Understanding individual differences in learning and consolidating new vocabulary

Emma Louise James

Doctor of Philosophy

University of York
Psychology

February 2019

Abstract

Sleep plays an active role in consolidating new words into vocabulary, in line with Complementary Learning Systems models that describe "offline" integration of rapidly acquired memories into longer-term stores. However, these models describe average learning processes, and do not account for individual variability. In this thesis, it is proposed that existing vocabulary knowledge may be one source of variation in supporting rapid integration of new words into memory, based on extant evidence that existing vocabulary knowledge predicts overnight improvements in new word memory. To test causal hypotheses of this relationship, eight experiments manipulated the extent to which trained pseudowords were similar to existing lexical items. Semantic and word-form lexical similarity to existing English words both influenced new learning, regardless of whether this learning took place in explicit teaching or incidental learning contexts. Adults showed long-lasting benefits of lexical similarity, whereas children received greater benefits from offline consolidation that enabled memory for lexically dissimilar items to catch up. These greater offline improvements for children relative to adults were consistent across experiments, supporting the claim that the developing brain may benefit from richer sleep in learning novel information. Standardised measures of vocabulary knowledge strongly predicted overall performance in all five of the experiments that incorporated them, but showed limited relationships with lexical similarity or offline improvements. In a ninth experiment, children's memory was tracked over equivalent periods of wake and sleep, finding that sleep soon after learning had long-term benefits for new word knowledge. However, only when a day's wake intervened between learning and sleep was overnight consolidation predicted by existing vocabulary knowledge. To conclude, prior knowledge supports vocabulary consolidation in some but not all learning contexts, and perhaps influences what is later consolidated rather than the consolidation process itself. Implications for optimising word learning in those with vocabulary weaknesses are discussed.

List of Contents

List of Tables

List of Figures

List of Appendices

List of Accompanying Material

The experimental chapters each have an associated page on the Open Science Framework, containing pre-registrations, experimental materials and stimuli, data, and analyses.

Chapter 3: https://osf.io/35ftn/
Chapter 4: https://osf.io/s2628/
Chapter 5: https://osf.io/stx6q/
Chapter 7: https://osf.io/zqp8r/

Acknowledgements

First and foremost, I wish to thank Dr Lisa Henderson and Professor Gareth Gaskell, for being the best "academic parents" I could have wished for: striking the perfect balance in guiding and challenging me, and with such kindness and generosity. I am grateful for your encouragement to pursue new ideas and opportunities, and for helping me to develop into the kind of scientist I want to be. Thank you for making the last few years so enjoyable, I have truly had the best time working with you.

Thanks also go to Dr Aidan Horner and Dr Emma Hayiou-Thomas for their thought-provoking questions and valuable insights as part of my thesis advisory panel, and to the members of the Sleep, Language, and Memory Lab for providing a safe space to discuss ideas. I especially wish to thank Fay Fletcher for moral support, for countless stats chats, and for helping me to run (literally) from the mid-afternoon stresses.

I am indebted to the wonderful children and schools who took part in the research, and especially to the families who welcomed me into their homes at unsociable hours. I would like to thank the MSc students for their hard work towards a number of the experiments presented: Josie Tulip and Grainne Murphy (Chapter 3 Experiment 2), Claudia Stubington and Nina Romo (Chapter 4 Experiment 3), and Rhiannon Pearce, Caroline Korell, and Charlotte Dean (Chapter 5, Experiment 1). I am especially grateful to Rhiannon, without whom so much comprehension screening would not have been feasible.

These last years would have been very different without Scott by my side. Thank you for putting up with my incoherent rambles, for having dinner waiting after late night testing, and for doing more than your fair share of the chores lately. You have picked me up whenever my confidence crashed and supported me in pursuing every opportunity (even when it meant travelling to another continent). I can't imagine that this PhD would have been half as fun without you to hold my hand.

Last, but not least, I am incredibly grateful to my family. To my mum and my brother, for their incredible strength, support, and understanding. And to my dad, who taught me to be over-ambitious and to question everything, and with whom I would have loved to share this achievement.

Author's declaration

This thesis is a presentation of original work completed solely by the author, under the supervision of Dr Lisa Henderson and Professor Gareth Gaskell. This work has not been previously presented for an award at this or any other university. All sources are acknowledged as references. Some of the data were collected by MSc students under the author's co-supervision (Chapter 3 Experiment 2, Chapter 4 Experiment 3, Chapter 5 Experiment 1).

The research was supported by a 1+3 Studentship from the Economic and Social Research Council.

Publications

This thesis features two chapters that are published in scientific journals as follows:

Chapter 1:

James, E., Gaskell, M.G., Weighall, A., & Henderson, L.M. (2017). Consolidation of vocabulary during sleep: the rich get richer? *Neuroscience & Biobehavioral Reviews*, *77,* 1-13.

Chapter 4:

James, E., Gaskell, M.G., & Henderson, L.M. (2018). Offline consolidation supersedes prior knowledge benefits in children's (but not adults') word learning. *Developmental Science, e12776.*

Conference presentations

A number of the findings have also been presented at conferences as follows:

Chapter 3:

James, E., Gaskell, M.G., & Henderson, L.M. (2017, September). *How can a chicken sleep upside down? The influence of prior knowledge in learning novel concepts.* Poster presented at BPS Developmental Section Annual Conference, Stratford-Upon-Avon, UK.

James, E.L., Henderson, L.M., Gaskell, M.G. (2016, July). *Consolidating new semantic information during word learning: does prior knowledge help?* Poster presented at the 6th International Conference on Memory, Budapest, Hungary.

<u>Chapter 4 & Chapter 5</u>

James, E., Gaskell, M.G., & Henderson, L.M. (2018, June). *The role of prior knowledge in vocabulary consolidation.* Paper presented at Child Language Symposium, Reading, UK.

James, E., Gaskell, M.G., & Henderson, L.M. (2017, November). *Prior knowledge influences in learning and consolidating new words.* Poster presented at the Society for Ninth Annual Meeting of the Society for the Neurobiology of Language, Baltimore, USA.

James, E., Gaskell, M.G., & Henderson, L.M. (2017, September). *The rich get richer? The role of existing knowledge in learning new words.* Paper presented at BPS Developmental Section Annual Conference, Stratford-Upon-Avon, UK.

James, E., Gaskell, M.G., & Henderson, L.M. (2017, July). *Matthew effects in word learning: does prior knowledge aid consolidation?* Paper presented at the Experimental Psychology Society Meeting, Reading, UK.

<u>Chapter 7</u>

James, E., Gaskell, M.G., & Henderson, L.M. (2018, September). *Word learning and consolidation in poor comprehenders.* Poster presented at EPS Research Workshop on Mapping the Future for Poor Comprehender Research, York, UK.

James, E., Gaskell, M.G., & Henderson, L.M. (2018, August). *Word learning and consolidation in poor comprehenders.* Paper presented at 8th International Summer School on Literacy Research, Egmond aan Zee, The Netherlands.

Chapter 1. Introduction

Previously published as:

James, E., Gaskell, M.G., Weighall, A., & Henderson, L.M. (2017). Consolidation of vocabulary during sleep: the rich get richer? *Neuroscience & Biobehavioral Reviews, 77*, 1-13.

Note: Box 1 was created by the senior author, and is not presented for examination.

1.1 Abstract

Sleep plays a role in strengthening new words and integrating them with existing vocabulary knowledge, consistent with neural models of learning in which sleep supports hippocampal transfer to neocortical memory. Such models are based on adult research, yet neural maturation may mean that the mechanisms supporting word learning vary across development. Here, we propose a model in which children may capitalise on larger amounts of slow-wave sleep to support a greater demand on learning and neural reorganisation, whereas adults may benefit from a richer knowledge base to support consolidation. Such an argument is reinforced by the well-reported "Matthew effect", whereby rich vocabulary knowledge is associated with better acquisition of new vocabulary. We present a meta-analysis that supports this association between children's existing vocabulary knowledge and their integration of new words overnight. Whilst multiple mechanisms likely contribute to vocabulary consolidation and neural reorganisation across the lifespan, we propose that contributions of existing knowledge should be rigorously examined in developmental studies. Such research has potential to greatly enhance neural models of learning.

1.2 Introduction

Building a good vocabulary is a crucial task for the developing child, enabling successful communication with others in both spoken and written language. A poor vocabulary places constraints on understanding academic texts, thereby hindering success at school across a broad range of subjects (Biemiller, 2006). Unfortunately, early vocabulary deficits may not be easy to resolve: a long-standing hypothesis in literacy development is the existence of a *Matthew effect* (Stanovich, 1986). The theory holds that the 'rich' get 'richer' in literacy skills; children with better reading and language skills are equipped to further improve these skills, whereas struggling children progress at a slower rate. Although longitudinal studies have provided mixed evidence for Matthew effects in literacy (e.g., Scarborough, Catts, & Kamhi, 2005), some of the most convincing evidence has come from the domain of vocabulary, where the knowledge gap widens throughout the school years (Cain & Oakhill, 2011). Discovering the mechanisms underlying this developmental lag is a key challenge for language acquisition researchers if we are to understand how best to help prevent increasingly widespread problems for children with vocabulary difficulties.

Studies of Matthew effects have largely focused on reading experience and exposure as the underlying mechanism: children with better literacy skills enjoy reading more, will engage in literacy activities in their own time, and have the skills to learn new words from texts when doing so (Cain & Oakhill, 2011; Stanovich, 1993). However, when viewing word learning in the context of neurocognitive theories of memory (Davis & Gaskell, 2009; Wojcik, 2013), it is plausible that other non-environmental processes might also contribute to the effect. Davis and Gaskell (2009) applied the Complementary Learning Systems (CLS) framework (McClelland, McNaughton, & O'Reilly, 1995) to word learning, hypothesising that a new word is initially stored as a distinct episodic trace in the hippocampus, but becomes integrated with existing vocabulary in neocortical long-term memory over time, particularly during sleep. In the broader memory literature, prior knowledge has been shown to enhance the ease with which new information is integrated, and initial evidence suggests that this may also be the case for the overnight integration of newly learned words in childhood (Henderson, Devine, Weighall, & Gaskell, 2015; Horváth, Myers, Foster, & Plunkett, 2015b). Weaker vocabulary may therefore hinder further vocabulary development by constraining neocortical consolidation, as well as via limiting an individual's exposure to language.

If existing knowledge plays such an influential role in subsequent vocabulary learning, then how is it that children (who typically have limited levels of vocabulary knowledge relative to adults) are able to accumulate a mass of vocabulary knowledge at such a rapid rate? Here, we consider that different states of brain maturation elicit different mechanisms to support word learning. Namely, we will review evidence suggesting that whilst word learning in the adult system can benefit from enriched levels of existing knowledge, the sleep architecture of the typically developing system is optimised for sleep-associated memory consolidation. We will begin by summarising systems consolidation models of memory and applications to word learning across development, and review studies that directly compare consolidation processes in children and adults. We consider the proposal that prior knowledge can account for inconsistencies in these data, and present a meta-analysis of our own published data that supports a relationship between existing vocabulary knowledge and the consolidation of newly learned words. Finally, we will propose future directions for addressing the consolidation account of Matthew effects.

1.3 Systems consolidation and the role of sleep

It is well accepted that memory is not a unitary store in which all information is stored and accessed in the way it was initially encoded (McGaugh, 2000). Although the hippocampus and other regions of the medial temporal lobes are known to play crucial roles in memory, studies of patients with hippocampal damage demonstrated that individuals could retain some memory of earlier life experiences (e.g., Scoville & Milner, 1957). From this, it has been concluded that memories may become gradually independent of the hippocampal system over time (Squire & Alvarez, 1995; Squire & Zola-Morgan, 1991) via a process coined *systems consolidation.* Although the nature of the different memory systems and the mechanisms that enable their interaction remain hotly debated in memory research (e.g., Nadel, Winocur, Ryan, & Moscovitch, 2007), there is good evidence to suggest that memory reorganisation continues for the months and even years after first encountering new information (e.g., Takashima et al., 2006).

The time required for systems consolidation necessarily includes multiple opportunities for sleep, and evidence is now converging on the view that neural processes that occur during sleep actually play an active role in memory consolidation. In particular, a substantial body of research has focused on the role slow-wave sleep (SWS) in various aspects of declarative memory consolidation (e.g., Marshall & Born, 2007), suggesting that this stage of sleep enables the reactivation of hippocampal traces to promote slower learning and integration in the neocortex (Diekelmann & Born, 2010; Rasch & Born, 2013). In this section, we describe the key features of SWS and other related aspects of sleep architecture, before reviewing the evidence for its involvement in consolidating linguistic information.

1.3.1 Slow-wave sleep (SWS) and memory

SWS (non-rapid eye movement stages 3 and 4) is characterised by three components of sleep architecture: slow oscillations, spindles, and ripples. Slow oscillations are alternating states of widespread hyperpolarisation and depolarisation at approximately 0.8 Hz. This synchronous firing of neurons throughout the brain is thought to enable communication between hippocampal and neocortical systems (Marshall & Born, 2007; Sirota & Buzsáki, 2005). The hyperpolarised "up" states of slow oscillations feature sleep spindles: short bursts of ~10-15 Hz activity (also seen in Stage 2 sleep). These too have been linked to the communication and replay of information between memory systems, given their tight temporal relationship with cortically-driven slow oscillations and hippocampal activity (Sirota & Buzsáki, 2005). The third component - although one not detected by surface EEG – involves very fast bursts of 80-100 Hz activity originating from the hippocampus. Recent intracranial recordings by Staresina et al. (2015) have demonstrated that these hippocampal ripples are further nested within the troughs of spindles, providing evidence that ripples, spindles, and slow oscillations occur systematically together during SWS. Cross-regional coupling between hippocampal and neocortical measurements demonstrated that the phase of slower oscillations modulated the power of faster oscillations: hippocampal spindles increased in relation to cortically recorded slow oscillations, and hippocampal ripples increased in relation to cortical spindles. The authors concluded that this functional coupling hierarchy might subserve the transfer of information between hippocampal and neocortical memory systems during consolidation.

In support of a causal role for slow oscillations in coordinating memory processing, studies have shown that boosting slow oscillation activity using transcranial direct current stimulation during sleep can improve declarative memory retention (Marshall, Helgadóttir, Mölle, & Born, 2006). However, the relationship between slow oscillations and memory consolidation is likely to be bidirectional: a number of studies have also linked learning demands to neural activity during subsequent sleep (Mölle, Eschenko, Gais, Sara, & Born, 2009). For example, both SWS coherence (Mölle, Marshall, Gais, & Born, 2004) and spindle density (Gais, Mölle, Helms, & Born, 2002) have been shown to be increased in sleep following a word pair learning task compared to a visual processing task of equivalent visual input and duration. Converging evidence therefore suggests that sleep plays a reciprocal and important role in the learning and retention of new information.

