effect of either day, $[F(1,40) = 0.33,\ p = 0.57,\ BF_{10} = 0.05]$, or training, $[F(1,40 = 1.23,\ p = 0.27,\ BF_{10} = 0.08]$, on response time, nor any interaction between the two, $[F(1,40) = 0.00,\ p = 0.96,\ BF_{10} = 0.00]$.

Orthographic training. For the combined training data there was a large effect of run, with accuracy improving over each run of training: $[F(3,108) = 259.08,\ p < 0.001,\ \eta_p^2 = .88,\ BF_{10} = 2.64 \times 10^{51}]$ but no main effect of day $[F(1,36) = 0.87,\ p = 0.36,\ BF_{10} = 0.03]$. Even with this larger sample the improved performance for trained spoken words did not show strong evidence, as measured by the bayes factor: $[F(1,36) = 7.11,\ p = 0.01,\ \eta_p^2 = .17,\ BF_{10} = 0.45]$. There was no interaction between day and training $[F(1,36) = 1.25,\ p = 0.27,\ BF_{10} = 0.00]$, between day and run $[F(3,108) = 0.90,\ p = 0.45,\ BF_{10} = 0.00]$, nor between training and run $[F(3,108) = 0.28\ p = 0.84,\ BF_{10} = 0.00]$. The three way interaction of day, training, and run was also not significant $[F(3,108) = 0.32,\ p = 0.81,\ BF_{10} = 0.00]$.

Matching task. On the matching task there were no significant differences in accuracy across the four conditions: no main effect of day, $[F(1,36) = 0.83,\ p = 0.37,\ BF_{10} = 0.11]$, no main effect of training, $[F(1,36) = 2.11,\ p = 0.15,\ BF_{10} = 0.05]$, and no interaction, $[F(1,36) = 0.89,\ p = 0.35,\ BF_{10} = 0.00]$. Moreover response times to endorse matching probes showed neither main effect of day, $[F(1,36) = 0.19,\ p = 0.67,\ BF_{10} = 0.04]$, main effect of training, $[F(1,36) = 1.83,\ p = 0.18,\ BF_{10} = 0.07]$, nor interaction, $[F(1,36) = 0.86,\ p = 0.36,\ BF_{10} = 0.00]$. Response times to the mismatching probes again showed participants were much faster to detect errors in onset position, $[F(1,36) = 121.19,\ p < 0.001,\ \eta_p^2 = .77,\ BF_{10} = 8.18 \times 10^{16}]$. The effect of day was not significant, $[F(1,36) = 0.89,\ p = 0.35,\ BF_{10} = 0.05]$, nor was training status, $[F(1,36) = 2.95,\ p = 0.09,\ BF_{10} = 0.07]$. There were no significant two- or three-way interactions.

5.9 Discussion

5.9.1 Failure to observe spoken word consolidation

Here, in a second experiment on spoken word consolidation, we fail to observe overnight spoken word consolidation using a spoken word repetition task. Participants did not show any differences in response times for trained or untrained words from days 1 or 2, even after the inclusion of a semantic element to the training. Indeed, this was the case even when pooling the participants from Experiments 5 and 6. Moreover, Bayes factors for the combined data indicate that the result provides strong support for the null hypothesis ($BF_{10} = 0.05$ for the effect of training and $BF_{10} = 0.08$ for the effect of day). This outcome is problematic as it suggests the spoken word repetition task itself is not sensitive to word learning in this context. Indeed, the lack of differences between trained and untrained spoken words precludes any potential consolidation effects.

Overnight consolidation of spoken words has been shown by several studies using the tasks used here (e.g., Davis et al., 2009; Hawkins et al., 2015; Takashima et al., 2014a) and so the absence of findings here does not necessarily cast doubt on previous results. Instead, the lack of replication in the current data may help to determine the parameters that influence lexical consolidation. The most likely reason for the failure to replicate is the experimental stimuli. We use training and testing paradigms that have shown consolidation effects elsewhere, but an item set different to that used elsewhere. For example, the combination of phoneme monitoring during training and spoken word repetition during testing showed robust consolidation effects in Davis et al. (2009). Moreover, the number of lexical items used here (16) seems unlikely to have prevented consolidation. Takashima et al. (2014) for instance found consolidation effects after training 40 novel words on day 1. The behavioural experiment in Davis et al. (2009) taught 18 items each day while the subsequent fMRI experiment used 30 items on each day.

As highlighted previously one possibility has to do with the distinctiveness of the items. Most studies of lexical consolidation have deliberately trained distinctive words. For instance, Gaskell and Dumay (2003)note the stimuli in their study were chosen deliberately to be atypical, to be low frequency, and to be long enough to contain uniqueness points two phonemes before the end of the word. Short words occupy dense phonological neighbourhoods, a fact that proved challenging when selecting non-word stimuli for the experiments of this thesis. The limited phonological space available to CVC words means that the both existing words and the novel words used here form a very confusable set of items. The lack of consolidation may thus be a consequence of the lexical items being very similar to each other.

Consequently, it may be the case that these novel words do not impact on the memory system in a manner that is measurable.

We attempted to increase both the meaningfulness and distinctiveness of items in Experiment 6 by adding an object naming element to the experiment. Experiment 1 demonstrated that learning to name objects engages holistic processing, while computational accounts have suggested holistic learning may be complementary to componential processes (Plaut and Vande Velde, 2017). Thus we reasoned that the 16 individual words may be more individually memorable. Additionally, a variety of sources point to the importance of semantics when learning a new word (e.g., Hawkins, Astle, & Rastle, 2015 in the case of spoken words and Taylor et al., 2011 for written words). The visual objects associated with the spoken words here may provide at least some element of 'meaning' to the novel words. Nevertheless, we fail to see any effect of overnight consolidation even with this additional learning step. Rather than strengthening hippocampal representations and thus supporting the reinstatement of memories during sleep, it may be the case that that addition of semantic input may have impeded lexicalisation. Hawkins and Rastle (2016) suggest that semantic learning may weaken the phonological representations by diverting attention away from the phonological form of the words. We reasoned that including both phoneme monitoring and an object learning task would avoid this problem, however, participants may still have been biased to focus on the meaning of the words rather than their forms. Note that while d' scores in the phoneme monitoring task highlight participants were attentive to the task we do not have a measure of how well participants learned and remembered the novel spoken words.

Alternatively, it may be the case that overnight consolidation effects are reliant on existing schema to support the rapid (overnight) lexicalisation of novel spoken and written words. Havas et al. (2017) for example found learning a novel spoken word was supported by association with an object, but only if that object was familiar. The objects used here were unfamiliar by design. In addition, word learning was also supported if the phonology was of similar to L1 phonology; novel words are supported by existing words. The spoken items used here, by contrast, occupy an extremely dense phonological space and so it may be unlikely that the novel words map clearly to an existing word. Moreover, pre-existing knowledge plays an important role in shaping learning in the triangle model (Chang et al., 2017; Harm and Seidenberg, 2004; Plaut et al., 1996); the mappings between $(P \mapsto S)$ influences the character of $(O \mapsto P)$ learning.

It may be possible that allowing a longer period between training and testing would allow the novel words to be more completely lexicalised. Hawkins and Rastle (2016) found consolidation effects that only became apparent one week after initial training, a pattern seen elsewhere (e.g., Bakker et al., 2014). Given the extended timeframe of these results, it is possible that extending the interval between the initial learning of the spoken words and the test date would encourage lexicalisation. Future studies could extend the approach used here by training spoken words over multiple days and using multiple training methods. Once clear lexical consolidation has been established the approach used here may be valuable in identifying the contributions of spoken language to literacy acquisition. Although we fail to see consolidation effects here, the approach of using newly learned and lexicalised spoken words to investigate literacy acquisition has merit and may be used in future to investigate the impact of spoken language on reading.

Unreliable effects of spoken language on learning orthography. The lack of spoken word learning on the repetition task means that we cannot make strong claims about the role of recently learned words when learning to read. Unsurprisingly given the lack of differences for the spoken words, no differences emerge in either the training or matching tasks in Experiment 6 (Figure 5.9, p168 and Figure 5.10, p170), either in terms of accuracy, time to endorse correct matches for trained and untrained spoken words, or to reject mismatches.

In Experiment 5 participants were more accurate when reading words that they had listened to during the phoneme monitoring task. In Experiment 6, by contrast, they were not more accurate to name words learned both in the phoneme monitoring and object naming tasks. This is a counter-intuitive outcome given the object-learning task was included with the goal of promoting lexical-level learning. It is possible that the inclusion of the object-naming task, rather than aiding learning of the spoken forms, actually impeded learning of the mappings between written and spoken forms. Combining the orthographic training data from the two experiments indicates that if there is an effect of spoken language on learning to read, the task here is not sensitive to it. Although the combined data showed a statistically significant result ($p = 0.01$) the associated Bayes factor ($BF_{10} = 0.45$) suggested this evidence is inconclusive (Dienes, 2014).

5.10 Conclusion

This chapter attempted to make use of established research on spoken word consolidation to explore the role of spoken language on the earliest stages of learning to read. In two experiments we fail to replicate evidence of spoken word consolidation, though there are a number of reasons why this may be the case. The lack of replication of consolidation effects is valuable in discovering the boundaries of the effect. To date published work on consolidation of spoken words has been limited to several narrow experimental settings. In the current context, we saw no evidence for consolidation even after inclusion of an additional object naming task. An important positive finding, even in the absence of consolidation effects, is that even at this early stage of acquiring a novel orthography there is a benefit of spoken language; trained words in Experiment 5 had an advantage over untrained words. This outcome motivates the importance of trying to separate the role of spoken language from purely orthographic knowledge. However the lack of replication of this outcome in Experiment 6 limits the strength of conclusions that can be drawn.

Chapter 6

Thesis summary and conclusions

This thesis set out to offer a fine-grained perspective on the earliest stages of reading acquisition by combining an artificial orthography paradigm with a memory-consolidation framework. This approach was motivated by the difficulty of tracking the relative contributions of lexical and sublexical orthographic learning in naturalistic contexts. We suggested that the experimental control allowed by training an artificial orthography would overcome some of these issues, while the impact of memory consolidation may highlight the different trajectories of lexical and sublexical learning. In doing so we hoped to provide a complementary perspective on continuing debates in reading instruction and to validate a framework that may address additional questions in a rigorous and replicable manner. Here, we revisit the specific questions identified by the review of the literature before summarising the extent to which these questions were addressed by the thesis:

1. To what extent does learning a novel orthography draw upon the same neural systems as reading real words and pseudowords, and to what extent might this change with consolidation?