1.3.2 A Complementary Learning Systems (CLS) account of word learning

Dual systems approaches to memory and consolidation have been of particular interest to language researchers in considering apparent dissociations in performance in explicit and implicit measures of word learning (e.g., Henderson, Powell, Gaskell, & Norbury, 2014). In particular, the Complementary Learning Systems account has provided a useful framework in which to consider these differences (Davis & Gaskell, 2009). According to the CLS model of memory (McClelland, 2013; McClelland et al., 1995), the two memory systems feature different types of representation: the neocortical memory system consists of overlapping representations that are susceptible to spreading activation from incoming information, whereas the hippocampal system forms sparse memory representations that retain their specificity to the contexts in which they are learned, and are stored largely independently of other representations in memory. However, reinstatement of these hippocampal representations into the neocortex enables this new episodic information to become gradually incorporated into the neocortical system via re-experiencing, rehearsal, or sleep processes. This computational model of memory was proposed to account for the way in which the learning brain can protect existing knowledge from the possible interference of new information, yet remain plastic to new skills and information.

The CLS model thus provides a framework in which to consider how a language system can come to process known words with high speed and efficiency, and function despite substantial variation in the incoming speech signal (Davis & Gaskell, 2009). Much like the distributed representations featured in the CLS account of memory, some computational models of spoken word processing propose that automatic spoken word recognition is accomplished by a distributed system in which phonological and semantic information is stored separately but activated in parallel as speech input unfolds (Gaskell & Marslen-Wilson, 1997). In line with this view, studies suggest that incoming speech sounds initiate phonological competition among related word level representations until the word has been fully specified (Mattys & Clark, 2002). At a semantic level, recent work suggests that activation of a given word also results in the sustained activation of related words in order to facilitate continued language processing and comprehension (e.g., Rodd, Cutrin, Kirsch, Millar, & Davis, 2013). The lexicon is thus characterised as a highly interconnected system that enables the rapid processing of linguistic information for successful communication.

To become an established lexical entry, a new word must become "engaged" with this existing lexicon (Leach & Samuel, 2007) without causing disruption to the system. The CLS framework proposes that an initial encounter with a new word engages numerous cortical regions involved in speech processing that output to form a bound representation in the hippocampus. Initially, retrieving the meaning and phonological form of this new word requires hippocampal mediation, but this new word can become gradually integrated into the main neocortical recognition system over longer periods of time – particularly during sleep (Davis & Gaskell, 2009). A key prediction of this model is therefore that we should not see immediate automatic competition and priming effects for newly learned words, but that these key markers of a fully-fledged lexical item should emerge over longer periods of time (including sleep) as representations become integrated into a distributed system. Although abstracting and generalising linguistic information (as in the context of grammatical features) may be feasible from newly acquired hippocampal traces (Kumaran & McClelland, 2012), the automaticity with which this occurs should be enhanced after representations become integrated within the neocortex. A wealth of evidence now exists to suggest widespread benefits for sleep for the memory and processing of newly acquired language. These have been demonstrated across phonological (Dumay & Gaskell, 2007), semantic (Tham, Lindsay, & Gaskell, 2015) and grammatical domains (Nieuwenhuis, Folia, Forkstam, Jensen, & Petersson, 2013). Less attention has been given to the orthographic aspects of word learning in this area, particularly in developmental research, which limits our discussion of written language here (see Bakker, Takashima, van Hell, Janzen, & McQueen, 2014, for consolidation effects across spoken and written modalities).

Studies of spoken word learning have often examined declarative aspects of learning – i.e., the explicit recall of a word form. For instance, in novel word training studies, adults show an increase in the number of word forms they can successfully recall following a period of sleep, whereas no such increase is seen during an equivalent period of wake (Dumay & Gaskell, 2007). Gais, Lucas, and Born (2006) examined this from the perspective of foreign language learning, training native English adults on German vocabulary translations. Participants recalled more words when they slept shortly after learning compared to when they remained awake. Comparing sleep versus wake periods in behavioural paradigms thus supports that sleep can strengthen word representations for successful retrieval. Tamminen, Payne, Stickgold, Wamsley, and Gaskell (2010) used polysomnography to further specify that the overnight strengthening of word form representations – in this case indicated by improvements in a speeded recognition task - is associated with the amount of time participants spent in SWS.

Researchers have addressed the causal role for sleep in word form consolidation by experimentally manipulating memory reactivations during sleep. Targeted memory reactivation (TMR) paradigms replay previously associated sound cues to participants during SWS, under the assumption that this reactivates the individual memory traces from learning and thereby facilitates consolidation (see Schreiner & Rasch, 2016, for a review). Schreiner and colleagues have demonstrated that recall of newly learned foreign vocabulary translations can be improved by cueing and reactivating newly learned words during SWS (compared to recall of uncued translations; Schreiner & Rasch, 2014) but not during wake (Schreiner & Rasch, 2015). Cues presented during sleep were often followed by slow oscillations, and resulted in increased theta and spindle activity for successful cues only (Schreiner, Göldi, & Rasch, 2015). Consistent with the findings from Staresina et al. (2015) above, the authors suggested that slow oscillations may provide the temporal framework for stabilization processes to occur. Considered together, these behavioural, polysomnography, and TMR studies provide strong evidence for sleep processes in declarative aspects of language learning.

Studies have demonstrated that sleep is also important for the more implicit aspects of phonological word learning; key to the predictions of the CLS model, sleep has been shown to enhance the integration of a novel word form with existing vocabulary knowledge. According to distributed models of the lexicon (Gaskell &

Marslen-Wilson, 1997), a fully lexicalised or “engaged” (Leach & Samuel, 2007) word form can better interact with other entries in vocabulary, competing for activation during word processing. The CLS model predicts that this lexical competition primarily occurs after a period of consolidation, once the word has become integrated within the neocortical memory system. Clear evidence for lexical integration has been provided by studies that teach participants novel competitors (e.g., *cathedruke*) for existing word forms (e.g., *cathedral*) and show that participants become significantly slower to detect a pause inserted into the existing word form (versus detecting pauses inserted into control words for which no new competitor has been taught). Crucially, this slowing of response times does not occur immediately, but emerges after a longer time period if it is inclusive of sleep (Dumay & Gaskell, 2007; Dumay et al., 2004). These findings lend support to the proposal that a period of offline consolidation can enable a word to become integrated with existing vocabulary knowledge and compete during lexical processing (although competition effects between new and existing words have been demonstrated immediately after learning under certain circumstances; see Section 1.5.1 or McMurray, Kapnoula, and Gaskell (2016) for a discussion). Sleep recordings have demonstrated that larger overnight increases in lexical competition effects between novel and existing words are associated with greater levels of spindle activity during sleep (Tamminen et al., 2010). Consistent with the CLS proposal that consolidation strengthens cortical networks, Davis, Di Betta, Macdonald, and Gaskell (2009) used fMRI to demonstrate that words learned a day prior to scanning had become more independent of the hippocampus during retrieval than words learned the same day: words with the opportunity for sleep-associated consolidation processes to occur elicited greater neocortical activity (e.g., in the superior temporal gyrus) and reduced engagement of the hippocampus compared to unconsolidated words. The converging evidence therefore supports that sleep both strengthens new word forms, and enables systems consolidation processes to integrate new words with existing knowledge.

Other research has examined semantic and grammatical aspects of word learning, with support beginning to accumulate for a role of sleep in these domains. One approach has been to examine the emergence of interference effects caused by the automatic activation of semantic information. Clay, Bowers, Davis, and Hanley (2007) used a picture-word interference task in which picture naming slows in the presence of distractor words, particularly for words that are semantically related. This latter meaning-specific effect was not apparent for novel words immediately after learning, but emerged one week later. Similarly, Tham et al. (2015) showed that a semantic incongruency effect for newly learned words emerged only after a period of sleep (e.g., participants took longer to decide that a Malay translation of "fox" was bigger than a Malay translation of "bee" when the latter was presented in larger font). Consistent with sleep effects for phonological forms, the integration of semantic information has also been linked to both SWS duration (Tham et al., 2015) and spindle activity in the intervening night (Tamminen, Lambon Ralph, & Lewis, 2013; Tham et al., 2015).

The CLS model predicts that transfer of newly formed memory traces to the neocortex should facilitate the abstraction of linguistic regularities (e.g., grammatical properties) in a more automatic fashion as the memory traces become represented in a more distributed manner. Speaking to this hypothesis, sleep-associated consolidation has been demonstrated as particularly important when rules are presented only implicitly during the learning phase (Batterink, Oudiette, Reber, & Paller, 2014; Nieuwenhuis et al., 2013; Tamminen, Davis, Merkx, & Rastle, 2012) or when speeded access is required in generalising to new exemplars (Tamminen et al., 2012). For example, using a nap paradigm with a stimulus set in which novel prefixes predicted the animacy of existing referents, Batterink et al. (2014) reported fast learning of the rule made explicit during training, which was not further influenced by sleep. However, adults' ability to extract the hidden regularities in a speeded categorisation task improved after a nap, and was associated with the interaction between SWS and rapid eye-movement sleep. A recent TMR study further supported this role of sleep, demonstrating that auditory cues presented during SWS resulted in improvements in generalising grammatical rules (Batterink & Paller, 2017).

However, evidence for the role of sleep on the abstraction and generalisation of new linguistic information is mixed, and this may be partially due to the nature of the mappings to be learned. While the CLS account of word learning predicts that neocortical integration should facilitate the abstraction of rules, it also predicts that the learning of arbitrary mappings is more dependent on hippocampal mechanisms and thus greater influenced by subsequent sleep than systematic elements. Mirković and Gaskell (2016) tested this hypothesis by using both arbitrary elements (i.e., word-stem to picture mappings, e.g., *scoiff*-ballerina, *jor*-cowboy) and a more systematic element in the mapping between determiners/suffixes and common semantic features (e.g., *tib…esh/eem* and female; *ked…ool/aff* and male). Knowledge of the arbitrary stems improved for participants who took a nap, whereas – in contrast to the previous findings - the systematic grammatical aspects did not. Mirković and Gaskell (2016) suggested that arbitrary items may take priority early in consolidation processes, whereas systematic mappings may be later strengthened. The extent to which the grammatical mappings overlapped with existing mappings was also higher in this study as gender is a relatively salient feature in English language. This overlap may have facilitated neocortical integration, and thereby reduced the potential boost from sleep (see Section 1.5).

The extant evidence therefore suggests that sleep has widespread benefits in adult language learning, with the nature of the material to be learned influencing the extent to which sleep supports learning. Polysomnography recordings highlight that both time spent in SWS (and/or slow oscillation activity) and sleep spindles are associated with the explicit recall of new words and with integrating these words with existing knowledge to enable fast and efficient linguistic processing, especially in the spoken domain. What determines the involvement of sleep spindles and/or SWS duration in processes of language consolidation in the above studies remains an important question that future research should aim to untangle. However, considering recent evidence demonstrating the tight temporal coupling of spindles with other oscillations during SWS (Staresina et al., 2015; see Section 1.3.1), both are considered relevant in the present review, and these sleep-associated consolidation processes are a prime focus in considering language learning across development.

1.3.3 Consolidation of vocabulary earlier in development

An important theoretical question is whether sleep-associated consolidation processes are equally as – or even more – important during development, given the high demand on fast and efficient vocabulary acquisition in childhood. Interestingly, children show a much higher percentage of SWS than adults (Ohayon, Carskadon, Guilleminault, & Vitiello, 2004) and greater slow oscillation activity that reaches a peak at roughly 10-12 years (Feinberg & Campbell, 2010). Thus, it is plausible that sleep could support the enhanced rates of vocabulary learning earlier in development. First, we review the evidence for sleep-associated improvements in children's language learning, and will later consider how their enhanced levels of SWS might affect processes of consolidation across development (Section 1.4).

Thus far behavioural evidence suggests that there are indeed similar benefits of sleep for word learning and integration from infancy to adolescence. A number of studies have suggested similar overnight improvements in novel word form learning to those found in adults. For example, Ashworth, Hill, Karmiloff-Smith, and Dimitriou (2014) taught 6- to 12-year-old children novel names for animals, and found a 14 per cent improvement in recall after a period of sleep compared to wake. A 28 per cent overnight improvement compared for novel word recall was demonstrated in a similar age group by Henderson, Weighall, Brown, and Gaskell (2012), who also demonstrated that sleep enabled lexical competition to occur in a pause detection tasks (e.g., Dumay & Gaskell, 2007; see Box 1 for more details); no such improvements in recall or lexical competition were apparent across a period of wake.

Moving beyond behavioural findings, only one study to date has utilised polysomnography recordings to examine associations between sleep and vocabulary consolidation in school-aged children: Smith et al. (2018) demonstrated that slow-wave activity (the power of EEG activity in the 0.5-4 Hz range; SWA) predicted overnight improvements in cued novel word recall in typically developing children (e.g., "Which novel word began with "bisc"?", answer "biscal"). Sleep spindle activity was also associated with these overnight improvements, but was more strongly predictive of the overnight changes in lexical competition (as measured via the pause detection task). These findings are consistent with those of adult studies (i.e., Tamminen et al., 2010), providing initial evidence of similar underlying mechanisms to sleep-associated consolidation of language across development. Although there is a scarcity of work examining sleep-associated semantic integration in children, benefits in consolidation processes have been shown for training word forms alongside their meaning, and thus for the acquisition of a more complete lexical representation. Henderson, Weighall, and Gaskell (2013c) showed that training on new words with their meaning led to better longer term representations of their word forms compared to form-only training in 5- to 9-year-old children. Furthermore, the benefits of a consolidation period for word learning (for both explicit measures of recall/recognition and implicit measures of lexical competition) are apparent even when novel words are more naturalistically encountered within a story (Henderson et al., 2015; Williams & Horst, 2014), demonstrating that these mechanisms are not restricted to explicit training methods and are likely representative of everyday word learning processes (although see Fernandes, Kolinsky, & Ventura, 2009).