2. Will whole-word orthographic representations emerge in the context of learning to read an artificial orthography?

3. How will phonological, lexical, and semantic information impact on the ability to use and generalise an artificial orthography?

4. Will sleep consolidation impact on recently learned orthographic knowledge, and if so, will consolidation occur at a lexical or sub-lexical level?

6.1 Thesis summary

Experiment 1. Experiment 1 investigated the neural basis of learning to read by training participants to read novel words in a novel orthography and to name novel objects. Participants learned to read/name a set of words and objects on day 1 and a second set on day 2. Following training on day 2 participants completed a scanning session where they read/named all trained items as well as a set of untrained words, and the written forms of the trained objects. We reasoned that decoding novel words letter-by-letter would rely on the same regions involved in the componential reading of pseudowords, while naming novel objects would tap into the same regions involved in holistic recognition of familiar English words. A functional localiser run additionally identified brain regions involved in reading/naming familiar words and objects, as well as a set of matched pseudowords.

Behavioural results highlighted the difference between holistic and componential processing. Participants learned the letter-sound mappings from the trained words on day 1 and thus performed more accurately from the first run of training on day 2. This pattern extended to scanning where participants could generalise their componential letter-sound knowledge in order to read untrained words. During scanning, participants were equally accurate reading trained words from days 1 and 2, as well as the untrained words. Accordingly, we conclude that participants have learned representations of the novel letters but not the novel words.

Analysis of the functional localiser incorporated a recent approach to modelling response times by Taylor et al., (2014b). This allowed us to make comparisons between activation for reading and object naming while overcoming the potential confounds caused by differences in response times to the stimuli. This functional localiser allowed a unique comparison between reading real and artificial orthographies within the same participants, thus helping to validate the artificial orthography approach. This analysis demonstrated the neural systems involved in established reading are the same as those involved in early stages of learning to read a novel orthography.

Whole-brain imaging analysis for artificial words and objects highlighted the neural regions involved with learning holistic and componential visual-verbal mappings. In particular anterior vOT regions were activated when learning holistic mappings while posterior vOT and parietal regions contributed to componential mappings. Critically, the strong similarities between neural responses to real and artificial stimuli provides a clear validation of the experimental approach to studying literacy acquisition: the main contrast between holistic and componential processing of artificial objects and words overlapped not only with a within-subject localiser for words and pseudowords, but also with a meta-analysis of reading (Taylor et al., 2013), and with the outcomes of a previous study using artificial reading methods (Taylor et al., 2014a).

Finally, a hippocampal region of interest analysis identified reduced activation for trained words and objects from day 1. As the same novel letters were used for the words on day 1 and day 2 this reduction may indicate the participants that whole word representations have emerged following consolidation.

Experiment 2. Experiment 2 directly investigated the impact of overnight consolidation at the letter-level by training distinct orthographies on days 1 and 2. By testing participants' performance reading both trained and untrained words in the two orthographies we were able to separate any effects of lexical and sublexical consolidation. Improved performance when reading aloud trained words from days 1 and 2 indicated lexical representations had been learned. The reading task showed no effects of overnight consolidation, however, in either response times, accuracy, or in a repetition priming measure. In contrast, a spoken-to-written matching task showed clear effects of overnight consolidation at the level of letters; participants were faster to identify both trained and untrained words written in the day 1 orthography. Response times to detect errors in mismatching probes confirmed this finding. Finally, the inclusion of consistent and mixed orthography conditions tested whether responses varied according to whether the letters appeared in a familiar context. Responses were significantly slower in the mixed orthography condition, suggesting participants were already sensitive to the bigram statistics of the items.

Experiment 3. Experiment 3 extended this approach to directly investigate how bigram learning may interact with sleep consolidation. Participants again learned two distinct orthographies on days 1 and 2. Participants' ability to generalise consolidated and unconsolidated bigrams was tested on day 2 by including untrained words that varied according to the training status of their component bigrams. No effect of bigram familiarity was found for words from days 1 and 2. Results from the matching task, however, replicated the finding of sublexical consolidation; participants were again faster to identify trained and untrained words written in the day 1 orthography. A contrasting finding came from the recognition memory task; faster responses to identify trained words from day 1 indicated participants had consolidated representations of the *spoken* word forms.

Experiment 4. Experiment 4 manipulated the frequency of bigrams during training to probe whether overnight consolidation may support the learning of statistical structure in the novel orthography. Certain bigrams appeared more frequently in than others during training on days 1 and 2. Results from the matching task provided another replication of the finding that overnight consolidation supports sublexical orthographic learning and generalisation. In this case the faster responses were also found in the reading task; participants were faster to read aloud trained and untrained words written in the day 1 orthography. Finally, there was clear evidence for overnight consolidation of *visual* word forms on the recognition memory task in Experiment 4.

Experiment 5. Experiment 5 attempted to link the research on consolidation of orthographic knowledge with previous findings on the consolidation of spoken word forms. Participants were trained on novel spoken words on days 1 and 2 using a phoneme monitoring task. A spoken word repetition task on day 2 failed to replicate previous findings of overnight lexical consolidation. Consequently, while participants successfully learned to read novel orthographies corresponding to the words from days 1 and 2, there were no differences in accuracy or response times. We discuss potential reasons for the discrepancy between the negative results found here and those found in previous studies.

Experiment 6. Experiment 6 expanded on the approach used in Experiment 5 by adding object naming task to the training phase of both days. We reasoned that learning the novel spoken words in association with novel objects would make the words more distinctive and thus more likely to encourage memory formation. Here again, a spoken word repetition task showed no evidence that participants had consolidated representations of the words overnight. Combining data from Experiments 5 and 6 showed no experimental effects even with the larger sample size. Null findings were confirmed by the inclusion of Bayes factors. We suggest that the lack of even training effects for the phoneme monitoring task indicates it is not a suitable measure in this context.

6.2 Limitations of this thesis

There are a number of limitations to the conclusions that can be drawn from this body of work as a consequence of the methods used. First, by its very nature laboratory-based learning approaches sacrifice ecological validity in favour of experimental control. We justify this trade-off by appealing to the need for rigorous, well-controlled evidence to complement the more naturalistic evidence predominant in the literature. The most important caveat of this body of research then is that this body of findings on reading artificial orthographies is not intended to stand alone. Instead, the approach offers a useful addition to the wide range of research methods employed by educators, psychologists, and neuroscientists to investigate reading.

A second major limitation is that despite the goal of tracking lexical and sublexical learning in the earliest stages of reading acquisition, all participants had already well developed literacy skills. All of the learning that takes place here may therefore be dependent on the literacy skills developed across a lifetime. As noted previously the process of initially acquiring written language fundamentally alters how spoken language is perceived (Brady & Shankweiler, 1991; Byrne, 1998; Goswami & Bryant, 1990). Thus, it is likely that the orthographic learning in these experiments piggybacks on the participant's existing literacy skills. A related concern is that for practical purposes we limited our participants to adults. This experimental decision was made for pragmatic reasons, allowing rapid deployment and comparison of different training and testing scenarios in a short time period. Nevertheless, we are limited in the extent to which we can generalise our findings to children. At the same time, we as researchers should not limit our attention to children learning to read. There remain very large proportions of the adult population that have not achieved functional literacy (Morais, 2017).

Additional caveats relate to the experimental design. One of the most important considerations is that the findings presented throughout this PhD use a 'train twice, test once' design. This approach is efficient, allowing us to compare consolidated and unconsolidated items within the course of a single scanning session. The behavioural studies maintained this approach both for consistency with the initial fMRI study, but also because the testing regime was amenable to the time-course and resources available for a PhD project. This design decision does limit the conclusions that we can make about the results. While the overnight changes found here are framed in light of the sleep-related memory changes found elsewhere, we do not rule out the possibility of offline consolidation not related to sleep also plays a role. An alternative explanation could be 'age of acquisition' effects related to the order of training (c.f., Zevin and Seidenberg, 2002). Future research to address this alternative explanation might adapt the paradigm to a between-group nap study whereby half of the participants have a period of sleep between training and testing sessions. The benefit of this approach is that it holds constant the amount of time between training and testing. Replicating improvements in recognising an artificial orthography after sleep in this setting would provide stronger evidence that sleep, rather than time, is contributing the learning. Alternatively, adopting the AM/PM training/testing design used elsewhere would add support to the view that changes in orthographic processing come about due to sleep-related changes.

The statistical power of these experiments remains a concern. This thesis took an exploratory approach by marrying the experimental approaches of two very distinct literatures (reading and consolidation). The six experiments of the thesis attempt to broadly explore the parameter space of this novel domain. The number of the experiments, along with the limited time and resources available in a PhD thesis, do come at a cost to participant numbers. The relatively low sample size of the experiments, and the consequent impact on statistical power, may diminish the robustness of these findings.

When using mixed-effects methods, Brysbaert and Stevens (2018) recommend 1,600 observations per condition with the observations divided between participants and items. For example, 40 participants each seeing 40 trials. The experiments presented here straddle both sides of this recommendation. For the critical matching condition in Experiment 2, 32 items written in the day 1 orthography each appeared four times. With 19 participants, the experiment contained 2,432 observations. In contrast, only 17 participants completed the recognition memory task in Experiment 3 where there were only 64 trials per condition, thus 1,088 observations overall.

To reduce the impact of low sample size on the value of the results the sequence of experiments was designed deliberately to include replication at each step. Each of the experimental chapters includes a strong element of repetition, either closely mirroring previous research in the case of Experiments 1, 5, and 6; or internal replication in the case of Experiments 2, 3, and 4. This approach has proved valuable, demonstrating multiple repetitions of core experimental findings of the thesis. Nevertheless, the design and implementation of these experiments is clearly exploratory in nature. The strength of experimental results motivate confirmatory follow-up studies using larger sample sizes.