The sleeping brain also appears able to abstract and integrate information from learned words from an early age, relevant for both semantic and grammatical aspects of word learning. For example, Friedrich, Wilhelm, Born, and Friederici (2015) used EEG and event-related potentials as a measure of semantic word learning in infants. Infants that napped after learning new words retained an understanding of the specific word meanings, and also generalised these word meanings to novel exemplars. Infants who stayed awake over this interval showed no such markers of learning. Even at this early age, ability to generalise to new exemplars was correlated with sleep spindles during the nap, suggesting that similar mechanisms may be at play in word learning throughout development (see also Horváth, Liu, and Plunkett (2015a), but Werchan and Gómez (2014) for conflicting findings). Furthermore, sleep has been shown to benefit the abstraction of statistical regularities in strings of nonsense syllables in infants (Gómez, Bootzin, & Nadel, 2006; Hupbach, Gomez, Bootzin, & Nadel, 2009), suggesting that sleep may aid grammatical learning and consolidation from very early in child development.

1.4 Consolidation processes across development

A critical first step in interpreting the mechanisms underlying consolidation during development is to assess whether consolidation takes place via a similar systems transfer of information as in adults. In one of the few studies to test the underlying neural mechanisms in children, Urbain et al. (2016) found that hippocampal activity (measured via magnetoencephalography) during the successful immediate recall of new objects positively correlated with percentage of SWS in a subsequent nap in 8-12-year-olds. After sleep however, successful recall was *negatively* correlated with hippocampal activity, and was instead associated with higher activity in the prefrontal cortex. This study suggests that – as in adults – sleep plays a role in transferring newly acquired memory traces from the hippocampus to neocortical regions, and thus that these mechanisms are of interest across development.

While developmental studies have largely provided findings that are conceptually consistent with adult models of sleep-associated consolidation, more careful developmental comparisons have the potential to inform us about the processes involved (Wilhelm, Prehn-Kristensen, & Born, 2012). Children require more sleep than adults overall, and show a much higher percentage of SWS (e.g., ~40% of total sleep time) relative to adults (e.g., ~20% of total sleep time; Wilhelm et al., 2013) that gradually declines throughout adolescence (Jenni & Carskadon, 2004; Ohayon et al., 2004). These changes in sleep have been tightly linked to processes of cortical maturation (Buchmann et al., 2011) and a greater synaptic strength of neurons involved in the generation of slow-wave oscillations (Kurth et al., 2010). Less is known about developmental changes in sleep spindle activity, but there is evidence that the number and density of spindles also declines from adolescence to adulthood (Nicolas, Petit, Rompre, & Montplaisir, 2001), and some indication of an increasing trend during the first decade of life (Kurth et al., 2010).

Whilst ongoing neural development throughout childhood and adolescence has often been linked to increased sensitivity for learning (Knudsen, 2004), we now turn to consider the potentially important implications of these changes in the context of the CLS account, and review the behavioural studies that make direct developmental comparisons in consolidation processes.

1.4.1 Implications of brain development for consolidation processes

To understand the implications of brain development in consolidation processes, we must acknowledge changes that are happening in the two proposed memory systems across childhood and adolescence. First, we consider the development of the hippocampal memory system. Regions of the hippocampus are not fully matured in infants, but robust effects of sleep-associated consolidation are observed from approximately age 2.5 years (see Gómez & Edgin, 2015, for a review). In preschool children, the correlation between hippocampal volume and expressive language ability increases with age (Lee et al., 2015), suggesting that the maturing hippocampus may be a constraint on word learning in early infancy.

Later in childhood, it is less clear how ongoing subcortical maturation may impact learning and consolidation processes. Hippocampal mechanisms are thought to be in place by the time children reach school age (e.g., Gilmore et al., 2012; Seress, 2001), and longitudinal studies have not been able to pinpoint significant age-related changes in overall hippocampal volume during subsequent years (Giedd et al., 1996; Østby, Tamnes, Fjell, & Walhovd, 2011; Østby et al., 2009). However, there is some evidence of continued development throughout middle childhood and adolescence (Ghetti & Bunge, 2012), predominantly in a shift in relative mass towards posterior hippocampal regions (Gogtay et al., 2006). This corresponds to functional shifts apparent in both encoding (Ghetti, DeMaster, Yonelinas, & Bunge, 2010) and episodic retrieval tasks (DeMaster & Ghetti, 2013), during which adolescents and adults come to recruit more anterior regions of the hippocampus than children. Interestingly, a recent study suggests that the refinement of this anterior region is correlated with an increased ability to draw inferences across learning episodes (Schlichting, Guarino, Schapiro, Turk-Browne, & Preston, 2016). This ongoing development may therefore have important implications for learning strategies, and thus for teaching practices with different age groups.

Structural and functional differences in the hippocampus between children and adults *could* account for children's need for more sleep throughout development. For example, an immature hippocampus may be able to retain less information before requiring sleep, or may store weaker representations that require strengthening and linking to existing knowledge via sleep-associated processes. However, the implications of hippocampal changes for longer-term consolidation and sleep are supported by only tentative evidence. Østby et al. (2011) related the structural brain maturation of 8-19-year-olds to their immediate and delayed performance in a visuospatial memory task, and showed that hippocampal volume was predictive of memory performance one week later (but not of immediate performance). Furthermore, measures of structural hippocampal volume in children have shown positive correlations with weekday sleep duration (Taki et al., 2012), although the causal direction is unclear. These studies enable us to speculate that differences in hippocampal development could be impacting the relationship between learning and sleep in childhood. Nevertheless, there is a clear need for direct assessments between sleep, memory and hippocampal function in this age group, and it is important to acknowledge that learning itself will impact neural development (Blakemore & Bunge, 2012).

There is much clearer evidence for the protracted development of cortical regions throughout childhood and their associations with sleep. It has often been noted that the decrease in SWS during adolescence parallels continued changes in cortical grey matter at this age (e.g., Feinberg & Campbell, 2010). Buchmann et al. (2011) used structural magnetic resonance imaging (MRI) and overnight polysomnography measures to confirm a positive correlation between SWA and cortical grey matter throughout adolescence, with both factors decreasing with age. Regional analyses strengthened this link further: once controlling for overall decreases in SWA, the strongest decrease in SWA was observed in parietal regions that were undergoing the strongest decrease in grey matter volume, whereas relative increases in SWA were shown in regions of the prefrontal cortex still undergoing grey matter development. Slow-wave activity thus appears to be tightly linked to the developing brain, and could play a supporting role in the cortical reorganisation that occurs during this period of enhanced learning.

One study has spoken to the developing brain's capacity for sleep-associated neural reorganisation by combining neuroscientific measures with behavioural tasks. Wilhelm et al. (2014) found larger region-specific boosts in children's SWA after participants completed a visuomotor adaptation task, compared to adolescents and adults. Consistent with the findings above, baseline levels of SWA positively correlated with parietal grey matter volume. More interestingly, grey matter volume was also associated with the local increase in SWA following the adaption task, suggesting these developmental changes in SWA are linked to experience-dependent plasticity particularly in the maturing brain. Unfortunately, there was no follow-up task in this study to assess the behavioural implications of these enhanced sleep processes. Nevertheless, the study provides an insight into how sleep could play a key role in shaping cortical maturation processes across development.

1.4.2 Direct comparisons of consolidation effects between childhood and adulthood

The greater amounts of SWS seen in childhood and its connections to plasticity raise the possibility of superior consolidation processes: if SWS facilitates reactivation of hippocampal traces for stabilisation in the neocortex, then this should enable faster and/or larger consolidation effects in children. However, few studies have made direct comparisons between children and adults, particularly within the contexts of explicit and/or linguistic memory tasks relevant to word learning, and extant findings are mixed. Making such comparisons brings challenges to interpretation, as differences in the amount of information encoded could drive apparent differences in subsequent consolidation processes. From this perspective, it would be important to match groups in their baseline performance at encoding. However, matching the amount of information encoded could also lead to disparities in task difficulty for the groups of participants, suggesting that multi-faceted approaches will be important to address these questions.

Some of the most convincing evidence for enhanced sleep-associated processes in children has come from a study by Wilhelm et al. (2013), who looked at the extraction of explicit knowledge from an implicit motor sequence learned prior to sleep. Children aged 8-11 years and adults were given equal amounts of training on a motor task that required them to respond as quickly as possible to a sequence of light-up buttons on a response box, forming an implicitly learned motor sequence. After sleep, children were significantly better at explicitly recalling the next light buttons in the learned sequence, suggesting an enhanced ability to extract explicit knowledge from an implicit task, and performance was tightly correlated with levels of SWA on the night between training and test in both groups. In fact, children were so much better at this task than adults that the study was repeated in children with a more complex sequence, in order to better analyse the relationship with SWA in this population. The findings supported the proposal that, at least under certain conditions, greater amounts of SWA in children can support the high demands on learning that is characteristic of this stage of development.

Returning to the consideration of consolidation effects in language learning, a recent study by Weighall, Henderson, Barr, Cairney, and Gaskell (2016) also demonstrated a larger overnight benefit for children compared to adults in the explicit recall of newly learned words. In this study, 7-to-9-year-old children and adults both learned a total of 48 novel word-object pairings. Crucially, half of these pairings had been trained the day before – allowing for a night of sleep before testing – whereas the other half were learned on the same day as the test session. When given the task of completing the novel word forms from their stems (e.g., “which novel word began with *dol*?”), children showed a large advantage (36%) for words that had the opportunity for consolidation, whereas for adults this figure was significantly smaller (24%). In addition, a visual world eye-tracking paradigm was used to examine fixations to novel competitor objects (e.g., *dolpheg*) when asked to click on one of four pictures arranged in quadrants (e.g. “click on the *dolphin”*). Whilst both children and adults showed increased fixations to novel competitor objects (e.g., *dolpheg*), only children showed an enhanced overnight benefit of sleep (i.e., significantly greater competitor effects for consolidated than unconsolidated items). Although sleep recordings were not taken from children in this study, the behavioural evidence again supports that the characteristics of sleep during childhood could support rapid learning (and sleep spindles were clearly implicated for adults).

However, the differences in overnight sleep benefits for adults and children are not always evident: several studies have demonstrated comparable (Henderson, Weighall, Brown, & Gaskell, 2013b; Wilhelm, Diekelmann, & Born, 2008) or occasionally even larger (Henderson et al., 2015) overnight boosts in novel word recall performance for adults compared to children. For example, Wilhelm et al. (2008) had 6- to 8-year-old children and adults learn both verbal (semantically associated word pairs) and nonverbal (location pair) declarative stimuli. Sleep recordings showed that children had over double the amount of SWS than adults in the night between learning and test, yet children showed a comparable behavioural benefit to adult participants. These mixed findings highlight that the mechanisms and influences of learning may not be the same for adults and children, and point towards the need for more direct comparisons between adults and children to systematically address this question.

1.5 A role for existing knowledge

One proposal put forward by Wilhelm and colleagues (Groch et al., 2016; Wilhelm et al., 2008; Wilhelm et al., 2012) is that adults have greater amounts of existing knowledge to support the fast consolidation of new information. Thus, children benefit from greater amounts of SWS, but adults can often compensate for their decreased amounts of SWS because of the higher levels of existing knowledge available to support integration. This proposal is in line with theories that suggest information is more readily integrated when consistent with existing schemata (Tse et al., 2007). Indeed, the most recent account of the CLS model emphasises that neocortical learning is not slower *per se*, but prior knowledge-dependent: new information that is consistent with existing knowledge produces little interference, and thus does not require the same extent of reactivation for cortical learning (McClelland, 2013).

Lewis and Durrant (2011) considered sleep-dependent mechanisms of integration in their *information overlap to abstract* (iOtA) model. They proposed that a new memory representation can activate relevant parts of schematic knowledge during encoding. During subsequent sleep, hippocampal reactivation of the representation amplifies the response of these overlapping neocortical neurons, thereby facilitating the integration of the new information with schematic knowledge via Hebbian learning principles. The greater the overlap between new and existing information, the more efficiently the integration can proceed as fewer new neural connections are required. From a developmental perspective, this would suggest that consolidation can proceed more rapidly in adults due to superior levels of existing knowledge, with reduced demands on processes during sleep, providing that the new information in question can capitalise on this.

The prior knowledge account could partially explain the mixed findings in studies that have compared the consolidation processes of adults and children. For example, in the study by Wilhelm et al. (2008), adults' greater amount of prior knowledge available to support the consolidation of word pairs could account for their similar overnight benefits to children, who instead showed greater amounts of SWS. Despite attempts to make their stimuli of equivalent difficulty across the two age groups, the extent of related or supporting prior knowledge that may be activated during learning is practically impossible to control. Further, the protracted development of anterior hippocampal regions across middle childhood may mean that the activation and integration of any prior knowledge is less consistent in this age group (Schlichting et al., 2016; see Section 1.4.1). Importantly, when existing semantic knowledge could not be capitalised upon in a motor sequence task, children showed enhanced sleep-associated benefits in comparison to adults (Wilhelm et al., 2013).

Such an explanation is supported by recent data from van Kesteren, Rijpkema, Ruiter, and Fernández (2013), which highlighted that individual items are particularly susceptible to the influence of prior knowledge on consolidation processes, compared to associations between them. Participants learned visual motifs paired with related or unrelated tactile fabrics, and were tested for both visual item recognition and the paired associates at different time intervals. Recognition of the items themselves was boosted for groups that had a 20- or 48-hour delay before testing to allow for consolidation processes to take place, whereas prior knowledge of associations (congruent visuo-tactile pairings) could benefit learning immediately. As a result, the consolidation benefit for schematic knowledge on associations was not as prominent. This earlier influence of schematic knowledge can also help to account for adults' generally higher level of performance but often smaller overnight consolidation effects relative to children: whilst adults experience greater benefit from existing knowledge during learning and/or consolidation, children benefit from enhanced SWS that facilitates overnight consolidation processes.

1.5.1 Existing vocabulary knowledge in word learning

In learning a new spoken word, we can consider the benefit of existing knowledge on both phonological and semantic aspects. If a word shares a similar phonological structure to existing words, then it can benefit from existing phonemic contingencies. Likewise, if the new word relates to known semantic concepts, then it can capitalise on knowledge about those concepts and thus require fewer new neural connections to be made. In comparing adults with children in language learning studies, we see a similar pattern to that described above: when adults could link novel words to prior knowledge in a story learning context, they showed better overnight improvements in cued recall of the words (Henderson et al., 2015), whereas children show the biggest improvements when words are linked to entirely novel objects (Weighall et al., 2016).

Within studies of developmental language acquisition, an influence of existing vocabulary knowledge predicts that children with superior vocabulary should demonstrate more efficient consolidation of new words. In this instance, a child can benefit from both enhanced SWS and good levels of prior knowledge. Henderson et al. (2015) explored this possibility further in their study of word learning (see also Horváth et al., 2015b, for similar findings in infants). In children aged 7-10 years, expressive vocabulary scores were positively correlated with overnight changes in both cued recall of newly learned words and lexical competition effects (the extent to which they became integrated with and influenced the processing of existing lexical neighbours). Also consistent with a delayed benefit of existing knowledge, Wilkinson and Houston-Price (2013) demonstrated that existing vocabulary knowledge accounted for over 20 per cent of variance in novel word memory 24 hours after training and after a further two weeks. However, a lack of an immediate test means it is not possible to pinpoint initial learning and consolidation processes in this latter study.