6.3 Future directions

The successful combination of overnight consolidation and artificial orthography learning motivates further directions of research. One outstanding question across all experiments is whether the inclusion of inconsistent letter-sound mappings might prompt lexical consolidation overnight. Taylor et al. (2011) in a same day design showed that learners were able to extract and generalise inconsistent letter-sound mappings in just one day of training. As discussed in Chapter 4, the fact that all letter-sound mappings in these experiments may have encouraged a letter-level focus while discouraging whole word recognition. Including inconsistent mappings may rectify this situation and prompt whole word consolidation.

In the same vein, inconsistent spellings are often conditional on the consonantal context of the word (Treiman et al., 2006, 1995). Learning the first- and second-order statistical dependencies that determine the pronunciation of inconsistent letters may be affected by overnight consolidation. Overnight consolidation has been demonstrated to support learning and abstraction of statistical information in language, (e.g. Gaskell et al., 2014). Experiment 4 of this PhD went some way to exploring statistical consolidation by training some bigrams more frequently. A more direct test of this hypothesis would be to adapt the inconsistent stimuli of Taylor et al. (2011) to an overnight consolidation design. Testing generalisation of orthographic knowledge, as well as, lexical recognition memory may highlight changing orthographic representations. In addition, Experiment 2 of this thesis gave partial evidence for orthotactic learning. A very direct adaptation of the Gaskell et al. (2014) experiment to an orthographic setting would address the question of whether overnight consolidation supports learning of orthotactic constraints.

Multivariate approaches to fMRI analysis would allow a more sensitive measure of learning and a natural extension of both the imaging results from Experiment 1 and the later behavioural results. These multivariate approaches ignore the overall amplitude of BOLD signal and focus instead on the information carried within the BOLD signal. Calculating multivariate distance metrics for different stimuli allows quantification of how different brain signals process different stimuli. This approach has been adopted in exploring representational differences in a familiar orthography between words and pseudowords, and between orthographically similar words in the VWFA (Baeck et al., 2015). This analysis demonstrated that VWFA discriminates not only lexical status but also sublexical orthographic features. Glezer et al. (2015) tracked changing patterns of activation as pseudowords became word-like over 19 days of training, this time using repetition suppression rather than multivariate analysis. Combining the sensitivity of MVPA approaches with an overnight consolidation design may allow us to track the changing representations of novel written and spoken lexical items. In particular this approach would be suitable for experimental designs such as that in Experiment 3. This experiment was able to hold letter and lexical frequency constant while manipulating the bigram familiarity within the words. The deliberate similarity between items here makes differences in BOLD amplitude unlikely and so a multivariate approach would be valuable.

6.4 Conclusions

Implications for models of reading These findings challenge existing reading models to more accurately reflect the developmental trajectories inherent in all learning. We have demonstrated that it is possible to track the emergence of interdependent cognitive skills over a period of time. Even in the short time windows used in these experiments the impact of consolidation is apparent. Our increasing understanding of the biological constraints on learning add perspective on processes that have been excluded or ignored in many accounts of reading acquisition. Most notably, one of the primary areas of agreement across models of reading is that there is more than one route to reading. The work presented here emphasises the distinction between holistic and componential approaches to reading, but also goes a step further. By situating these processes relative to each other within a developmental trajectory we allow future research to ask direct questions about educational practice.

Both of the main models of reading discussed in the literature review are challenged by these findings. The DRC model does not attempt to incorporate learning into the model, instead attempting to explain the end-point of reading acquisition. The findings of the six experiments suggest that the interdependent contributions of spoken work knowledge, knowledge of letter-sound mappings, visual word recognition all contribute to learning at different rates. Memory consolidation provides an explanatory mechanism to link these skills together across learning. By failing to address these questions the explanatory power of the DRC account of reading is reduced.

The triangle model of reading seems better placed to incorporate the findings of this research, given the relationship between connectionist models and dual-systems accounts of learning. The data here suggest a more complex picture than the DRC model currently attempts. Indeed, certain aspects of the triangle model already go some way to capturing the trajectory of learning, pre-training knowledge of spoken word forms before introducing orthographic learning. To fully engage with consolidation-related findings however, the triangle model should be informed by behavioural evidence like that shown here. Rather than training all elements of the model simultaneously, introducing different elements across learning may capture the staged learning shown by consolidation effects.

Implications for education Using the process of consolidation to provide insight into reading acquisition has proved fruitful. The value of this approach is not limited to reading however. The core insight of this approach is that all learning is inherently developmental but developmental trajectories are difficult to examine under experimental conditions. By making use of the cognitive changes that take place in the days and weeks subsequent to learning, however, we can find a balance between cost-effective research and experimental rigour that has been difficult to attain in educational research to date. The data presented here hold promise of an important change in how society improves education provision - improving instruction with experimental evidence.

There is increasing recognition that experimental rigour is the only method by which we can assess the efficacy of intervention in complex situations. While the medical field has whole-heartedly adopted experimental research, education has been slow to adopt this perspective. In recent years organisations such as the Educational Endowment Foundation have begun programmes of randomised controlled-trials to evaluate the efficacy of different approaches. These randomised controlled trials (RCTs) offer the gold-standard for evaluation of the relative efficacy of educational interventions (c.f. Torgerson et al., 2013, for a discussion of the methodological considerations of implementing RCTs in an educational setting). As with medical research research, there is a complementary relationship between the outcomes of population-level interventions and the fine-grained analysis of the cognitive factors that influence individual learning trajectories.

The difficulty incorporation these approaches into mainstream education policy may reflect the additional difficulties found in educational settings. Schanzenbach (2012) lists many of the pragmatic limitations of experimental research in education settings. In many cases it is not feasible to design and implement an intervention in a setting already lacking resources. Moreover, educational theories tend towards larger systems of thought and consequently complex treatments that cannot be easily assessed in an experimental context. The timing of experiments is also challenging. Results must come quickly enough to inform current policy perspectives but must also be informative from a developmental viewpoint. Cost and ethical concerns further hamper the ability of policy makers to use RCTs and these questions must be balanced against the need to ensure external validity across different geographical and populations. In sum, conducting high-quality experimental research in school settings is difficult. This has been particularly true for reading, where decades of research has faced these issues.

Research using artificial orthographies offers a rapid, low-cost, robust, and flexible complement to the larger field of literacy research. The scale of artificial reading studies can be a short as a matter of hours or be extended across months. Unlike interventions that might require >100 schools to modify their teaching policy the approaches used here can be implemented with different populations with samples sizes that reflect the advancement of the issue being tested. The increased feasibility and reduced costs mean that many more experimental questions can be tested in a more rigorous manner. A perennial difficulty with educational interventions is that complex treatment regimes lead to difficult to interpret outcomes. This can be seen from the lack of clarity about whether synthetic or analytic approaches to phonics instruction are better. These complex interventions can also lead to criticisms that the implementation of an intervention was not valid. Testing a larger number of simpler interventions is likely to offer a productive route to establishing more rigorous interventions.

Perhaps most importantly, artificial reading approaches address the core difficulty of education interventions, the black box problem. Evidence suggests basing reading instruction on phonics is more effective than other methods of instruction but does not tell us why this is the case, or indeed, if phonics is therefore the best *possible* approach. Because school-level research focusses on the outcomes of learning it does not give us insight into the learning mechanisms themselves. Consequently, debate about reading instruction has tended to become polarised into pro- or anti-phonics perspectives. The approach laid out here instead encourages a more fine-grained approach where we can track the shift from novice decoding to skilled reading. It is easy to imagine incorporating the abundant evidence for the impact of spoken language on reading using the experimental approached laid out here (Castles et al., 2006; Nation, 2017; Nation and Cocksey, 2009; Nation and Snowling, 2004; Ricketts et al., 2007; Stuart et al., 1999), and in doing so improve literacy provision.

To revisit the starting point of the thesis, *reading is the core skill* that enables participation in modern society. A child who reaches the end of school without learning to read is unlikely ever to catch up with their peers. Alongside the large economic benefits, we have an ethical duty to identify and adopt best practice in literacy instruction in society. To achieve this we must challenge our existing methods and assumptions with robust, replicable research. This thesis demonstrates that this research is both viable and valuable, and should encourage further efforts.

References

Acemoglu, D. (2002). Technical change, inequality, and the labor market. *Journal of economic literature*, 40(1):7–72.

Anderson, M. C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of memory and language*, 49(4):415–445.

Andrews, S. and Scarratt, D. R. (1996). What comes after phonological awareness? Using lexical experts to investigate orthographic processes in reading. *Australian Journal of Psychology*, 48(3):141–148.

Ashburner, J. and Friston, K. J. (2005). Unified segmentation. *Neuroimage*, 26(3):839–851.

Atienza, M., Cantero, J. L., and Stickgold, R. (2004). Posttraining Sleep Enhances Automaticity in Perceptual Discrimination. *Journal of Cognitive Neuroscience*, 16(1):53–64.

Baayen, R. and Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, 3(2).

Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, 59(4):390–412.

Baeck, A., Kravitz, D., Baker, C., and de Beeck, H. P. O. (2015). Influence of lexical status and orthographic similarity on the multi-voxel response of the visual word form area. *Neuroimage*, 111:321–328.

Baeck, A., Rentmeesters, N., Holtackers, S., and de Beeck, H. P. O. (2014). The effect of sleep in perceptual learning with complex objects. *Vision Research*, 99:180–185.

Bakker, I., Takashima, A., van Hell, J. G., Janzen, G., and McQueen, J. M. (2014). Competition from unseen or unheard novel words: Lexical consolidation across modalities. *Journal of Memory and Language*, 73:116–130.

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, 68(3):255–278.

Bates, D., Kliegl, R., Vasishth, S., and Baayen, H. (2015a). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., and others (2015b). lme4: Linear mixed-effects models using Eigen and S4, 2014. *R package version*, 1(4).

Beech, J. R. (2005). Ehri's model of phases of learning to read: A brief critique. *Journal of Research in Reading*, 28(1):50–58.

Ben-Shachar, M., Dougherty, R. F., Deutsch, G. K., and Wandell, B. A. (2007). Contrast responsivity in MT+ correlates with phonological awareness and reading measures in children. *Neuroimage*, 37(4):1396–1406.

Ben-Shachar, M., Dougherty, R. F., Deutsch, G. K., and Wandell, B. A. (2011). The development of cortical sensitivity to visual word forms. *Journal of Cognitive Neuroscience*, 23(9):2387–2399.