If existing vocabulary knowledge facilitates the processes of learning a new word, then this account is highly relevant for Matthew effects in word learning. In light of this proposal, we have conducted a meta-analysis of our existing novel word learning data from five previous studies that analysed the predictive relationship between existing vocabulary knowledge and overnight changes in phonological integration (Box 1). Standardised vocabulary scores were a unique predictor of lexical competition effects the next day (accounting for 10% of variance) after controlling for age, explicit retrieval of the word forms, and reaction times to control words. This relationship held regardless of whether the study included semantic elements of word learning; although the association was numerically stronger (albeit not significantly) when words had been trained in the context of meaning. Although we cannot conclude a causal direction for this relationship, and there are likely to be additional factors at play, the findings are consistent with a facilitatory effect of prior vocabulary knowledge in lexical consolidation, and we propose a number of studies to address this hypothesis in Section 1.6.

Furthermore, new words have also been demonstrated to integrate more quickly with existing vocabulary knowledge when both the novel word and existing neighbours are co-activated during learning. As previously mentioned, the neocortical system is proposed to be slower or *prior knowledge-dependent*, such that substantial links between existing knowledge and new information can lead to more rapid consolidation, without the need for sleep. For example, in contrast to studies that use the pause detection paradigm, new words tend to show immediate competition effects if they are learned using a "referent selection" procedure (Coutanche & Thompson-Schill, 2014). In these studies, participants identify the referent of a novel word by eliminating the known objects present, such that accessing prior knowledge during word learning appears to fast-track the consolidation of novel words. Further, words learned via referent selection do not further benefit from sleep processes (Himmer, Müller, Gais, & Schönauer, 2017). Whether children could also experience this immediate benefit in word learning via this procedure remains an important open question, with potential practical implications for vocabulary teaching methods.

1.5.2 Experimental evidence for the role of existing vocabulary knowledge

The consolidation literature points to an additional means by which existing vocabulary can facilitate the acquisition of new words, and thus could partially account for Matthew effects found in development alongside enhanced exposure to novel vocabulary (Cain & Oakhill, 2011). It remains highly likely that the environmental factors of experience and exposure play key roles in helping the 'rich' get 'richer', but the contribution of prior knowledge to lexical consolidation suggests that the underlying neural mechanisms might also facilitate this effect.

Box 1. Vocabulary ability as a unique predictor of overnight integration effects

Studies with school-aged children have shown that sleep works to integrate new phonological forms with existing lexical knowledge (Henderson, Devine, Weighall, & Gaskell, 2015; Henderson, Powell, Gaskell, & Norbury, 2014; Henderson, Weighall, Brown, & Gaskell, 2013a; Henderson, Weighall, & Gaskell, 2013b; Henderson, Weighall, Brown, & Gaskell, 2012). Such conclusions are based on the assumption that once a novel word has been integrated into long-term language networks, it should compete for recognition with known words. Studies have captured this 'lexical competition' effect with the pause detection task (Mattys & Clark, 2002). In this task, participants make speeded judgements on whether a 200ms pause is present or absent in a set of basewords (e.g., "dolph_in" or "dolphin", respectively) for which a new competitor has been taught (e.g., "dolpheg") and a set of control words for which no new competitors have been taught. Lexical competition (i.e., significantly slower responses to basewords than control words) seems to emerge after a consolidation period that includes sleep. This is the case when children are taught only the phonological forms of words via explicit instruction (Henderson et al., 2014; Henderson et al., 2013a; Henderson et al., 2012), when they are taught real words with meaning (Henderson et al., 2013b), and when they learn novel words via more implicit encounters in stories (Henderson et al., 2015). The latter study reported that children with better existing vocabulary knowledge show larger overnight gains in lexical competition. This provides some evidence that existing vocabulary knowledge might work to bolster the consolidation of new language, that superior consolidation processes facilitate the growth of vocabulary, or both. However, it remains possible that the correlation between existing vocabulary and overnight consolidation of new vocabulary occurred as a consequence of teaching novel words in stories. Namely, children with richer vocabulary knowledge may be better at comprehending the story, leaving more resources available for novel word learning and/or consolidation.

To address this issue, we combined data from five of our previous studies, three of which trained novel phonological forms (e.g., dolpheg) via phonological training tasks (e.g., repetition, initial and final phoneme segmentation, phoneme monitoring) without including any reference to novel word meaning (Henderson et al., 2012; Henderson et al., 2014; Henderson et al., 2013a), and two of which taught novel words with meanings (i.e., real words with definitions, Henderson et al., 2013b; novel words in spoken stories, Henderson et al., 2015). A total of 158 children participated in these studies: 90 in the 'no meaning' studies (mean age 9.61 years, SD=1.69, range 7-13 years) and 68 in the 'meaning' studies (mean age 8.38 years, SD=1.18, 6-10 years). It should be noted that in Henderson et al (2014) only the typically developing children (and not the children with diagnoses of autism spectrum disorder) were included in the present analyses, but all other child participants were included.

Given that the magnitude of overnight change can depend on baseline performance, hierarchical regression analyses were conducted predicting Day 2 lexical competition while controlling for Day 1 lexical competition, Day 1 pause detection RT for the control condition, and Day 1 cued recall performance, with standardised vocabulary scores as the key predictor. Vocabulary was measured via the Peabody Picture Vocabulary Test in all studies except Henderson et al. (2015), which used the Vocabulary subtest from the Weschsler Abbreviated Test of Intelligence. As shown in Table 1, vocabulary knowledge accounts for significant variance in lexical competition effects on the day after training when pooling data across all studies (Model 1), and when word learning occurs with meaning (Model 2) or without meaning (Model 3). The unstandardized regression coefficients for Models 2 and 3 did not significantly differ (Fischer's r-z transformed z score = .50), confirming that vocabulary was a significant predictor of lexical competition on Day 2, regardless of whether words were taught in the context of their meanings or not. Partial correlations, controlling for age and Day 1 lexical competition effects further demonstrate that children with better existing vocabulary knowledge showed larger overnight gains in lexical competition from Day 1 to Day 2 ($r(154)=.27$, $p<.001$). Although vocabulary appeared to account for twice as much variance in studies that provided meanings versus no meanings, these correlations did not significantly differ in magnitude ('no meaning' studies: $r(86)=.22$, $p<.05$; 'meaning' studies: $r(64)=.35$, $p<.01$) (Fischer's r-z transformed z score = -.86) (see Figure 1).

Table 1. Hierarchical regression analyses predicting Day 2 lexical competition (Competitor pause detection RT - Control pause detection RT) from standardised vocabulary scores.

Model	Step	Predictors	R^2	ΔR^2	F change	β
1 - All studies	1	Lexical competition Day 1	.02	.02	1.07	.04
(n = 158)		Control RT Day 1				.02
		Cued Recall Day 1				.16
	2	Vocabulary	.11	.09	14.46***	.31***
2 - Meaning studies	1	Lexical competition Day 1	.08	.08	1.79	.03
(n = 68)		Control RT Day 1				-.19
		Cued Recall Day 1				.24
	2	Vocabulary	.19	.12	9.14**	.36**
No Meaning studies	1	Lexical competition Day 1	0.2	.02	0.48	.05
(n = 90)		Control RT Day 1				.10
		Cued Recall Day 1				.18
	2	Vocabulary	.07	.05	4.97*	.24*

Note. Analyses control for Day 1 lexical competition, Day 1 Control pause detection RT, and Day 1 cued recall performance. Results are presented separately for a combined analysis across all studies, and for the 'meaning' and 'no meaning' studies. *p < .05, **p < .01, *** p < .001.

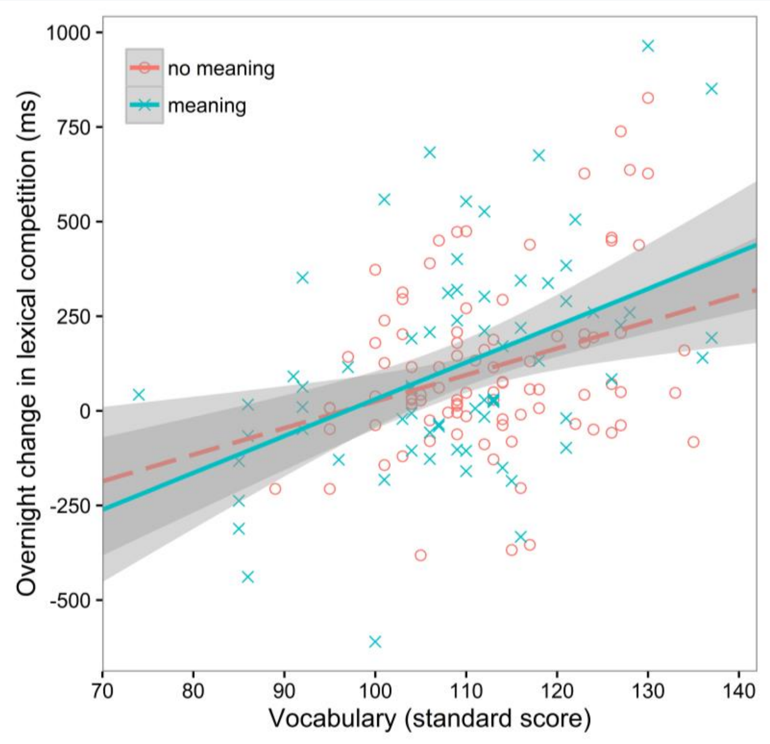


Figure 1. Scatterplot showing the positive correlation between overnight changes in lexical competition from Day 1 to Day 2 and standardised vocabulary scores, for the 'meaning' and 'no meaning' studies separately. Overnight changes in lexical competition = (Competitor RT – Control RT Day 2) – (Competitor RT – Control RT Day 1). Grey bands represent 95% confidence intervals.

Although this view of consolidation is a novel proposal for explaining vocabulary development of school-aged children, the facilitatory effect of existing knowledge on word learning gains support from areas of infant language acquisition. Computational analyses of early acquired semantic networks have led to the proposal of a preferential attachment theory, whereby highly connected words or concepts are more likely to acquire new connections (Hills, Maouene, Maouene, Sheya, & Smith, 2009; Steyvers & Tenenbaum, 2005). Borovsky, Ellis, Evans, and Elman (2015) built on this idea to propose a lexical leverage hypothesis in infant word learning: a given word should be more easily learned if it is entering a densely occupied semantic space, as a child can use their existing knowledge to make inferences about the new concept rather than build a new representation from scratch. Consistent with this hypothesis, they showed that an infant's existing knowledge of a semantic category (e.g., animals, clothes, fruit) was predictive of which words were learned more easily when taught new words from the same categories.

Perry, Axelsson, and Horst (2015) further demonstrated that the structure of an infant's existing vocabulary knowledge guides them towards *what* they learn about a new object. In this study, toddlers remembered more features about new objects if their vocabulary included more shape-based nouns, suggesting that their previous experiences helped to guide them towards what to learn about new concepts in order to successfully distinguish between them. These studies support the proposal that prior knowledge can indeed influence word learning in young children, and that this is a plausible factor in word learning throughout subsequent development.

1.6 The rich get richer: future directions

The evidence points towards an additional means by which children with good vocabulary knowledge could advance at a faster rate than those with poorer vocabulary. The consolidation account provides a testable explanation as to how Matthew effects might arise, suggesting that such effects could be a product of internal learning mechanisms as well as the environmental factors typically considered in previous research. Our meta-analysis supports a link between existing vocabulary knowledge and word learning ability, but it has yet to be tested experimentally in school-aged children to establish a causal influence. Here we propose some future directions for exploring this hypothesis further, and argue that consolidation effects should be considered as a factor in any complete account of vocabulary acquisition.

1.6.1 Developmental comparisons

First, there is a clear need for more direct and careful comparisons of sleep-associated consolidation effects across development. A key proposal made here and previously by others (e.g., Wilhelm et al., 2012) is that children are equipped for faster and superior consolidation effects due to enhanced levels of SWS and accompanying capacity for cortical reorganisation. Later in development, adults are advantaged by greater amounts of pre-existing knowledge that can in some instances bolster the integration of new information. Varied approaches are required to thoroughly test whether these different mechanisms are responsible for similar behavioural findings. In word learning, we might expect adults to always be able to gain from their superior language knowledge where words share phonological, orthographic or semantic neighbours, whereas overnight consolidation benefits would be stronger for children where new words and concepts share few similarities with existing knowledge.

As highlighted earlier, it will be important to draw behavioural comparisons when similar quantities of information are presented for learning (leaving variable prior knowledge contributions across participants), as well as when the to-be-learned information is manipulated to ensure equivalent levels of difficulty across younger and older participants. A comprehensive approach in language learning would thus be to compare consolidation effects in developmental groups when the groups are trained to criterion (e.g., successful performance on a given number of words) to when groups receive the same amount of exposures to the new words. An alternative approach would also be to include pre-training on novel material to create equivalent levels of prior knowledge across groups and observe subsequent consolidation effects of experimental items trained into them. Together, these comparisons would help to better specify the relationship between the demands of learning and subsequent sleep parameters, and their combined influence on overnight consolidation. Worthy comparisons could also be made regarding *when* existing knowledge plays a role on different aspects of language learning. Given evidence to suggest that prior knowledge contributes to a larger overnight consolidation effect for individual items compared to associations (van Kesteren et al., 2013), one might suggest that developmental differences in consolidation effects will be larger in word form recall than in associating new words with meanings. Furthermore, adults may show larger differences in overnight consolidation effects between items and associations than children, given the weaker influence of existing knowledge in the latter age group.

The engagement of these different processes over the course of learning and consolidation could be further elucidated by using functional magnetic resonance imaging (fMRI) to compare the engaged neural mechanisms between adults and children.

Rather than manipulating the influence of prior knowledge in these studies, an alternative approach could focus on comparing the performance of adults and children following manipulation during sleep. A number of methods can be used to influence and enhance SWS architecture in adults, with consequences for memory performance: transcranial direct current stimulation has been used to successfully boost slow oscillation activity (Marshall et al., 2006), and auditory stimulation delivered in phase with slow oscillation up-states enhances subsequent slow oscillation activity and phase-locked spindle activity (Ngo, Martinetz, Born, & Mölle, 2013; Ngo et al., 2015). Thus, in word-learning designs that have minimised the influence of prior knowledge, it may be possible to bring the superior sleep-associated memory benefits of children to adults by enhancing their sleep architecture in this way. This would provide further support of the two contributing mechanisms to consolidation across development.

1.6.2 Manipulating the connections of new words to existing vocabulary knowledge

We can look for clearer evidence regarding the impact of existing knowledge on new word learning by manipulating the extent to which new information can capitalise on prior knowledge. If our findings of a relationship between existing vocabulary knowledge and overnight gains in word learning and lexical competition are due to the ease at which the new words can be integrated, then children with better vocabulary should show an advantage when learning words that are richly linked to their body of existing knowledge, compared to words that are less well linked to existing knowledge. However, if the source of individual differences lies elsewhere – for example, as a consequence of more general differences in the learning mechanisms or other variables that were not included in the present analyses (e.g., IQ, differences in sleep architecture) – then children with superior vocabulary should continue to show better gains regardless of the words they are learning.

We propose three ways by which connections with existing knowledge could be manipulated in word learning studies. First, as in the infant language studies described above, vocabulary across different semantic categories could be used categorise novel items as having weak or strong links to existing knowledge on an individual basis (Borovsky et al., 2015). For example, a child whose hobbies are primarily musical might show greater overnight benefits for instrument names compared to sport terminology, whereas a child who spends their weekends playing football might show the opposite effect.

Second – and perhaps most feasibly in school-aged children – novel items can be created that link to low or high density phonological, orthographic and/or semantic neighbourhoods. This manipulation makes broader predictions about the ease at which certain items should be integrated, and has already been used in one study of new semantic knowledge in adults (Tamminen et al., 2013). If sensitive enough to changing neighbourhoods across development, this may interact further with individual differences in vocabulary knowledge and provide an even stronger assessment of existing knowledge on word learning ability (see Storkel & Hoover, 2011, for a similar approach to immediate word learning in infants).

Third, more carefully controlled studies can manipulate the existing knowledge itself, such that later trained novel items can feature strong or weak links to existing knowledge. Although time intensive, similar approaches have been highly successful in unpicking the ease of assimilation effects in other areas of memory research (Hennies, Lambon Ralph, Kempkes, Cousins, & Lewis, 2016; Sommer, 2016). For example, Hennies et al. (2016) first taught participants a new schema over the course of two weeks. Participants were then presented with a series of facts to learn that were either consistent with their new knowledge or completely unrelated. Spindle density during the following night's sleep was predictive of a memory benefit for the related facts only, and predicted a decreased involvement of the hippocampus as shown by functional MRI scans the following day. A similar approach could therefore be taken to word learning, by creating sparse and high density phonological, orthographic and/or semantic neighbourhoods prior to training experimental items for analysis of consolidation effects.

1.6.3 Studies of atypical development

Another potentially informative approach will be to explore the learning and consolidation of new words in children with developmental disorders, especially considering the prevalence of sleep difficulties within these populations (e.g., Malow et al., 2006; Sadeh, Pergamin, & Bar-Haim, 2006). Sleep-associated consolidation differences have already been a topic of interest in children with developmental disorders, including children with autism (Henderson et al., 2014; Maski et al., 2015), ADHD (Prehn-Kristensen et al., 2011), and Williams Syndrome (Dimitriou, Karmiloff-Smith, Ashworth, & Hill, 2013). Here, it is important to consider the role of prior knowledge as well as sleep difficulties in order to better understand how to remediate and support learning in these groups. Again, multiple comparisons that match relative difficulty and the amount of knowledge learned across groups will be important to consider.

One group of particular interest in studying the influence of prior knowledge on vocabulary acquisition processes will be poor comprehenders: children who struggle to understand and make inferences from text or discourse, despite otherwise adequate phonological skills that support accurate word identification (Nation & Snowling, 1998; Stothard & Hulme, 1995). Such specific comprehension problems are apparent in approximately 5-10 per cent of school-aged children and constitute the largest proportion of reading deficits that emerge in later schooling (Catts, Compton, Tomblin, & Bridges, 2012). Research has shown that these children exhibit vocabulary deficits that are largely linked to the semantic component of word learning (Nation, Snowling, & Clarke, 2007). These vocabulary deficits clearly worsen over time (Cain & Oakhill, 2011), highlighting the importance of understanding word learning difficulties in these children at an early age.

Studies of word learning in poor comprehenders have demonstrated that these children show equivalent learning to typically developing children initially, but do not retain their new lexical knowledge well over time (Nation et al., 2007; Ricketts, Bishop, & Nation, 2008). Considered differently, poor comprehenders have the skills to learn new words – even when it places demands on their comprehension skills to infer their meanings from text (Ricketts et al., 2008) – but their impairment arises at the consolidation stage of learning. An fMRI study by Cutting et al. (2013) further reported increased hippocampal and parahippocampal involvement in word reading in children with specific reading comprehension difficulties, suggesting anomalies in connections between basic language-related areas (e.g., BA 44) and declarative memory systems. Given the role of hippocampal and parahippocampal regions in the initial encoding of episodic and semantic memories (Moscovitch et al., 2005), the authors speculated that poor comprehenders may have difficulty with cortical consolidation, or rely on hippocampal connections as a compensatory mechanism. A prime question here will therefore be whether poor comprehenders can be characterised as having a problem localised to the specific processes of consolidation, or whether these deficits are accounted for by their pre-existing deficits in vocabulary knowledge that provide weakened support for consolidation and integration of new words into long-term memory.

1.7 Conclusions

Sleep plays an important role in the stabilisation of newly learned memories and their integration with existing knowledge. Numerous studies have demonstrated this sleep-associated benefit in word learning, and have accumulated support for the specific roles of SWA and sleep spindles. We have reviewed evidence that suggests enhanced levels of SWS during childhood may support the greater amounts of learning experience at this time, enabling neural reorganisation as cortical networks continue to develop into adolescence. Consistent with Wilhelm and colleagues' proposal, we have suggested that links to prior knowledge can also facilitate consolidation during word learning, and the reviewed evidence of adults and children supports this suggestion. Furthermore, a meta-analysis of our previously published data has shown that individual differences in vocabulary knowledge are predictive of overnight consolidation effects during word learning. This provides a novel and robust demonstration of the Matthew Effect within the context of lexical consolidation.

The influence of existing vocabulary in supporting word learning has important implications for studying the trajectory of vocabulary development, and particularly in considering the means by which the 'rich get richer'. The reviewed studies suggest that neurological mechanisms could contribute to such Matthew effects in vocabulary, alongside differences in environment and exposure. Accounting for both types of influence is important in developing a complete model of word learning, and understanding how best to prevent children with poor vocabulary falling further behind.

However, there is a clear need for more direct and experimental approaches to this question, and we have provided a number of suggestions for future research in both typical and atypically developing populations. It is hoped that these will help to further our understanding of the mechanisms at play during word learning, and unpick the directional relationships between new information, existing knowledge, sleep processes and neural reorganisation.

Chapter 2. Approaches

In Section 1.6, we outlined a number of suggested approaches for better understanding the role of prior linguistic knowledge in learning and consolidating new vocabulary. This chapter briefly summarises the experimental and statistical approaches taken forward in this thesis, aimed at addressing two over-arching theoretical questions:

1) How and when does prior knowledge influence the learning and consolidation of new vocabulary?
2) Do the influences of prior knowledge and offline consolidation in new vocabulary learning differ from childhood to adulthood?

2.1 Experimental approaches

2.1.1 Manipulating "local" prior knowledge

As described in Section 1.6.2, one way to test for causal contributions of prior knowledge to vocabulary learning and consolidation is to train new words that vary in their similarity to existing language. This enables us to compare memory for new vocabulary with greater or fewer connections to existing knowledge, and assess how these influences change with offline consolidation. We predicted that benefits of prior knowledge would increase with consolidation, following increased opportunities for new words to engage with an individual's existing vocabulary.

These manipulations are central to three sets of experiments presented in this thesis. We refer to these kinds of manipulations as "local" prior knowledge, to distinguish from analyses relating to individual differences in prior vocabulary knowledge (Section 2.1.2). In Chapter 3, we trained pseudowords with associated concepts that varied in their semantic neighbourhood density. The reasons for starting with semantic connections were two-fold. First, previous experiments showing relations between existing vocabulary knowledge and longer-term consolidation have used *expressive* vocabulary measures (e.g., Henderson et al., 2015). These tasks arguably probe depth and richness of an individual's lexical-semantic knowledge, suggesting that it may be connections to semantic knowledge that are important for supporting vocabulary consolidation. Second, there was evidence to suggest that presence (versus absence) of semantic information during training can enhance consolidation of word-forms (Henderson et al., 2013c). As such, the first experiments sought to test the hypothesis that training new concepts into dense semantic networks might enable new vocabulary to benefit from semantic richness during learning and consolidation.

Alternatively, it may be that it is similarity to known word-*forms* that might underlie the correlations observed in previous studies: the meta-analysis included in Chapter 1 was conducted on data from a pause detection task, proposed to measure the integration of a novel word-form (e.g., *cathedruke*) with its word-form neighbours (*cathedral*). In Chapter 4, we trained pseudowords with and without word-form neighbours (phonological and orthographic), and further assessed contributions of a single neighbour to new vocabulary learning. Both Chapter 3 and Chapter 4 manipulated these local prior knowledge connections in the context of explicit vocabulary instruction. However, in Chapter 5, we assessed whether access to word-form neighbours during learning differs when pseudowords are encountered incidentally in stories. This change in presentation was intended to inform the extent to which local prior knowledge benefits result from strategic engagement of known words at encoding. In all cases, we examined memory performance on the same day, the next day, and one week after learning, to assess prior knowledge benefits before and after opportunities for consolidation.

2.1.2 Differences in "global" prior knowledge

Importantly, the manipulations described above were designed to further our understanding of the relationship between existing vocabulary knowledge and overnight consolidation in new word memory. We refer to this existing vocabulary ability as "global" prior knowledge, measuring the prior knowledge any individual brings to the task of word learning. We take two approaches to examining global prior knowledge in this thesis: assessing differences between children and adults, and measuring an individual's vocabulary knowledge using standardised vocabulary assessments.

Developmental comparisons

Comparing adults and children in their acquisition of new vocabulary allows us to examine two aspects of our hypotheses. First, adults are assumed to bring larger amounts of global prior knowledge to vocabulary learning than children, and are proposed to use this prior knowledge to support new learning (Section 1.6.1). Children generally have weaker prior knowledge at this earlier stage of development, but are proposed to benefit more from consolidation processes associated with sleep. In Chapters 3, 4, and 5, we present experiments comparing both children (aged 7-10 years) and adults (18-35 years), to examine developmental differences in influences of prior knowledge and offline consolidation on vocabulary learning.

Measuring individual differences

To better understand how the consolidation of new words might be supported by an individual's global vocabulary knowledge, experiments in Chapter 4 and Chapter 5 also incorporated standardised assessments of vocabulary into the analyses. This enabled us to test the prediction that those with good global vocabulary would be relatively better at consolidating words with more local connections to prior knowledge than those with fewer connections. Individuals with good vocabulary knowledge are predicted to have superior knowledge of the semantic and/or word-form neighbours, and therefore show a larger benefit for learning items that capitalise upon them.

2.1.3 Atypical development

In Section 1.6.3, we suggested that studying poor comprehenders had potential to inform consolidation in the context of impoverished prior semantic knowledge. Chapter 6 summarises the screening procedures used to identify children with good decoding skills but poor language comprehension. In Chapter 7, we present a study that tracked new word memory across equivalent periods of wake and sleep in poor comprehenders compared to good comprehenders. This experiment allowed us to test the hypothesis that poor comprehenders have specific difficulty in consolidating new words into longer-term vocabulary knowledge.

2.2 Statistical approaches

2.2.1 Use of mixed effects models

For all experiments, we used mixed effects models to incorporate participant- and item-level variability into a single analysis. This is in contrast to traditional ANOVA analyses, in which it is common to enter each participant's average score per experimental condition. The error term in an ANOVA represents variability in participant performance, enabling us to draw inferences about the population we are sampling from: not all participants will perform at the same level and show the same extent of experimental effects, but we want to infer that the effects are likely true for the population as a whole and not just the specific sample tested. However, the same has also been argued for the *stimuli* used in language experiments (Clark, 1973): we want to be able to conclude that the findings reflect language as a whole, and not just the specific words selected for our experiment.

To address this "fixed effects fallacy" (Clark, 1973), a common approach in linguistics has been to compute two separate analyses: one in which item-level data is averaged for each participant, and one in which participant-level data is averaged for each item. However, this approach still ignores some aspect of variability within each analysis. Mixed effects models permit participant- and item-level data in a single analysis. Compared to previous approaches, mixed effects models have greater statistical power, are less affected by missing data, and allow for both continuous and categorical variables within the model (Baayen, Davidson, & Bates, 2008). This latter point is particularly beneficial if we are to relate local prior knowledge manipulations (categorical) to individuals' global prior knowledge (continuous) in understanding vocabulary learning success.

2.2.2 Model fitting

There are not currently any broadly agreed guidelines to fitting mixed effects models. In modelling effects of experimental manipulations, a logical option is to retain all fixed effects of interest to the experimental hypothesis. Barr, Levy, Scheepers, and Tily (2013) recommend that confirmatory analyses of these datasets should incorporate maximal random effects as the gold standard – i.e., including random slopes for every fixed effect of interest. However, others have argued that maximal models can reduce statistical power in complex models (Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017), and favour a more parsimonious approach (Bates, Kliegl, Vasishth, & Baayen, 2015a). We frequently found that maximal models suffered with non-convergence, especially as the complexity of the experimental design increased in later experiments. To ensure a consistent approach across experiments presented in this thesis, we took two approaches to simplifying the models. First, we pruned higher-order interactions from the model if there was no evidence that they predicted performance. Second, we used a data-driven approach to model random effects structures. Barr et al. (2013) argued that data-driven approaches could obtain reasonable results providing that a liberal threshold is used to determine the inclusion of random slopes. Upon this guidance, we took a forwards best path approach to building the structure of random effects from an intercepts-only model. At each stage, we used likelihood ratio tests to select the remaining slope that best improved the fit of the model, before testing for further slope inclusion. We retained only random slopes that improved model fit under a liberal α-criterion ($p < .20$).

Although there is considerable variability in current practices of model fitting, Matuschek et al. (2017) highlight that *best* practice in such circumstances is a transparent one: ensuring that all data and code is released upon publication to enable others to investigate consequences of analytical decisions. As such, all experimental data and code is available online (see Section 2.3.3).