Bernstein, S. E. and Treiman, R. (2001). Learning a novel grapheme: Effects of positional and phonemic context on children's spelling. *Journal of Experimental Child Psychology*, 79(1):56–77.

Bertels, J., Franco, A., and Destrebecqz, A. (2012). How implicit is visual statistical learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(5):1425.

Bitan, T., Burman, D. D., Chou, T.-L., Lu, D., Cone, N. E., Cao, F., Bigio, J. D., and Booth, J. R. (2007a). The interaction between orthographic and phonological information in children: An fMRI study. *Human brain mapping*, 28(9):880–891.

Bitan, T., Burman, D. D., Chou, T.-L., Lu, D., Cone, N. E., Cao, F., Bigio, J. D., and Booth, J. R. (2007b). The interaction between orthographic and phonological information in children: An fMRI study. *Human brain mapping*, 28(9):880–891.

Bitan, T., Burman, D. D., Lu, D., Cone, N. E., Gitelman, D. R., Mesulam, M.-M., and Booth, J. R. (2006). Weaker top–down modulation from the left inferior frontal gyrus in children. *Neuroimage*, 33(3):991–998.

Bitan, T., Cheon, J., Lu, D., Burman, D. D., Gitelman, D. R., Mesulam, M.-M., and Booth, J. R. (2007c). Developmental changes in activation and effective connectivity in phonological processing. *Neuroimage*, 38(3):564–575.

Bitan, T. and Karni, A. (2003). Alphabetical knowledge from whole words training: effects of explicit instruction and implicit experience on learning script segmentation. *Cognitive Brain Research*, 16(3):323–337.

Bitan, T. and Karni, A. (2004). Procedural and declarative knowledge of word recognition and letter decoding in reading an artificial script. *Cognitive Brain Research*, 19(3):229–243.

Bolger, D. J., Perfetti, C. A., and Schneider, W. (2005). Cross-cultural effect on the brain revisited: Universal structures plus writing system variation. *Human brain mapping*, 25(1):92–104.

Bookheimer, S. (2002). Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. *Annual review of neuroscience*, 25(1):151–188.

Booth, J. R., Burman, D. D., Meyer, J. R., Gitelman, D. R., Parrish, T. B., and Mesulam, M. (2003). Relation between brain activation and lexical performance. *Human brain mapping*, 19(3):155–169.

Booth, J. R., Cho, S., Burman, D. D., and Bitan, T. (2007). Neural correlates of mapping from phonology to orthography in children performing an auditory spelling task. *Developmental science*, 10(4):441–451.

Boros, M., Anton, J.-L., Pech-Georgel, C., Grainger, J., Szwed, M., and Ziegler, J. C. (2016). Orthographic processing deficits in developmental dyslexia: Beyond the ventral visual stream. *NeuroImage*, 128:316–327.

Bowers, J. S., Davis, C. J., and Hanley, D. A. (2005). Interfering neighbours: The impact of novel word learning on the identification of visually similar words. *Cognition*, 97(3):B45–B54.

Breitenstein, C., Jansen, A., Deppe, M., Foerster, A.-F., Sommer, J., Wolbers, T., and Knecht, S. (2005). Hippocampus activity differentiates good from poor learners of a novel lexicon. *Neuroimage*, 25(3):958–968.

Brem, S., Bach, S., Kucian, K., Kujala, J. V., Guttorm, T. K., Martin, E., Lyytinen, H., Brandeis, D., and Richardson, U. (2010). Brain sensitivity to print emerges when children learn letter–speech sound correspondences. *Proceedings of the National Academy of Sciences*, 107(17):7939–7944.

Brysbaert, M. and Stevens, M. (2018). Power analysis and effect size in mixed effects models: a tutorial. *Journal of Cognition*, 1(1).

Byrne, B. J. (1998). *The foundation of literacy: The child's acquisition of the alphabetic principle*. Psychology Press.

Cao, F., Bitan, T., Chou, T.-L., Burman, D. D., and Booth, J. R. (2006). Deficient orthographic and phonological representations in children with dyslexia revealed by brain activation patterns. *Journal of Child Psychology and Psychiatry*, 47(10):1041–1050.

Cao, F., Brennan, C., and Booth, J. R. (2015). The brain adapts to orthography with experience: evidence from English and Chinese. *Developmental science*.

Caporale, N. and Dan, Y. (2008). Spike timing–dependent plasticity: a Hebbian learning rule. *Annu. Rev. Neurosci.*, 31:25–46.

Carreiras, M. and Perea, M. (2002). Masked priming effects with syllabic neighbors in a lexical decision task. *Journal of Experimental Psychology: human perception and performance*, 28(5):1228.

Carreiras, M., Quiñones, I., Hernández-Cabrera, J. A., and Duñabeitia, J. A. (2014). Orthographic coding: brain activation for letters, symbols, and digits. *Cerebral Cortex*, pages 4748–60.

Carreiras, M., Seghier, M., Baquero, S., Estévez, A., Lozano, A., Devlin, J., and Price, C. (2009). An anatomical signature for literacy. *Nature*, 461(7266):983–986.

Cassar, M. and 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., Nation, K., and others (2006). How does orthographic learning happen. *From inkmarks to ideas: Current issues in lexical processing*, pages 151–179.

Chang, Y.-N., Taylor, J., Rastle, K., and Monaghan, P. J. (2017). Exploring the relations between oral language and reading instruction in a computational model of reading. In *Proceedings of the 39th Annual Conference of the Cognitive Science Society.*, Austin, Tx. Cognitive Science Society.

Chetail, F. (2017). What do we do with what we learn? Statistical learning of orthographic regularities impacts written word processing. *Cognition*, 163:103–120.

Christle, C. A., Jolivette, K., and Nelson, C. M. (2005). Breaking the school to prison pipeline: Identifying school risk and protective factors for youth delinquency. *Exceptionality*, 13(2):69–88.

Church, J. A., Coalson, R. S., Lugar, H. M., Petersen, S. E., and Schlaggar, B. L. (2008). A developmental fMRI study of reading and repetition reveals changes in phonological and visual mechanisms over age. *Cerebral Cortex*, 18(9):2054–2065.

Cohen, L. and Dehaene, S. (2004). Specialization within the ventral stream: the case for the visual word form area. *Neuroimage*, 22(1):466–476.

Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M.-A., and Michel, F. (2000). The visual word form area. *Brain*, 123(2):291–307.

Cohen, L., Dehaene, S., Vinckier, F., Jobert, A., and Montavont, A. (2008). Reading normal and degraded words: Contribution of the dorsal and ventral visual pathways. *Neuroimage*, 40(1):353–366.

Cohen, L., Lehéricy, S., Chochon, F., Lemer, C., Rivaud, S., and Dehaene, S. (2002). Language-specific tuning of visual cortex? Functional properties of the Visual Word Form Area. *Brain*, 125(5):1054–1069.

Coltheart, M. (2005). Modeling Reading: The Dual-Route Approach. In Snowling, M. J. and Hulme, C., editors, *The science of reading: A handbook*. John Wiley & Sons.

Coltheart, M. (2006). Dual route and connectionist models of reading: An overview. *London Review of Education*, 4(1):5–17.

Coltheart, M., Curtis, B., Atkins, P., and Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological review*, 100(4):589.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., and Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological review*, 108(1):204.

Conway, C. M. and 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.

Craik, F. I. and Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behavior*, 11(6):671–684.

Cusack, R., Vicente-Grabovetsky, A., Mitchell, D. J., Wild, C. J., Auer, T., Linke, A. C., and Peelle, J. E. (2014). Automatic analysis (aa): efficient neuroimaging workflows and parallel processing using Matlab and XML. *Frontiers in neuroinformatics*, 8.

Davis, M. H., Di Betta, A. M., Macdonald, M. J., and Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. *Journal of cognitive neuroscience*, 21(4):803–820.

Davis, M. H. and Gaskell, M. G. (2009). A complementary systems account of word learning: neural and behavioural evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1536):3773–3800.

Dehaene, S. and Cohen, L. (2007). Cultural recycling of cortical maps. *Neuron*, 56(2):384–398.

Dehaene, S. and Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in cognitive sciences*, 15(6):254–262.

Dehaene, S., Cohen, L., Morais, J., and Kolinsky, R. (2015). Illiterate to literate: behavioural and cerebral changes induced by reading acquisition. *Nature Reviews Neuroscience*, 16(4):234–244.

Dehaene, S., Cohen, L., Sigman, M., and Vinckier, F. (2005). The neural code for written words: a proposal. *Trends in cognitive sciences*, 9(7):335–341.

Dehaene, S., Le Clec'H, G., Poline, J.-B., Le Bihan, D., and Cohen, L. (2002). The visual word form area: a prelexical representation of visual words in the fusiform gyrus. *Neuroreport*, 13(3):321–325.

Dehaene, S., Pegado, F., Braga, L., Ventura, P., Nunes, F., Jobert, A., Dehaene-Lambertz, G., Kolinsky, R., Morais, J., and Cohen, L. (2010). How learning to read changes the cortical networks for vision and language. *Science*, 330(6009):1359–1364.

Devlin, J. T., Matthews, P. M., and Rushworth, M. F. (2003). Semantic processing in the left inferior prefrontal cortex: a combined functional magnetic resonance imaging and transcranial magnetic stimulation study. *Journal of cognitive neuroscience*, 15(1):71–84.

Diamond, A. (2007). Interrelated and interdependent. *Developmental science*, 10(1):152–158.

Diekelmann, S. and Born, J. (2010). The memory function of sleep. *Nature reviews. Neuroscience*, 11(2):114.

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in psychology*, 5.

Djonlagic, I., Rosenfeld, A., Shohamy, D., Myers, C., Gluck, M., and Stickgold, R. (2009). Sleep enhances category learning. *Learning & Memory*, 16(12):751–755.

Dudai, Y. (2004). The neurobiology of consolidations, or, how stable is the engram? *Annu. Rev. Psychol.*, 55:51–86.

Dufau, S., Lété, B., Touzet, C., Glotin, H., Ziegler, J. C., and Grainger, J. (2010). A developmental perspective on visual word recognition: New evidence and a self-organising model. *European Journal of Cognitive Psychology*, 22(5):669–694.

Dumay, N. and Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science*, 18(1):35–39.

Durrant, S. J., Cairney, S. A., and Lewis, P. A. (2016). Cross-modal transfer of statistical information benefits from sleep. *cortex*, 78:85–99.

Durrant, S. J., Taylor, C., Cairney, S., and Lewis, P. A. (2011). Sleep-dependent consolidation of statistical learning. *Neuropsychologia*, 49(5):1322–1331.

Ehri, L. C. (1995). Phases of development in learning to read words by sight. *Journal of research in reading*, 18(2):116–125.

Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of reading*, 9(2):167–188.

Ehri, L. C., Nunes, S. R., Stahl, S. A., and Willows, D. M. (2001). Systematic phonics instruction helps students learn to read: Evidence from the National Reading Panel's meta-analysis. *Review of educational research*, 71(3):393–447.

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

Ellis, N. C. and Hooper, A. M. (2001). Why learning to read is easier in Welsh than in English: Orthographic transparency effects evinced with frequency-matched tests. *Applied Psycholinguistics*, 22(04):571–599.

Elman, J. L., Bates, E., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., and Plunkett, K. (1996). Rethinking innateness: Connectionism in a developmental framework. *Cambridge, MA: MIT Press. Embick, D., Marantz, A., Miyashita, Y., O'Neil W., Sakai, KL (2000). A syntactic specialization for Broca's area. Proceedings of the National Academy of Sciences*, 97(11):6150–6154.

Feld, G. B. and Born, J. (2017). Sculpting memory during sleep: concurrent consolidation and forgetting. *Current Opinion in Neurobiology*, 44:20–27.

Fiez, J. A. (1997). Phonology, semantics, and the role of the left inferior prefrontal cortex. *Human brain mapping*, 5(2):79–83.

Fischer, S., Hallschmid, M., Elsner, A. L., and Born, J. (2002). Sleep forms memory for finger skills. *Proceedings of the National Academy of Sciences*, 99(18):11987–11991.

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

Fiser, J. and Aslin, R. N. (2005). Encoding multielement scenes: statistical learning of visual feature hierarchies. *Journal of Experimental Psychology: General*, 134(4):521.

Forster, K. I. and Chambers, S. M. (1973). Lexical access and naming time. *Journal of verbal learning and verbal behavior*, 12(6):627–635.

Forster, K. I. and Davis, C. (1984). Repetition priming and frequency attenuation in lexical access. *Journal of experimental psychology: Learning, Memory, and Cognition*, 10(4):680.

Frankland, P. W. and Bontempi, B. (2005). The organization of recent and remote memories. *Nature reviews. Neuroscience*, 6(2):119.

French, R. M. (1992). Semi-distributed representations and catastrophic forgetting in connectionist networks. *Connection Science*, 4(3-4):365–377.

French, R. M. (1999). Catastrophic forgetting in connectionist networks. *Trends in cognitive sciences*, 3(4):128–135.

Friedmann, N. and Gvion, A. (2001). Letter position dyslexia. *Cognitive Neuropsychology*, 18(8):673–696.

Friston, K. J., Ashburner, J., Frith, C. D., Poline, J.-B., Heather, J. D., Frackowiak, R. S., and others (1995). Spatial registration and normalization of images. *Human brain mapping*, 3(3):165–189.

Gais, S., Mölle, M., Helms, K., and Born, J. (2002). Learning-dependent increases in sleep spindle density. *Journal of Neuroscience*, 22(15):6830–6834.

Gais, S., Plihal, W., Wagner, U., and Born, J. (2000). Early sleep triggers memory for early visual discrimination skills. *Nature Neuroscience*, 3(12):1335–1339.

Gakidou, E., Cowling, K., Lozano, R., and Murray, C. J. (2010). Increased educational attainment and its effect on child mortality in 175 countries between 1970 and 2009: a systematic analysis. *The Lancet*, 376(9745):959–974.

Gaskell, M. G. and Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition*, 89(2):105–132.

Gaskell, M. G. and Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and cognitive Processes*, 12(5-6):613–656.

Gaskell, M. G., Warker, J., Lindsay, S., Frost, R., Guest, J., Snowdon, R., and Stackhouse, A. (2014). Sleep underpins the plasticity of language production. *Psychological science*, 25(7):1457–1465.

Giroux, I. and Rey, A. (2009). Lexical and sublexical units in speech perception. *Cognitive Science*, 33(2):260–272.

Glezer, L. S., Jiang, X., and Riesenhuber, M. (2009). Evidence for highly selective neuronal tuning to whole words in the “visual word form area”. *Neuron*, 62(2):199–204.

Glezer, L. S., Kim, J., Rule, J., Jiang, X., and Riesenhuber, M. (2015). Adding words to the brain’s visual dictionary: novel word learning selectively sharpens orthographic representations in the VWFA. *Journal of Neuroscience*, 35(12):4965–4972.

Goodale, M. A. and Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in neurosciences*, 15(1):20–25.

Goswami, U. and Bryant, P. (1990). *Phonological skills and learning to read*. Wiley Online Library.

Gough, P. B. and Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and special education*, 7(1):6–10.

Gough, P. M., Nobre, A. C., and Devlin, J. T. (2005). Dissociating linguistic processes in the left inferior frontal cortex with transcranial magnetic stimulation. *Journal of Neuroscience*, 25(35):8010–8016.

Grainger, J. (2008). Cracking the orthographic code: An introduction. *Language and Cognitive Processes*, 23(1):1–35.

Grainger, J., Bouttevin, S., Truc, C., Bastien, M., and Ziegler, J. (2003). Word superiority, pseudoword superiority, and learning to read: A comparison of dyslexic and normal readers. *Brain and Language*, 87(3):432–440.

Grainger, J. and van Heuven, W.J.B. (2004). Modeling letter position coding in printed word perception. In *Bonin, P. (ed.) The mental lexicon*, pages 2–23. New York : Nova Science.

Gusnard, D. A. and Raichle, M. E. (2001). Searching for a baseline: functional imaging and the resting human brain. *Nature reviews. Neuroscience*, 2(10):685.

Hagoort, P., Indefrey, P., Brown, C., Herzog, H., Steinmetz, H., and Seitz, R. J. (1999). The neural circuitry involved in the reading of German words and pseudowords: a PET study. *Cognitive Neuroscience, Journal of*, 11(4):383–398.

Harm, M. W. and Seidenberg, M. S. (2004). Computing the meanings of words in reading: cooperative division of labor between visual and phonological processes. *Psychological review*, 111(3):662.

Hashimoto, R. and Sakai, K. L. (2004). Learning letters in adulthood: direct visualization of cortical plasticity for forming a new link between orthography and phonology. *Neuron*, 42(2):311–322.

Havas, V., Taylor, J. S. H., Vaquero, L., Diego-Balaguer, R. d., Rodríguez-Fornells, A., and Davis, M. H. (2017). Semantic and phonological schema influence spoken word learning and overnight consolidation. *The Quarterly Journal of Experimental Psychology*, 0(ja):1–29.

Hawkins, E., Astle, D. E., and Rastle, K. (2015). Semantic advantage for learning new phonological form representations. *Journal of cognitive neuroscience*, 27(4):775–786.

Hawkins, E. A. and Rastle, K. (2016). How does the provision of semantic information influence the lexicalization of new spoken words? *The Quarterly Journal of Experimental Psychology*, 69(7):1322–1339.

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological approach*. John Wiley & Sons.

Henderson, L., Weighall, A., and Gaskell, G. (2013). Learning new vocabulary during childhood: Effects of semantic training on lexical consolidation and integration. *Journal of Experimental Child Psychology*, 116(3):572–592.

Henson, R., Shallice, T., and Dolan, R. (2000). Neuroimaging evidence for dissociable forms of repetition priming. *Science*, 287(5456):1269–1272.

Henson, R. N. (2001). Repetition effects for words and nonwords as indexed by event-related fMRI: A preliminary study. *Scandinavian journal of psychology*, 42(3):179–186.

Hickok, G. and Small, S. L. (2015). *Neurobiology of language*. Academic Press.

Hirshorn, E. A. and Fiez, J. A. (2014). Using artificial orthographies for studying cross-linguistic differences in the cognitive and neural profiles of reading. *Journal of Neurolinguistics*, 31:69–85.

Hoeft, F., Meyler, A., Hernandez, A., Juel, C., Taylor-Hill, H., Martindale, J. L., McMillon, G., Kolchugina, G., Black, J. M., Faizi, A., and others (2007). Functional and morphometric brain dissociation between dyslexia and reading ability. *Proceedings of the National Academy of Sciences*, 104(10):4234–4239.

Hothorn, T., Bretz, F., Westfall, P., and Heiberger, R. M. (2008). Multcomp: Simultaneous Inference in General Parametric Models—R Package Version 1.0–0. *R Foundation for Statistical Computing. Vienna, Austria*.

Hunt, R. R. (1995). The subtlety of distinctiveness: What von restorff really did. *Psychonomic Bulletin & Review*, 2(1):105–112.

Jackson, N. E. and Coltheart, M. (2001). *Macquarie monographs in cognitive science. Routes to reading success and failure: Toward an integrated cognitive psychology of atypical reading*. New York: Psychology Press.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of memory and language*, 59(4):434–446.

James, T. W., Humphrey, G. K., Gati, J. S., Menon, R. S., and Goodale, M. A. (2002). Differential effects of viewpoint on object-driven activation in dorsal and ventral streams. *Neuron*, 35(4):793–801.

Ji, D. and Wilson, M. A. (2007). Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nature neuroscience*, 10(1):100.

Judd, C. M., Westfall, J., and Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: a new and comprehensive solution to a pervasive but largely ignored problem. *Journal of personality and social psychology*, 103(1):54.

Kapnoula, E. C. and McMurray, B. (2016). Newly learned word forms are abstract and integrated immediately after acquisition. *Psychonomic bulletin & review*, 23(2):491–499.

Kapnoula, E. C., Packard, S., Gupta, P., and McMurray, B. (2015). Immediate lexical integration of novel word forms. *Cognition*, 134:85–99.

Karni, A. and Sagi, D. (1993). The time course of learning a visual skill. *Nature*, 365(6443):250–252.