2.3 Reproducible science

2.3.1 Pre-registration

To enhance the transparency of the research process, the majority of the experiments presented were pre-registered on the Open Science Framework (Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012). These are accessible from the web links presented at the beginning of each chapter, and links to the pre-registered hypotheses are present in the methods sections. The majority of experiments made use of the Pre-Registration Challenge form created by the Center for Open Science (see Veldkamp et al., 2018, for discussion of pre-registration templates). Two exceptions are from data collected via MSc research projects: Chapter 3 Experiment 2 (no pre-registration), and Chapter 5 Experiment 1 (pre-registered analyses only).

Sometimes it was necessary to deviate from the pre-registered plans. In early studies, this generally resulted from ongoing learning about mixed effects models. In many respects, the changes reflected learning about a new statistical approach – e.g., understanding the need to set contrasts for factorial predictors and the relative merits of different approaches. In others, they reflected current and ongoing debates about best practice for these analyses (e.g., Matuschek et al., 2017), and different considerations for dealing with convergence issues. In later experiments, we were able to be more specific in our predictions and analysis plans. Deviations from initial plans are noted in all cases.

2.3.2 Experimental software

To facilitate sharing of materials, I prioritised use of open source software where possible. Experiments were usually programmed using DMDX (Forster & Forster, 2003) or OpenSesame (Mathôt, Schreij, & Theeuwes, 2012). Some experiments made use of remote data collection online. For early experiments, these were programmed using Qualtrics and QRTEngine (Barnhoorn, Haasnoot, Bocanegra, & van Steenbergen, 2015), for which experimental scripts are not in a shareable format. Later studies made use of Testable (Rezlescu, 2015; scripts shared online) or Gorilla (Anwyl-Irvine, Massonnié, Flitton, Kirkham, & Evershed, 2018; partial sharing available). For all studies, we made the stimuli separately available unless we did not have copyright permissions to do so (standardised assessments; images used in Chapter 7).

2.3.3 Open data and analyses

All experimental datasets are available online. For early experiments, these data are pre-processed and do not include excluded data. For later datasets (e.g., Chapter 5, Chapter 7), we progressed to making open as much of the data as possible. However, some pre-processing was always necessary to preserve anonymity, and vocal responses could not be made available.

Similarly, R Markdown was used to make the analyses for each experiment public. For the earlier published work (Chapter 4), this was limited to presenting how to produce the results from the data files made available. More recently, I have made increased attempts to make the data processing and model fitting processes more transparent. These are available as R Markdown scripts, as well as html files with integrated output.

Chapter 3. Manipulating Access to Semantic Knowledge

All experiment pre-registrations, materials, data, and analyses are available on the Open Science Framework: https://osf.io/35ftn

3.1 Abstract

A word's semantic neighbourhood is known to influence the processing of its form, highlighting the automaticity with which semantic knowledge is activated in the vocabulary system. We conducted three experiments to explore how this semantic activation influences new vocabulary acquisition, for both adults (Experiments 1, 3) and children (Experiment 2) who are proposed to bring different levels of prior knowledge to the task of learning. Participants were taught pseudowords (e.g., *oggice, marpan*) assigned to novel concepts with low versus high semantic neighbourhood density. These novel concepts were developed by adding a feature to base concepts selected to have low- versus high- density based on feature norms – for example, a *chicken* (base) *that sleeps upside down* (feature). Memory for the new items was tested on the same day, the next day, and one week later, to assess influences of semantic knowledge before and after opportunities for consolidation. There was no influence of semantic neighbourhood density on performance when accuracy was high (Experiment 1). However, at lower levels of performance (Experiments 2, 3) there was a benefit for recalling vocabulary from low- versus high-density neighbourhoods, for either recall of word-forms (adults) or their meanings (children). These density effects were apparent immediately after learning and did not change with opportunities for consolidation. We discuss the similarities and differences in children's and adults' activation of semantic knowledge during vocabulary learning.

3.2 Introduction

The ultimate goal of language is to convey meaning, and thus semantic representations play a central role in both word learning and processing. Semantic information is posited as essential for new phonological forms to become lexicalised and engage with existing vocabulary knowledge (Leach & Samuel, 2007), thereby facilitating the long-term consolidation of these new words in memory (Henderson et al., 2013c). Once part of an individual's vocabulary, the semantic properties of a word influence the speed at which words are recognised even in tasks where meaning is irrelevant for successful performance (e.g., Buchanan, Westbury, & Burgess, 2001), highlighting the automaticity with which semantic knowledge is activated upon encountering word-forms. Here, we explored ways in which new words might enter and engage with both developing and mature semantic systems, which presumably differ in their existing knowledge. Specifically, we asked whether a new concept could benefit from the rich semantic knowledge of its related concepts.

3.2.1 Conceptualising semantic knowledge in the vocabulary system

Evidence for a distributed semantic structure in the vocabulary system comes from speeded word recognition tasks, which show that rapid lexical processing can be influenced by a word's semantic properties. Semantic space can be conceptualised according to two broad sets of principles: language-based and object-based similarities (Buchanan et al., 2001). Language-based similarities relate to the co-occurrence of concepts in spoken and/or written language, resulting in a broad and rich set of linguistic associations. Object-based similarities document the content of the concepts themselves – for example, similarities in their physical properties. Both types of measure have been consistently demonstrated to influence word processing, even in tasks that place minimal demands on accessing semantic knowledge (e.g., Buchanan et al., 2001; Grondin, Lupker, & McRae, 2009; Pexman, Lupker, & Hino, 2002; Yates, Locker, & Simpson, 2003). For example, Pexman, Hargreaves, Siakaluk, Bodner, and Pope (2008) showed that words higher in semantic richness were responded to more quickly in a lexical decision task, and that both language- and object-based measures of richness contributed unique variance in reaction time. These studies demonstrate that semantic knowledge is automatically activated when processing incoming linguistic information; and that greater activation can facilitate efficient language processing.

3.2.2 Engaging semantic knowledge during learning

If pre-existing semantic knowledge is automatically engaged in lexical processing, then what might its role be in the learning of new linguistic information? One model of vocabulary learning draws upon the Complementary Learning Systems (CLS) model of memory (Davis & Gaskell, 2009; McClelland et al., 1995). The CLS model originally specified that gradual consolidation processes were required for new information to be strengthened from the rapid learning hippocampal system into the neocortical memory storage. More recently, this neocortical system has been re-conceptualised as “prior knowledge dependent” (Kumaran, Hassabis, & McClelland, 2016; McClelland, 2013) suggesting that related semantic knowledge should facilitate the acquisition of new information. From a language learning perspective, James, Gaskell, Weighall, and Henderson (2017) similarly proposed that existing linguistic knowledge may facilitate the consolidation of new vocabulary.

What is less clear is *how* or *when* the influence of prior knowledge plays out in acquiring new information. Whilst many studies have addressed how existing knowledge might help in the initial processes of identifying novel words to be learned (see Mitchell & McMurray, 2009), less is known about how it impacts the memory mechanisms engaged. From one perspective, new information that is closely related to and consistent with known information does not pose the same risk of interfering with existing knowledge, and thus places reduced demand on careful integration processes (Kumaran et al., 2016). As such, it may be that neocortical learning can proceed immediately, without the need for gradual consolidation. Alternatively, it may be that consolidation is still required, but that the rich connections made with existing knowledge during learning speed the rate at which neocortical learning can occur offline (e.g., Lewis & Durrant, 2011). In the present study, we assess how existing semantic knowledge influences the acquisition of related novel concepts, and aim to determine whether this influence requires offline consolidation to emerge.

3.2.3 Benefiting from semantic knowledge during learning

A number of developmental studies have suggested that a child’s semantic knowledge may assist them in learning new vocabulary. Borovsky et al. (2015) described a *lexical leverage hypothesis,* in which children benefit from recognising similarities between known concepts and new ones in learning. In an experimental study, they demonstrated that infants were more able to learn and recognise new words from categories that they had more knowledge about compared to categories for which they had lower levels of existing knowledge. Similarly, Perry et al. (2015) found that preschool children with larger shape-based noun vocabularies were more likely to remember object shapes during word learning. These studies provide direct evidence that a child's existing semantic knowledge may help them in the initial acquisition of new related concepts, although cannot address questions of consolidation as children were tested only on the same day as learning.

3.2.4 Competition from semantic knowledge during learning

However, not all studies find a benefit for existing semantic knowledge: others have demonstrated that existing knowledge may cause *interference* during learning. Tamminen et al. (2013) taught adults new pseudowords and concepts that had either sparse or dense semantic connections. To create novel concepts with these semantic knowledge connections, they selected existing concepts from either low- or high-density semantic neighbourhoods (as quantified by word association norms) and added a novel feature (e.g., *bee whose sting feels pleasant, crab that has a beak*). They tested participants' knowledge of the new pseudowords immediately after learning and after opportunities for consolidation (next day, one week later). Across all test sessions, participants were slower to respond and made more errors in a synonym judgement task for pseudowords with high-density novel concepts, suggesting that related semantic knowledge interfered with new lexical processing. Participants also completed a semantic categorisation task (animacy decision), for which a slowing of responses to high-density items did not emerge until after a more prolonged period of consolidation. Polysomnography recordings showed differences in sleep spindles – sleep architectural features associated with memory integration - during the night following learning, which the authors suggested may reflect the ease at which items in the sparse condition can integrate due to less inconsistent knowledge. By these findings, integrating into a dense neighbourhood triggers a slower consolidation process to work around competing information.

Whilst Tamminen et al. (2013) did not show an influence of semantic neighbourhoods in their explicit form and meaning recall measures, similar semantic interference has been demonstrated in explicit memory tasks when returning to the developmental literature. Storkel and Adlof (2009) quantified semantic set sizes of novel objects by collecting free associations from their pictures, largely influenced by visual similarity. In a subsequent learning task, preschool children were more accurate in identifying the names of objects from small semantic set sizes, suggesting that connections to existing knowledge can interfere in learning new information. In this study, the effect only emerged after opportunities for consolidation, consistent with previous adult studies demonstrating later engagement of new words with existing semantic knowledge (Clay et al., 2007; Tham et al., 2015).

3.2.5 The present study

The available evidence suggests that an abundance of related semantic knowledge can sometimes facilitate and sometimes interfere with new learning, yet what drives these differences and when they emerge during learning and consolidation is not well understood. One notable difference between the studies reviewed is that those which demonstrated interference used stimuli based on semantic associations, largely reflecting language usage. On the contrary, those which showed facilitation from related semantic knowledge used stimuli selected from object categories, drawing similarities across features. These object-based measures convey more information regarding the physical properties of the referent, which could arguably be more beneficial in learning about a new stimulus. For example, “cat” is a more frequent lexical associate of “bird” than “robin” (Nelson, McEvoy, & Schreiber, 2004), yet a person without prior knowledge of a bird would be easily misled in trying to learn from this association. These different ways of conceptualising semantic relationships have been shown to make unique contributions to different types of word processing tasks (Pexman et al., 2008), making it plausible that they may exert different influences on word learning. Furthermore, the age groups in the reviewed studies also vary from pre-schoolers to adults, making it challenging to interpret differences related to semantic knowledge in light of likely developmental differences. Using the same tasks and materials across age groups is thus also important to better understanding contributions of semantic knowledge to new vocabulary learning.

The present study used a similar design to Tamminen et al. (2013), but instead drew upon shared features as a measure of semantic neighbourhoods to address whether this better captures existing semantic knowledge that may facilitate word learning. We taught participants pseudowords and associated definitions, formed by adding a feature to known concepts from low and high feature density neighbourhoods. We tested explicit recall of the pseudowords and definitions immediately after learning, the next day, and one week later, to examine the influence of existing knowledge before and after opportunities for consolidation. A speeded semantic categorisation task was also used to index integration of the new words into neocortical vocabulary, given that known words with high semantic neighbourhood density are responded to more quickly in this task than words with low semantic density (Mirman & Magnuson, 2008). Tamminen et al. (2013) found that these more implicit semantic neighbourhood effects for trained pseudowords emerged only after a period of consolidation.

Although the studies described above span a broad age range from infants to adults, there is no clear developmental divide in the influence of semantic neighbours, and the differences in methodology and lack of direct comparisons presents a challenge to considering developmental differences. A study of known words taken from linguistic corpora suggested that young infants start by learning words from sparse semantic neighbours but increasingly benefit from dense neighbours as they age (Storkel, 2009). However, developmental differences have not been tested experimentally and no studies to our knowledge have considered the influence of semantic neighbours in school-aged children. There are two possibilities here: first, children may show smaller semantic density effects given that the measure is created from adult norms, and children may not have yet acquired the rich knowledge about concepts to have such extreme differences in low- versus high-density items. An effect in this direction would be in line with the proposal that adults can rely more on their greater amounts of prior knowledge to support new learning than children (James et al., 2017). Alternatively, we can consider that children could show larger semantic density effects, under the possibility an underdeveloped system may be more sensitive to the influence of existing knowledge. For example, Davies, Arnell, Birchenough, Grimmond, and Houlson (2017) showed that effects of psycholinguistic variables on lexical processing decline across the lifespan as the lexical system accumulates experience and maximises learning efficiency. We examine these possibilities across three experiments with adults (Experiments 1, 3) and children (Experiment 2).

3.3 Experiment 1

Our first experiment set out to address three main questions for adult word learning. First, whether newly trained semantic information can acquire the lexical properties of its neighbours, benefiting from rich semantic connections in speeded reaction time tasks. Second, whether novel words benefit from or are hindered by links to existing semantic knowledge during word learning. Third, what might be the time course of engagement with semantic knowledge?

These research questions were addressed by testing three hypotheses that were pre-registered on the Open Science Framework (http://osf.io/3vnsg) as follows: 1) Novel concepts that share lots of features with existing concepts should show a reaction time advantage in an animal decision task; 2) A large number of shared features will facilitate word learning, as demonstrated by superior performance in recall and recognition tasks; and 3) Effects of neighbourhood density will emerge only after a night's sleep (24-hour test) or longer period of consolidation (week follow-up test).

3.3.1 Experiment 1 Methods

Participants

Seventy-one participants (10 male; mean age = 19.99 years) were recruited through the University of York Psychology participant pool according to the following criteria: aged 18-35 years old, native monolingual English speakers, with normal or corrected-to-normal vision and hearing abilities, and without documented reading or language disorders. Of these, 66 participants completed all three follow-up tests at appropriate times, with five participants contributing only partial data to the analyses (2/3 tests).

The study was approved by the Research Ethics Committee of the Department of Psychology, University of York. Participants received £10 or course credit for their time.