Karni, A., Tanne, D., Rubenstein, B. S., Askenasy, J. J., and Sagi, D. (1994). Dependence on REM sleep of overnight improvement of a perceptual skill. *Science (New York, N.Y.)*, 265(5172):679–682.

Kherif, F., Josse, G., and Price, C. J. (2011). Automatic top-down processing explains common left occipito-temporal responses to visual words and objects. *Cerebral Cortex*, 21(1):103–114.

Kim, R., Seitz, A., Feenstra, H., and Shams, L. (2009). Testing assumptions of statistical learning: is it long-term and implicit? *Neuroscience letters*, 461(2):145–149.

Korman, M., Doyon, J., Doljansky, J., Carrier, J., Dagan, Y., and Karni, A. (2007). Daytime sleep condenses the time course of motor memory consolidation. *Nature neuroscience*, 10(9):1206.

Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., and Mishkin, M. (2013). The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends in cognitive sciences*, 17(1):26–49.

Laing, E. and Hulme, C. (1999). Phonological and semantic processes influence beginning readers' ability to learn to read words. *Journal of experimental child psychology*, 73(3):183–207.

Leach, L. and Samuel, A. G. (2007). Lexical configuration and lexical engagement: When adults learn new words. *Cognitive psychology*, 55(4):306–353.

Levy, F. and Murnane, R. J. (2004). Education and the Changing Job Market. *Educational leadership*, 62(2):80.

Lindsay, S. and Gaskell, M. G. (2010). A complementary systems account of word learning in L1 and L2. *Language Learning*, 60(s2):45–63.

Lobier, M., Peyrin, C., Le Bas, J.-F., and Valdois, S. (2012). Pre-orthographic character string processing and parietal cortex: a role for visual attention in reading? *Neuropsychologia*, 50(9):2195–2204.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychological bulletin*, 109(2):163.

Magnié, M. N., Besson, M., Poncet, M., and Dolisi, C. (2003). The Snodgrass and Vanderwart set revisited: Norms for object manipulability and for pictorial ambiguity of objects, chimeric objects, and nonobjects. *Journal of Clinical and Experimental Neuropsychology*, 25(4):521–560.

Mano, Q. R., Humphries, C., Desai, R. H., Seidenberg, M. S., Osmon, D. C., Stengel, B. C., and Binder, J. R. (2013). The role of left occipitotemporal cortex in reading: reconciling stimulus, task, and lexicality effects. *Cerebral Cortex*, 23(4):988–1001.

Marr, D. (1970). A theory for cerebral neocortex. *Proceedings of the Royal Society of London B: Biological Sciences*, 176(1043):161–234.

Martin, A., Schurz, M., Kronbichler, M., and Richlan, F. (2015). Reading in the brain of children and adults: A meta-analysis of 40 functional magnetic resonance imaging studies. *Human brain mapping*, 36(5):1963–1981.

Masson, M. E. and Isaak, M. I. (1999). Masked priming of words and nonwords in a naming task: Further evidence for a nonlexical basis for priming. *Memory & Cognition*, 27(3):399–412.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., and Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, 94:305–315.

Maye, J., Werker, J. F., and Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82(3):B101–B111.

McArthur, G., Castles, A., Kohnen, S., Larsen, L., Jones, K., Anandakumar, T., and Banales, E. (2015). Sight word and phonics training in children with dyslexia. *Journal of learning disabilities*, 48(4):391–407.

McClelland, J. L. and Johnston, J. C. (1977). The role of familiar units in perception of words and nonwords. *Attention, Perception, & Psychophysics*, 22(3):249–261.

McClelland, J. L., McNaughton, B. L., and O'reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological review*, 102(3):419.

McClelland, J. L. and Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological review*, 88(5):375.

McCloskey, M. and Cohen, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. *Psychology of learning and motivation*, 24:109–165.

McDevitt, E. A., Rowe, K. M., Brady, M., Duggan, K. A., and Mednick, S. C. (2014). The benefit of offline sleep and wake for novel object recognition. *Experimental Brain Research*, 232(5):1487–1496.

McKague, M., Pratt, C., and Johnston, M. B. (2001). The effect of oral vocabulary on reading visually novel words: A comparison of the dual-route-cascaded and triangle frameworks. *Cognition*, 80(3):231–262.

McKay, A., Davis, C., Savage, G., and Castles, A. (2008). Semantic involvement in reading aloud: evidence from a nonword training study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(6):1495.

McNorgan, C., Alvarez, A., Bhullar, A., Gayda, J., and Booth, J. R. (2011). Prediction of reading skill several years later depends on age and brain region: Implications for developmental models of reading. *Journal of Neuroscience*, 31(26):9641–9648.

Mechelli, A., Crinion, J. T., Long, S., Friston, K. J., Ralph, M., Patterson, K., McClelland, J. L., Price, C. J., and others (2005). Dissociating reading processes on the basis of neuronal interactions. *Cognitive Neuroscience, Journal of*, 17(11):1753–1765.

Mei, L., Xue, G., Lu, Z.-L., He, Q., Zhang, M., Xue, F., Chen, C., and Dong, Q. (2013a). Orthographic transparency modulates the functional asymmetry in the fusiform cortex: an artificial language training study. *Brain and language*, 125(2):165–172.

Mei, L., Xue, G., Lu, Z.-L., He, Q., Zhang, M., Xue, F., Chen, C., and Dong, Q. (2013b). Orthographic transparency modulates the functional asymmetry in the fusiform cortex: An artificial language training study. *Brain and language*, 125(2):165–172.

Meital, N., Korinth, S. P., and Karni, A. (2013). Plasticity in the adult oculomotor system: offline consolidation phase gains in saccade sequence learning. *Brain Research*, 1528:42–48.

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

Merkx, M., Rastle, K., and Davis, M. H. (2011). The acquisition of morphological knowledge investigated through artificial language learning. *The Quarterly Journal of Experimental Psychology*, 64(6):1200–1220.

Milner, A. D. and Goodale, M. A. (1993). Visual pathways to perception and action. *Progress in brain research*, 95:317–337.

Milner, A. D. and Goodale, M. A. (2008). Two visual systems re-viewed. *Neuropsychologia*, 46(3):774–785.

Mishkin, M. and Ungerleider, L. G. (1982). Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behavioural brain research*, 6(1):57–77.

Monaghan, J. and Ellis, A. W. (2002). What exactly interacts with spelling-sound consistency in word naming? *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 28(1):183–206.

Monaghan, P., Christiansen, M. H., and Fitneva, S. A. (2011). The arbitrariness of the sign: Learning advantages from the structure of the vocabulary. *Journal of Experimental Psychology: General*, 140(3):325.

Morais, J. (2017). Literacy and democracy. *Language, Cognition and Neuroscience*, 0(0):1–22.

Morais, J. and Kolinsky, R. (2005). Literacy and cognitive change. *The science of reading: A handbook*, pages 188–203.

Morey, R. D., Rouder, J. N., Jamil, T., and Morey, M. R. D. (2015). Package 'BayesFactor'. *URL< http://cran. r-project. org/web/packages/BayesFactor/BayesFactor. pdf>(accessed 10.06. 15)*.

Morgan, M. and Kett, M. (2003). The prison adult literacy survey. *Results and implications*.

Nation, K. (2009). Form–meaning links in the development of visual word recognition. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 364(1536):3665–3674.

Nation, K. (2017). Nurturing a lexical legacy: reading experience is critical for the development of word reading skill. *npj Science of Learning*, 2(1):3.

Nation, K. and Cocksey, J. (2009). The relationship between knowing a word and reading it aloud in children's word reading development. *Journal of Experimental Child Psychology*, 103(3):296–308.

Nation, K. and Snowling, M. J. (2004). Beyond phonological skills: Broader language skills contribute to the development of reading. *Journal of Research in Reading*, 27(4):342–356.

Ossmy, O., Ben-Shachar, M., and Mukamel, R. (2014). Decoding letter position in word reading. *Cortex*, 59:74–83.

O'Brien, B. A. (2014). The development of sensitivity to sublexical orthographic constraints: An investigation of positional frequency and consistency using a wordlikeness choice task. *Reading Psychology*, 35(4):285–311.

Pacton, S., Fayol, M., and Perruchet, P. (2005). Children's implicit learning of graphotactic and morphological regularities. *Child development*, 76(2):324–339.

Pattamadilok, C., Chanoine, V., Pallier, C., Anton, J.-L., Nazarian, B., Belin, P., and Ziegler, J. C. (2017). Automaticity of phonological and semantic processing during visual word recognition. *NeuroImage*, 149:244–255.

Peelle, J. E. (2014). Methodological challenges and solutions in auditory functional magnetic resonance imaging. *Frontiers in neuroscience*, 8.

Peigneux, P., Laureys, S., Fuchs, S., Collette, F., Perrin, F., Reggers, J., Phillips, C., Degueldre, C., Del Fiore, G., Aerts, J., and others (2004). Are spatial memories strengthened in the human hippocampus during slow wave sleep? *Neuron*, 44(3):535–545.

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, 162(1):8–13.

Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics*, 2.

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific studies of reading*, 11(4):357–383.

Perrachione, T. K. and Ghosh, S. S. (2013). Optimized design and analysis of sparse-sampling fMRI experiments. *Frontiers in neuroscience*, 7.

Perry, C., Ziegler, J. C., and Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: the CDP+ model of reading aloud. *Psychological review*, 114(2):273.

Peyrin, C., Démonet, J. F., N'Guyen-Morel, M. A., Le Bas, J. F., and Valdois, S. (2011). Superior parietal lobule dysfunction in a homogeneous group of dyslexic children with a visual attention span disorder. *Brain and language*, 118(3):128–138.

Peyrin, C., Lallier, M., and Valdois, S. (2008). Visual attention span brain mechanisms in normal and dyslexic readers. *Neuropsychology and cognition of language: Behavioural, neuropsychological and neuroimaging studies of oral and written language*, pages 22–40.

Pinheiro, J. C. and Bates, D. M. (2000). Linear mixed-effects models: basic concepts and examples. *Mixed-effects models in S and S-Plus*, pages 3–56.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., and Patterson, K. (1996). Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychological review*, 103(1):56.

Plaut, D. C. and Vande Velde, A. K. (2017). Statistical learning of parts and wholes: A neural network approach. *Journal of Experimental Psychology: General*, 146(3):318.

Plihal, W. and Born, J. (1997). Effects of early and late nocturnal sleep on declarative and procedural memory. *Journal of Cognitive Neuroscience*, 9(4):534–547.

Price, C. J. (2012). A review and synthesis of the first 20years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage*, 62(2):816–847.

Price, C. J. and Devlin, J. T. (2003). The myth of the visual word form area. *Neuroimage*, 19(3):473–481.

Price, C. J. and Devlin, J. T. (2011a). The interactive account of ventral occipitotemporal contributions to reading. *Trends in cognitive sciences*, 15(6):246–253.

Price, C. J. and Devlin, J. T. (2011b). The interactive account of ventral occipitotemporal contributions to reading. *Trends in cognitive sciences*, 15(6):246–253.

Price, C. J. and Mechelli, A. (2005). Reading and reading disturbance. *Current opinion in neurobiology*, 15(2):231–238.

Price, C. J., Wise, R. J. S., and Frackowiak, R. S. J. (1996). Demonstrating the Implicit Processing of Visually Presented Words and Pseudowords. *Cerebral Cortex*, 6(1):62–70.

Protopapas, A. (2007). CheckVocal: a program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior Research Methods*, 39(4):859–862.

Protopapas, A., Orfanidou, E., Taylor, J. S. H., Karavasilis, E., Kapnoula, E. C., Panagiotaropoulou, G., Velonakis, G., Poulou, L. S., Smyrnis, N., and Kelekis, D. (2016). Evaluating cognitive models of visual word recognition using fMRI: Effects of lexical and sublexical variables. *NeuroImage*, 128:328–341.

Qin, Y.-L., McNaughton, B. L., Skaggs, W. E., and Barnes, C. A. (1997). Memory reprocessing in corticocortical and hippocampocortical neuronal ensembles. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 352(1360):1525–1533.

Raichle, M. E. (2015). The brain's default mode network. *Annual review of neuroscience*, 38:433–447.

Rastle, K. (2007). Visual word recognition. In Gaskell, M. G., editor, *The Oxford handbook of psycholinguistics*. Oxford University Press.

Rastle, K. and Coltheart, M. (1999). Serial and strategic effects in reading aloud. *Journal of Experimental Psychology: Human perception and performance*, 25(2):482.

Rastle, K. and Coltheart, M. (2000). Serial processing in reading aloud: Reply to Zorzi (2000). *Journal of Experimental Psychology: Human Perception and Performance*, 26(3):1232–1235.

Rastle, K., Croot, K. P., Harrington, J. M., and Coltheart, M. (2005). Characterizing the motor execution stage of speech production: consonantal effects on delayed naming latency and onset duration. *Journal of Experimental Psychology: Human Perception and Performance*, 31(5):1083.

Rastle, K., Davis, M. H., and New, B. (2004). The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin & Review*, 11(6):1090–1098.

Rastle, K., Harrington, J., and Coltheart, M. (2002). 358,534 nonwords: The ARC nonword database. *The Quarterly Journal of Experimental Psychology: Section A*, 55(4):1339–1362.

Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., and Seidenberg, M. S. (2001). How Psychological Science Informs the Teaching of Reading. *Psychological Science in the Public Interest*, 2(2):31–74.

Reilhac, C., Peyrin, C., Démonet, J.-F., and Valdois, S. (2013). Role of the superior parietal lobules in letter-identity processing within strings: FMRI evidence from skilled and dyslexicreaders. *Neuropsychologia*, 51(4):601–612.

Ricketts, J., Nation, K., and Bishop, D. V. (2007). Vocabulary is important for some, but not all reading skills. *Scientific Studies of Reading*, 11(3):235–257.

Rosazza, C., Cai, Q., Minati, L., Paulignan, Y., and Nazir, T. A. (2009). Early involvement of dorsal and ventral pathways in visual word recognition: an ERP study. *Brain research*, 1272:32–44.

Rose, J. (2006). Independent Review of the Teaching of Early Reading, Final Report, March 2006. *Annesley, Nottingham: DfES.*

Rouder, J. N., Morey, R. D., Verhagen, J., Swagman, A. R., and Wagenmakers, E.-J. (2017). Bayesian analysis of factorial designs. *Psychological Methods*, 22(2):304.

Roy, B. C., Frank, M. C., DeCamp, P., Miller, M., and Roy, D. (2015). Predicting the birth of a spoken word. *Proceedings of the National Academy of Sciences*, 112(41):12663–12668.

Rumelhart, D. E. and McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: II. The contextual enhancement effect and some tests and extensions of the model. *Psychological review*, 89(1):60.

Saffran, J. R., Aslin, R. N., and Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, pages 1926–1928.

Scarborough, D. L., Cortese, C., and Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human perception and performance*, 3(1):1.

Schanzenbach, D. W. (2012). Limitations of experiments in education research. *Education Finance and Policy*, 7(2):219–232.

Scoville, W. B. and Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of neurology, neurosurgery, and psychiatry*, 20(1):11.

Seghier, M. L., Lee, H. L., Schofield, T., Ellis, C. L., and Price, C. J. (2008). Inter-subject variability in the use of two different neuronal networks for reading aloud familiar words. *Neuroimage*, 42(3):1226–1236.

Seidenberg, M. S. and McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological review*, 96(4):523.

Shafritz, K. M., Gore, J. C., and Marois, R. (2002). The role of the parietal cortex in visual feature binding. *Proceedings of the National Academy of Sciences*, 99(16):10917–10922.

Shapiro, L. R. and Solity, J. (2016). Differing effects of two synthetic phonics programmes on early reading development. *British Journal of Educational Psychology*, 86(2):182–203.

Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition*, 55(2):151–218.

Share, D. L. (1999). Phonological recoding and orthographic learning: A direct test of the self-teaching hypothesis. *Journal of experimental child psychology*, 72(2):95–129.

Share, D. L. (2008). On the Anglocentricities of current reading research and practice: the perils of overreliance on an" outlier" orthography. *Psychological bulletin*, 134(4):584.

Sibley, D. E., Kello, C. T., Plaut, D. C., and Elman, J. L. (2008). Large-Scale Modeling of Wordform Learning and Representation. *Cognitive Science*, 32(4):741–754.

Sibley, D. E., Kello, C. T., and Seidenberg, M. S. (2010). Learning orthographic and phonological representations in models of monosyllabic and bisyllabic naming. *European Journal of Cognitive Psychology*, 22(5):650–668.

Snodgrass, J. G. and Corwin, J. (1988). Pragmatics of measuring recognition memory: applications to dementia and amnesia. *Journal of Experimental Psychology: General*, 117(1):34.

Snoeren, N. D., Gaskell, M. G., and Di Betta, A. M. (2009). The perception of assimilation in newly learned novel words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(2):542.

Snow, P. C. (2016). Elizabeth Usher Memorial Lecture: Language is literacy is language-Positioning speech-language pathology in education policy, practice, paradigms and polemics. *International journal of speech-language pathology*, 18(3):216–228.

Snowling, M. J., Adams, J. W., Bowyer-Crane, C., and Tobin, V. (2000). Levels of literacy among juvenile offenders: The incidence of specific reading difficulties. *Criminal Behaviour and Mental Health*, 10(4):229–241.

Snowling, M. J. and Hulme, C. (2012). Annual Research Review: The nature and classification of reading disorders–a commentary on proposals for DSM-5. *Journal of Child Psychology and Psychiatry*, 53(5):593–607.

Song, Y., Bu, Y., Hu, S., Luo, Y., and Liu, J. (2010a). Short-term language experience shapes the plasticity of the visual word form area. *Brain research*, 1316:83–91.

Song, Y., Hu, S., Li, X., Li, W., and Liu, J. (2010b). The role of top-down task context in learning to perceive objects. *The Journal of Neuroscience*, 30(29):9869–9876.

Squire, L. R. and Alvarez, P. (1995). Retrograde amnesia and memory consolidation: a neurobiological perspective. *Current opinion in neurobiology*, 5(2):169–177.

Stark, C. E. and McClelland, J. L. (2000). Repetition priming of words, pseudowords, and nonwords. *Journal of experimental psychology: Learning, memory, and cognition*, 26(4):945.

Stickgold, R., Whidbee, D., Schirmer, B., Patel, V., and Hobson, J. A. (2000). Visual discrimination task improvement: A multi-step process occurring during sleep. *Journal of Cognitive Neuroscience*, 12(2):246–254.

Stuart, M., Masterson, J., Dixon, M., and Quinlan, P. (1999). Inferring sublexical correspondences from sight vocabulary: Evidence from 6-and 7-year-olds. *The Quarterly Journal of Experimental Psychology: Section A*, 52(2):353–366.

Sutherland, G. R. and McNaughton, B. (2000). Memory trace reactivation in hippocampal and neocortical neuronal ensembles. *Current opinion in neurobiology*, 10(2):180–186.

Szwed, M., Dehaene, S., Kleinschmidt, A., Eger, E., Valabrègue, R., Amadon, A., and Cohen, L. (2011). Specialization for written words over objects in the visual cortex. *Neuroimage*, 56(1):330–344.

Taft, M. (1992). The body of the BOSS: Subsyllabic units in the lexical processing of polysyllabic words. *Journal of Experimental Psychology: Human Perception and Performance*, 18(4):1004.

Takashima, A., Bakker, I., Van Hell, J. G., Janzen, G., and McQueen, J. M. (2014a). Richness of information about novel words influences how episodic and semantic memory networks interact during lexicalization. *NeuroImage*, 84:265–278.

Takashima, A., Bakker, I., Van Hell, J. G., Janzen, G., and McQueen, J. M. (2017). Interaction between episodic and semantic memory networks in the acquisition and consolidation of novel spoken words. *Brain and language*, 167:44–60.

Takashima, A., Hulzink, I., Wagensveld, B., and Verhoeven, L. (2016). Emergence of representations through repeated training on pronouncing novel letter combinations leads to efficient reading. *Neuropsychologia*, 89:14–30.

Takashima, A., Wagensveld, B., Van Turennout, M., Zwitserlood, P., Hagoort, P., and Verhoeven, L. (2014b). Training-induced neural plasticity in visual-word decoding and the role of syllables. *Neuropsychologia*, 61:299–314.