Design and procedure

Participants attended a single training session in the University of York Psychology Department that lasted approximately 45 minutes. All participants learned stimuli from two semantic neighbourhood density conditions (low vs. high), and were asked to complete the memory tests from home at three time points: the same day as training (T1), the next day (T2), and one week later (T3). They were asked to complete the first memory test within two hours of completing the training session, and to complete each subsequent session at a similar time. We analysed data from all sessions completed on the correct day.

To identify differences in attention and general engagement across laboratory and home sessions, a short vigilance task featured at the beginning of the training and each test session. In this task, an X was presented on screen at randomly occurring intervals (2000-8000 ms; programmed in 5 ms intervals) and participants were instructed to press the spacebar as soon as they saw it. The task continued until participants had responded to 20 stimuli. They were also given the opportunity at the end of each test session to report anything that might have affected their performance during the tasks (e.g., interruptions, technical problems). No sessions were excluded on these bases.

Stimuli

Pseudowords were selected using the English Lexicon Project (Balota et al., 2007) according to the following criteria: 5-6 letters long, no orthographic neighbours, and a nonword rejection Z-score of -0.45 to 0.45 (i.e., an average range response time for rejection in a lexical decision task). Twenty-four bisyllabic pseudowords were selected such that each began with different vowels or consonant clusters, and that were judged to be easily pronounceable (Appendix A1).

Each pseudoword was assigned a novel concept of either high or low semantic neighbourhood density (counterbalanced across participants). Novel concepts were created by adding an additional feature to an existing (base) concept. For example, a *gorilla* (base concept) *that has green skin* (added feature). Critically, these base concepts were selected for having high (n = 12) or low (n = 12) semantic neighbourhood density according to the McRae, Cree, Seidenberg, and McNorgan (2005) feature norms[1]. Low-density base concepts have fewer features listed in the norms (≤ 16), and fewer of these listed features (≤ 14%) co-occur in other normed concepts. High-density base concepts have more features overall (≥ 18) and more of these (≥ 25%) also co-occur in other concepts. The two groups of stimuli were otherwise well-matched on measures of frequency, age of acquisition, imageability, concreteness and word length (Table 2). A pilot study of these base concepts supported a reaction time benefit for high-density concepts (mean difference = 14 ms; $t(70)$ = 2.56, p = .01).

[1] Only 18 of the 24 items were also entries in the Florida Free Association Norms. These indicated that the two sets would likely differ in semantic neighbourhood density by this measure, with high-density concepts having more associates (M = 17.33) than low-density concepts (M = 12.22; p= .05).

Table 2. Properties of stimuli in the low and high semantic neighbourhood density conditions.

	No. of features[a]	% features correlated[a]	AoA[b]	Frequency[c]	Log10 freq[c]	Imageability[d]	Concreteness[e]	No. of phonemes
Low	12.75	5.58	5.14	16.41	1.02	607.56	4.89	4.33
High	18.92	40	5.17	16.38	1.13	616	4.89	4.5
p	<.001*	<.001*	.96	1	.56	.61	.9	.79

[a]McRae et al. (2005). [b]Kuperman et al. (2012). [c]CELEX English linguistic database (Baayen et al., 1995). [d]MRC Psycholinguistic database (Coltheart, 1981). [e]Brysbaert et al. (2014). *significant difference between low vs high semantic density items at $p<.05$.

The added features that made each concept novel were also selected from the McRae et al. (2005) norms, and each occurred only once in the norms to minimise the influence of additional semantic neighbourhoods. The features were drawn from a range of perceptual, behavioural and functional categories, which were matched in type and counterbalanced in assignment to low- and high-density base concepts items (Appendix A2). To ensure that these combinations of base concepts and features did not differ in plausibility across low- and high-density conditions, 58 adults completed online ratings of how plausible they would find each item in a children's storybook. High- and low-density items did not differ in plausibility in either of these base-feature counterbalanced groups (*p*s >.2). Each counterbalanced base-feature group could be assigned either of the two pseudoword lists, making four counterbalancing conditions in total.

Training tasks

All training tasks were presented using DMDX software v5.1.3.4 (Forster & Forster, 2003). A voice recorder in the experiment booth was used to check that participants were engaging with the training and vocalising the new word-forms during the first two tasks.

Form repetition. Each item was presented simultaneously over headphones and in the centre of the screen (1500 ms), before being replaced by a visual cue for participants to repeat the pseudoword out loud. Each word was presented and repeated three times.

Definition. Each item was presented as above, followed by a cue to repeat the pseudoword aloud. After 2000 ms, the definition of the pseudoword appeared on screen beneath it. Participants were given 8000 ms to try and learn the meaning of the word, and were encouraged to visualise the novel concept to help them. Each item was presented only once during this task.

Sentence creation. Participants were presented with the pseudoword and definition onscreen, and were asked to type a sentence containing the new word. Each item was presented only once, and there was no time limit for completion.

Meaning matching. The pseudoword and four possible options for its definition were presented on screen, and participants were asked to select the correct meaning. The distractors for each item were selected from the other newly learned definitions. Each item was presented twice, with feedback on the correct match.

In total, adults had seven exposures to each of the word-forms, and ten exposures to each novel definition.

Test tasks

All test tasks were hosted online using Qualtrics (Qualtrics, 2014) and QRTEngine version 18 (Barnhoorn et al., 2015). The tasks were presented in a fixed order for all participants, as listed below.

Cued form recall. Participants were presented with the first consonant(s) and vowel of the word-form on screen, and were asked to type the whole word into the computer. Instructions encouraged partial answers even if participants were not certain. Items were presented in a randomised order, and there was no time limit for completion. Responses were subsequently scored on the basis of whole word accuracy (0, 1), with phonologically equivalent spellings also marked as correct (e.g., ‘rejeel’ instead of ‘rejele’, ‘oggis’ instead of ‘oggice’).

Cued meaning recall. Participants were given the written word-form and asked to type as much of the definition as they could remember. Items were presented in a randomised order, and there was no time limit for completion. A total of two points could be awarded per item for correctly recalling the base concept and the added feature.

Semantic categorisation. Participants were presented with the written word-form and were asked to make speeded judgements about whether the concepts were animals by pressing the Z key for Yes and M key for No. Items were presented in a randomised order. Participants were asked to respond as quickly and accurately as possible, and each trial terminated upon response or after 3 seconds. To allow for adjustment to the task and responses, the experimental task was preceded by 24 practice trials using existing English words, providing feedback for erroneous responses.

Analyses

Data were analysed in R (R Core Team, 2015), using *lme4* (Bates, Maechler, Bolker, & Walker, 2015b) and *ordinal* (Christensen, 2015) to fit mixed effects models. For binomial models, Wald's Z was used to determine statistical significance. For linear models, we report significance computed using *lmerTest* (Kuznetsova, Brockhoff, & Christensen, 2017). Although we pre-registered an initial plan for maximal models with all random effects and slopes (Barr et al., 2013), our analyses frequently suffered convergence problems and we adopted a more parsimonious modelling approach for later studies (osf.io/yk3d5; Bates et al., 2015a). To ensure a consistent approach across all experiments presented here, we take this parsimonious approach to all models: first fitting an intercepts-only model with fixed effects of test session, semantic density, and their interaction, and then pruning away the interaction if not contributing to model fit ($p < .2$). We then use a forward "best-path" approach to test for the inclusion of appropriate random slopes (Barr et al., 2013). The results presented are from the most complex model supported by the data, and the model tables are presented in Appendix A (A3-A6). The data and details of the full modelling procedure for each analysis are available on the OSF (https://osf.io/35ftn).

Fixed effects were deviance coded to enable interpretation of each predictor in relation to the overall mean. Test session is a three-level factor, and we set two orthogonal contrasts to interpret the data: *delay1* tested for differences in memory performance without versus with opportunities for consolidation (T1 vs. T2&T3); *delay2* tested for continued changes across the week (T2 vs. T3).

3.3.2 Experiment 1 Results

Cued form recall

On the same day as training (T1), participants could successfully recall an average proportion of .30 (*SD* = .46) word-forms. Recall improved after opportunities for consolidation (*delay1:* $\beta = 0.30$, *SE* = 0.03, *Z* = 11.70, $p < .001$), leading to higher performance at both T2 (*M* = .43, *SD* = .50) and T3 (*M* = .46, *SD* = .50). This continued improvement across the course of the week was also statistically significant (*delay2:* $\beta = 0.11$, *SE* = 0.04, *Z* = 2.65, *p* = .008). There was no indication that semantic neighbourhood density influenced recall of the word-forms (*p*s > .2).

Cued meaning recall

Participants could score up to two points for each definition, and achieved an average point score of 1.32 (*SD* = 0.94) per item at T1. Performance declined after the first day (*delay1:* β = -0.10, *SE* = 0.02, *Z* = -4.09, *p* <.001), and between T2 (*M* = 1.27, *SD* = 0.95) and T3 (*M* = 1.16, *SD* = 0.97; *delay2:* β = -0.14, *SE* = 0.04, *Z* = -3.34, *p* = .001). There was no influence of semantic neighbourhood density on memory for the new word-forms, before or after opportunities for consolidation (*p*s > .5).

Semantic categorisation

Accuracy. Mean accuracy on the semantic categorisation task was .77 (*SD* = .42), which remained stable over time (*p*s > .4). There was no effect of semantic neighbourhood density, either alone (*p* = .701) or in interaction with test session (*p*s > .15).

Reaction time. One participant was removed from the RT analyses due to chance levels of performance. We log-transformed the RT data to remediate issues of non-normality (although report raw means for ease of interpretation); and also removed responses that were < 200 ms or ≥ 2.5 standard deviations above the participant's condition mean. We analysed RTs to correct responses only, leaving 75.69% of the original scores for analysis.

Reaction times were slowest at T1 (*M* = 1041 ms, *SD* = 404 ms; *delay1:* β = -0.05, *SE* = 0.01, *t* = -8.28, *p* < .001). Performance continued to speed between T2 (*M* = 917 ms, *SD* = 323 ms) and T3 (*M* = 856 ms, *SD* = 297 ms; *delay2*: β = -0.04, *SE* = 0.01, *t* = -4.19, *p* < .001). However, there was no influence of semantic neighbourhood density on response times *(p* = .99).

3.3.3 Experiment 1 Summary

Experiment 1 looked at the learning and consolidation of pseudowords that had been assigned novel semantic concepts in adults. Recall of the new word-forms was weaker than recall for the meanings, but improved with opportunities for offline consolidation. Recall of the associated meanings was much higher, but declined slightly across the week. This pattern of findings is consistent with previous studies showing offline benefits for recall of word-forms but not semantic knowledge trained via presenting definitions (e.g., Tamminen & Gaskell, 2013).

Our primary research questions related to the ways in which the new words would engage with existing semantic knowledge, as indicated by performance differences related to the semantic neighbourhood density of the novel concepts. Consistent with the findings of Tamminen et al. (2013), there was no influence of semantic neighbourhood density in either explicit memory measure for adults. However, we also found no effect of semantic neighbourhood density in the semantic categorisation task, suggesting that the new words did not adopt the implicit lexical processing properties of their related semantic concepts. Perhaps new words do not benefit from existing shared features when accessed only via an existing concept (i.e., learning that the concept is a *gorilla*), but rather build up connections in these networks through direct experiences with the concepts, building feature similarities independently that converge on known concepts (e.g., *is strong, lives in jungles, eats bananas*). Alternatively, it may be that processing benefits do not emerge without significantly more time and exposure than included in the presented study[2].

3.4 Experiment 2

Experiment 2 was designed to test the same research questions as Experiment 1, but in school-aged children. The experimental materials were adapted to make them suitable for 7-to-9-year-olds, allowing us to compare influences of semantic knowledge in children's and adults' word learning. Although Experiment 2 was not pre-registered, this experiment was run in parallel to Experiment 1 - with the same hypotheses - and we maintain a consistent approach to analysis for comparability.

3.4.1 Experiment 2 Methods

Participants

Two classes of children took part in the study, and were recruited via two schools in North Yorkshire. The resulting sample included 51 children (25 male) aged 7-10 years (*M* = 8.67 years). One additional child was excluded from analyses due to hearing difficulties. Two of the included children were absent on the second day of testing, and thus only contributed data for two out of the three follow-up tests.

[2] To further explore this possibility, we invited participants for a delayed follow-up test three months later. Only 28 participants completed the activities. There remained marginal main effects of density in the stem completion task, and weak statistical evidence for density effects emerging in semantic categorisation accuracy at the later test point. Given the exploratory nature and weak statistical power, these are presented as supplementary materials on the Open Science Framework.

The study was approved by the Research Ethics Committee of the Department of Psychology, University of York. Consent was obtained from the school head teachers, and parents were given the opportunity to opt their child out of taking part.

Design and procedure

Children completed a single training session in a whole-class setting, which lasted approximately 45 minutes. Test sessions were then conducted individually in a quiet setting outside the classroom at three time points (as Experiment 1): the same day (T1), the next day (T2), and one week later (T3). Measures of vocabulary and matrix reasoning from the Wechsler Abbreviated Scale of Intelligence II (Wechsler, 2011) were also collected during these sessions.

Stimuli

The stimuli were a subset of the items used in Experiment 1[3]. Children learned 16 new pseudowords and concepts, 8 from each density condition. These two density groups remained closely matched on their base concept characteristics, as above.

Training tasks

Training tasks were adapted from Experiment 1 to make them more appropriate for the younger age group and suitable for whole-class administration. Children were given workbooks to support their learning, and were guided through a number of tasks using a PowerPoint presentation projected at the front of the classroom. For this training, the first three tasks were completed for each item in turn. Meaning matching was completed afterwards.

Form repetition. Children heard each new word-form spoken by the experimenter, with its orthographic form projected on the PowerPoint at the front of the classroom. They repeated the pseudoword aloud twice, and subsequently copied it into their workbooks.

Definition repetition. Children were introduced to the definition of each word-form, and again repeated it aloud twice.

Drawing task. A drawing task was used in place of the sentence creation task to reduce demands on children's writing ability. Children were given 30 seconds per

[3] Due to an error, there were minor differences in the novel word forms and pronunciations assigned to the concepts (see Appendix A1).

item to draw a picture of the new concept, designed to help them to engage with its different features.

Meaning matching. After the workbooks had been collected, further learning and feedback took place via a multiple choice quiz. In the first round, a pseudoword and three possible options for its definition were presented on screen, and children had to show their answer by raising one, two or three fingers. In the second round, the definition was presented and the children had to choose the correct word-form to match. Each item was presented once in each round, with the correct answer provided after each one.