Talamini, L. M., Nieuwenhuis, I. L., Takashima, A., and Jensen, O. (2008). Sleep directly following learning benefits consolidation of spatial associative memory. *Learning & Memory*, 15(4):233–237.

Tamminen, J., Davis, M. H., Merkx, M., and Rastle, K. (2012). The role of memory consolidation in generalisation of new linguistic information. *Cognition*, 125(1):107–112.

Tamminen, J., Davis, M. H., and Rastle, K. (2015). From specific examples to general knowledge in language learning. *Cognitive psychology*, 79:1–39.

Tamminen, J., Ralph, M. A. L., and Lewis, P. A. (2013). The role of sleep spindles and slow-wave activity in integrating new information in semantic memory. *Journal of Neuroscience*, 33(39):15376–15381.

Tamminen, J., Ralph, M. A. L., and Lewis, P. A. (2017). Targeted memory reactivation of newly learned words during sleep triggers REM-mediated integration of new memories and existing knowledge. *Neurobiology of learning and memory*, 137:77–82.

Taylor, J. S. H., Davis, M. H., and Rastle, K. (2017). Comparing and validating methods of reading instruction using behavioural and neural findings in an artificial orthography. *Journal of Experimental Psychology: General*, 146(6):826.

Taylor, J. S. H., Plunkett, K., and Nation, K. (2011). The influence of consistency, frequency, and semantics on learning to read: an artificial orthography paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(1):60.

Taylor, J. S. H., Rastle, K., and Davis, M. H. (2013). Can cognitive models explain brain activation during word and pseudoword reading? A meta-analysis of 36 neuroimaging studies. *Psychological bulletin*, 139(4):766.

Taylor, J. S. H., Rastle, K., and Davis, M. H. (2014a). Distinct neural specializations for learning to read words and name objects. *Journal of cognitive neuroscience*, 26(9):2128–2154.

Taylor, J. S. H., Rastle, K., and Davis, M. H. (2014b). Interpreting response time effects in functional imaging studies. *NeuroImage*, 99:419–433.

Thompson, G. B., Cottrell, D. S., and Fletcher-Flinn, C. M. (1996). Sublexical orthographic–phonological relations early in the acquisition of reading: The knowledge sources account. *Journal of Experimental Child Psychology*, 62(2):190–222.

Torgerson, C., Brooks, G., and Hall, J. (2006). *A systematic review of the research literature on the use of phonics in the teaching of reading and spelling*. DfES Publications Nottingham.

Torgerson, C. J., Torgerson, D. J., and Director, Y. T. U. (2013). Randomised trials in education: An introductory handbook. *London: Education Endowment Foundation*.

Treiman, R., Kessler, B., Zevin, J. D., Bick, S., and Davis, M. (2006). Influence of consonantal context on the reading of vowels: Evidence from children. *Journal of Experimental Child Psychology*, 93(1):1–24.

Treiman, R., Mullennix, J., Bijeljac-Babic, R., and Richmond-Welty, E. D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology: General*, 124(2):107.

Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., Witter, M. P., and Morris, R. G. (2007). Schemas and memory consolidation. *Science*, 316(5821):76–82.

Tucker, M. A., Hirota, Y., Wamsley, E. J., Lau, H., Chaklader, A., and Fishbein, W. (2006). A daytime nap containing solely non-REM sleep enhances declarative but not procedural memory. *Neurobiology of learning and memory*, 86(2):241–247.

Turkeltaub, P. E., Flowers, D. L., Lyon, L. G., and Eden, G. F. (2008). Development of ventral stream representations for single letters. *Annals of the New York Academy of Sciences*, 1145(1):13–29.

Twomey, T., Kawabata Duncan, K. J., Price, C. J., and Devlin, J. T. (2011). Top-down modulation of ventral occipito-temporal responses during visual word recognition. *Neuroimage*, 55(3):1242–1251.

Ungerleider, L. G. and Haxby, J. V. (1994). 'What'and 'where'in the human brain. *Current opinion in neurobiology*, 4(2):157–165.

Van Assche, E., Duyck, W., and Gollan, T. H. (2016). Linking recognition and production: Cross-modal transfer effects between picture naming and lexical decision during first and second language processing in bilinguals. *Journal of Memory and Language*, 89:37–54.

van de Ven, G. M., Trouche, S., McNamara, C. G., Allen, K., and Dupret, D. (2016). Hippocampal offline reactivation consolidates recently formed cell assembly patterns during sharp wave-ripples. *Neuron*, 92(5):968–974.

van den Boer, M. and de Jong, P. F. (2015). Parallel and serial reading processes in children's word and nonword reading. *Journal of Educational Psychology*, 107(1):141.

Van Heuven, W. J., Mandera, P., Keuleers, E., and Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology*, 67(6):1176–1190.

Vinckier, F., Dehaene, S., Jobert, A., Dubus, J. P., Sigman, M., and Cohen, L. (2007). Hierarchical coding of letter strings in the ventral stream: dissecting the inner organization of the visual word-form system. *Neuron*, 55(1):143–156.

Vinckier, F., Naccache, L., Papeix, C., Forget, J., Hahn-Barma, V., Dehaene, S., and Cohen, L. (2006). "What" and "Where" in Word Reading: Ventral Coding of Written Words Revealed by Parietal Atrophy. *Journal of cognitive neuroscience*, 18(12):1998–2012.

Viner, R. M., Ozer, E. M., Denny, S., Marmot, M., Resnick, M., Fatusi, A., and Currie, C. (2012). Adolescence and the social determinants of health. *The Lancet*, 379(9826):1641–1652.

Vousden, J. I., Ellefson, M. R., Solity, J., and Chater, N. (2011). Simplifying Reading: Applying the Simplicity Principle to Reading. *Cognitive Science*, 35(1):34–78.

Vuilleumier, P., Henson, R. N., Driver, J., and Dolan, R. J. (2002). Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. *Nature neuroscience*, 5(5):491.

Walker, M. P. and Stickgold, R. (2004). Sleep-dependent learning and memory consolidation. *Neuron*, 44(1):121–133.

Wang, H.-C., Savage, G., Gaskell, M. G., Paulin, T., Robidoux, S., and Castles, A. (2016). Bedding down new words: Sleep promotes the emergence of lexical competition in visual word recognition. *Psychonomic bulletin & review*, pages 1–8.

Warker, J. A. and Dell, G. S. (2006). Speech errors reflect newly learned phonotactic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(2):387.

Waters, G. S. and Seidenberg, M. S. (1985). Spelling-sound effects in reading: Time-course and decision criteria. *Memory & Cognition*, 13(6):557–572.

Weekes, B. S. (1997). Differential effects of number of letters on word and nonword naming latency. *The Quarterly Journal of Experimental Psychology: Section A*, 50(2):439–456.

Whitehouse, A. J., Line, E. A., Watt, H. J., and Bishop, D. V. (2009). Qualitative aspects of developmental language impairment relate to language and literacy outcome in adulthood. *International Journal of Language & Communication Disorders*, 44(4):489–510.

Wilson, M. A. (2002). Hippocampal memory formation, plasticity, and the role of sleep. *Neurobiology of learning and memory*, 78(3):565–569.

Wilson, M. A., McNaughton, B. L., and others (1994). Reactivation of hippocampal ensemble memories during sleep. *Science*, 265(5172):676–679.

Wonnacott, E., Brown, H., and Nation, K. (2017). Skewing the evidence: The effect of input structure on child and adult learning of lexically based patterns in an artificial language. *Journal of Memory and Language*, 95:36–48.

Wonnacott, E. and Newport, E. L. (2005). Novelty and regularization: The effect of novel instances on rule formation. In *BUCLD 29: Proceedings of the 29th annual Boston University conference on language development*, pages 663–673.

Wonnacott, E., Newport, E. L., and Tanenhaus, M. K. (2008). Acquiring and processing verb argument structure: Distributional learning in a miniature language. *Cognitive psychology*, 56(3):165–209.

Wyse, D. and Goswami, U. (2008). Synthetic phonics and the teaching of reading. *British Educational Research Journal*, 34(6):691–710.

Xue, G., Chen, C., Jin, Z., and Dong, Q. (2006). Language experience shapes fusiform activation when processing a logographic artificial language: an fMRI training study. *Neuroimage*, 31(3):1315–1326.

Yang, G., Lai, C. S. W., Cichon, J., Ma, L., Li, W., and Gan, W.-B. (2014). Sleep promotes branch-specific formation of dendritic spines after learning. *Science*, 344(6188):1173–1178.

Yoncheva, Y. N., Blau, V. C., Maurer, U., and Ph.D, B. D. M. (2010). Attentional Focus During Learning Impacts N170 ERP Responses to an Artificial Script. *Developmental Neuropsychology*, 35(4):423–445.

Yoncheva, Y. N., Wise, J., and McCandliss, B. (2015). Hemispheric specialization for visual words is shaped by attention to sublexical units during initial learning. *Brain and language*, 145:23–33.

Zeelenberg, R., Wagenmakers, E.-J., and Shiffrin, R. M. (2004). Nonword repetition priming in lexical decision reverses as a function of study task and speed stress. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(1):270.

Zevin, J. D. and Seidenberg, M. S. (2002). Age of acquisition effects in word reading and other tasks. *Journal of Memory and language*, 47(1):1–29.

Zevin, J. D. and Seidenberg, M. S. (2006). Simulating consistency effects and individual differences in nonword naming: A comparison of current models. *Journal of Memory and Language*, 54(2):145–160.

Ziegler, J. C. and Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, 131(1):3–29.

Ziegler, J. C., Grainger, J., and Brysbaert, M. (2010). Modelling word recognition and reading aloud. *European Journal of Cognitive Psychology*, 22(5):641–649.

Ziegler, J. C., Perry, C., Jacobs, A. M., and Braun, M. (2001). Identical words are read differently in different languages. *Psychological Science*, 12(5):379–384.

Zoccolotti, P., De Luca, M., Di Pace, E., Gasperini, F., Judica, A., and Spinelli, D. (2005). Word length effect in early reading and in developmental dyslexia. *Brain and language*, 93(3):369–373.

Zorzi, M., Houghton, G., and Butterworth, B. (1998). Two routes or one in reading aloud? A connectionist dual-process model. *Journal of Experimental Psychology: Human Perception and Performance*, 24(4):1131.