In total, children heard each new word-form nine times, and each definition six times.

Test tasks

Children sat the same three test tasks as adults, and an additional form recognition task that was designed to capture word-form learning should the recall task be too challenging. All test tasks were presented using DMDX, with item order randomised. They were presented in the following fixed order.

Cued form recall. As in Experiment 1, participants had to complete the word from a partial cue. In this version, children were simultaneously provided with auditory and visual presentations of the cue, and produced oral responses that were transcribed by the experimenter.

Form recognition. Children were presented with auditory and orthographic presentations of the pseudoword alongside a corresponding foil in which the final vowel was changed (Appendix A1). Both of the written stimuli remained on screen for up to 7 seconds, or until the child had selected their answer with a key press response.

Semantic categorisation. Children completed a speeded animal judgement task as in Experiment 1, but with a simultaneous auditory presentation of the stimuli. In this version, each item remained on screen for up to 7 seconds or until a response, and children responded with a key press.

Cued meaning recall. Children were given an auditory and visual presentation of the word-form, and asked to provide as much of the definition as they could remember (as Experiment 1). Verbal responses were transcribed by the experimenter.

Analyses

Analyses were conducted as for Experiment 1 (model tables can be found in Appendix A: A7-A11). Graphs were made using *ggplot2* (Wickham, 2009) with *ggpirate* (Braginsky, 2018).

3.4.2 Experiment 2 Results

Cued form recall

Children recalled a mean proportion of .20 (*SD* = .40) of the word-forms at T1 but performance improved substantially with opportunities for offline consolidation (Figure 2a; *delay1:* β = 0.95, *SE* = 0.05, *Z* = 21.10, *p* <.001). Recall continued to improve between T2 (*M* = .51, *SD* = .50) and T3 (*M* = .80, *SD* = .40; *delay2:* β = 0.91, *SE* = 0.07, *Z* = 13.35, *p* <.001). There was no influence of semantic neighbourhood density in recall of word-forms, alone or in interaction with test session (*p*s > .6).

Form recognition

Children could successfully recognise the new word-forms at above chance levels at T1 (*M* = .83, *SD* = .38), and improved at subsequent tests (T2: *M* = .92, *SD* = .28; T3: *M* = .94, *SD* = .24). This effect of test session was statistically significant across both contrasts (*delay1:* β = 0.39, *SE* = 0.05, *Z* = 8.07, *p* < .001; *delay2:* β = 0.21, *SE* = 0.10, *Z* = 2.08, *p* = .037), again demonstrating significant improvements in form knowledge with opportunities for offline consolidation. As with the recall of word-forms, there was no influence of semantic neighbourhood density on their recognition (*p*s > .18).

Cued meaning recall

Children showed much poorer learning of the definitions than adults, scoring an average of .36 out of two for each item at T1 (*SD* = .76). There were no significant changes in performance across test sessions (*p*s > .36), but a significant difference in memory for words from different semantic neighbour conditions (β = -0.48, *SE* = 0.18, *Z* = -2.62, *p* = .009). Children were better at recalling definitions with low semantic neighbourhood density (*M* = .47, *SD* = .84) than high semantic neighbourhood density (*M* = .26, *SD* = .67; Figure 2c). There was no evidence of an interaction between test session and semantic neighbourhood density (pruned from the final model; *p* = .687).

Semantic categorisation

Accuracy. Performance was very low on the semantic categorisation task (*M* = .59, *SD* = .49). Neither test session nor semantic neighbourhood density influenced accuracy on this task (all *p*s > .4).

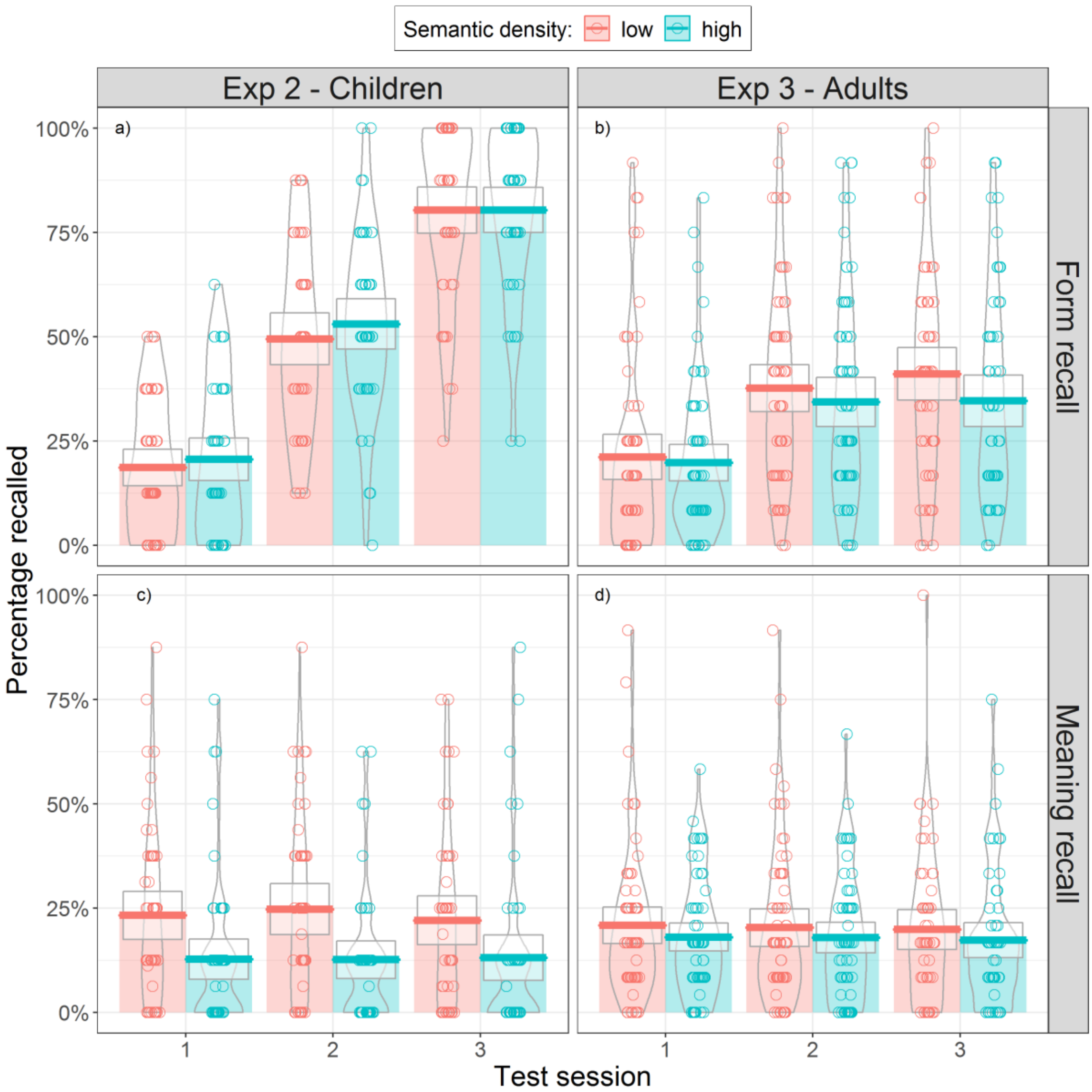


Figure 2. RDI plots of the percentage of items recalled in the explicit recall tasks for Experiments 2 (Children) and 3 (Adults). RDI plots incorporate Raw data, Descriptive statistics, and Inference. As such, circles represent an individual participant's condition mean, with grey outlines marking overall density of the data. Thick horizontal lines represent condition means, and the boxes 95% confidence intervals. Note that children learned fewer items (*n* = 16) than adults (*n* = 24).

Reaction time. We were cautious in analysing the RT data considering that performance accuracy was so low in this task, but removed participants who were at/below chance performance (n = 11). The data were log-transformed to remediate issues of skewness in model fitting. We also removed responses <200 ms or ≥2.5 standard deviations above each participant's condition mean. We analysed RTs to correct responses only, leaving 49.05% of original trials.

Responses were significantly slower at the first test point (M = 2154 ms, SD = 1211 ms) compared to later test points (β = -0.07, SE = 0.01, t = -7.66, p < .001), with weak statistical evidence of further speeding between the day (M = 1833 ms, SD = 1150 ms) and week (M = 1696 ms, SD = 977 ms) memory tests (β = -0.03, SE = 0.02, t = -1.84, p = .066). There was no influence of semantic neighbourhood density on reaction times (ps > .14).

3.4.3 Experiment 2 Summary

As with adults in Experiment 1, children showed improvements in their memory for the new word-forms after opportunities for consolidation. Children were much poorer in their learning of the word meanings: they showed low performance in both the meaning recall and semantic categorisation tasks that neither improved nor declined across test sessions. However, it should also be noted that they had fewer exposures to the definitions than adults (6 vs. 10).

Most interestingly, Experiment 2 demonstrated that existing semantic knowledge can influence new vocabulary acquisition in school-aged children: recall of novel concepts from low-density semantic neighbourhoods was higher than for those from high-density semantic neighbourhoods. This finding is more in line with studies that show interference from existing knowledge in learning related concepts (Storkel & Adlof, 2009; Tamminen et al., 2013), despite using a feature-based manipulation. These effects did not require consolidation to emerge, nor did they change with consolidation, suggesting that semantic knowledge was activated automatically during learning and/or retrieval.

Whilst Experiment 2 showed clear effects of semantic neighbourhood in the meaning recall task, there was no evidence of such an effect in our adult experiment. This difference raises interesting possibilities regarding a developmental difference in activating and/or inhibiting semantic knowledge during learning. However, two key issues prevent interpretation of these differences. First, there were a number of methodological differences between the two studies, and differences in training tasks may have led to the use of different learning strategies. Second, children's memory performance was much lower than adults - particularly for the semantic aspects of their new learning - such that performance differences may also account for the differences across experiments. To investigate these questions further, Experiment 3 aimed to reduce adult levels of learning to investigate whether semantic neighbourhood effects emerge with weaker memory traces in adults, using comparable methodology.

3.5 Experiment 3

Three hypotheses were pre-registered on the OSF (http://osf/io/yk3d5): 1) Cued recall for word-forms will improve over time, consistent with Experiments 1 and 2, and with extant evidence supporting strengthening of novel word-forms with offline consolidation; 2) Where a neighbourhood density effect emerges, we predict that low-density items will be better learned than high-density items; and 3) If the absence of a density effect in the definitions task for adults was driven by their higher performance, then we would expect a neighbourhood density effect to emerge at lower performance levels in this task[4]. However, if the difference is driven by developmental differences in the semantic system, we would expect no effect of density in the definitions task for adults regardless of performance levels.

3.5.1 Experiment 3 Methods

Participants

70 participants were recruited via the University of York Psychology Department participant pool according to the following criteria: native monolingual English speakers, aged 18-35, with normal or corrected-to-normal hearing and vision, and no reading or language disorders. Three participants did not complete more than one of the three follow-up sessions, and were thus excluded from analyses. The final sample consisted of 67 participants (14 male), with a mean age of 20.33 years (*SD* = 2.54). Nine participants contributed only partial data (2/3 sessions) having not completed one of the sessions on the correct day.

[4] Note that the pre-registration refers to a significant effect of semantic neighbourhood density for cued form recall in Experiment 1. This was due to an analysis error in which test session was entered as a continuous rather than categorical predictor.

Participants received £10 or course credit for their time. The study was approved by the Department of Psychology Research Ethics Committee at the University of York.

Design and procedure

To make Experiment 3 as comparable as possible to Experiment 2, we conducted training in a group setting lasting approximately 45 minutes. Test sessions were then completed online at the same three test points: the same day (T1), next day (T2), and one week later (T3). Participants were asked to complete the first session within 2 hours of training, and complete each session at a similar time (by 6pm at the latest). We included all sessions completed on the correct day for analysis. Test sessions were identical in format to Experiment 1, with the addition of the form recognition to ensure equivalent spaced exposures to the child experiment.

Stimuli

As Experiment 1, but simplified to two counterbalancing conditions to facilitate group training. The two versions altered the pseudoword assigned to each base concept, as well as the added feature that made each concept novel.

Training tasks

The training tasks were identical to Experiment 2, apart from form and definition repetitions were reduced to one per item. Only one round of meaning matching was administered, presenting each definition only once, with three options for its word-form on each occasion. This meant that participants had five exposures to the new word-forms in total, and only two exposures to the definitions, intended to reduce performance levels in line with children. Participants circled their meaning matching answers (1, 2, or 3) in an additional training booklet.

Test tasks

The four test tasks were set up as Experiment 1, except that the form recognition and semantic categorisation tasks were programmed using Testable (Rezlescu, 2015) and accessed via a link in the Qualtrics survey. This re-programming was due to QRTEngine being discontinued.

Analyses

Analyses were conducted as in Experiments 1 and 2. Full model tables can be found in Appendix A (A12-A16).

3.5.2 Experiment 3 Results

Cued form recall

The proportion of word-forms recalled on the same day of learning (M = .21, SD = .40) was highly comparable to Experiment 2 (M = .20, SD = .40), and significantly improved after opportunities for offline consolidation (*delay1:* β = 0.39, SE = 0.03, Z = 13.32, p < .001; Figure 2b). Unlike Experiments 1 and 2, there were no significant improvements between T2 (M = .36, SD = .48) and T3 (M = .38, SD = .49; p = .122).

There was also a small but statistically significant effect of density for this task (β = -0.11, SE = 0.05, Z = -2.254, p = .025): word-forms associated with low neighbourhood density concepts were better recalled (M = .33, SD = .47) than those associated with high-density concepts (M = .29, SD = .46), despite no explicit demand on meaning retrieval in this task. This density effect did not change over time, and the interaction was pruned from the final model (p = .383).

Form recognition

A technical issue meant that T1 form recognition and semantic categorisation data from the first set of participants was not saved from Testable (n = 9), and this issue also affected a later session for two participants. Unfortunately it was not possible to replace these participants due to timing constraints, and our main hypotheses related to the explicit recall measures for this experiment. We removed any participants who did not have data from at least two of the three sessions, leaving 65 participants for these analyses.

Recognition of the new word-forms was much higher than participants' ability to recall them. Performance was lowest at the first test point (M = 0.91, SD = 0.29; *delay1:* β = 0.11, SE = 0.04, Z = 2.59, p = .010), but there were no further changes in performance between the day (M = .94, SD = .24) and week (M = .93, SD = .26; p = .578) tests. There was a significant effect of neighbourhood density (β = 0.21, SE = 0.11, Z = 1.97, p = .049): performance was marginally higher for high-density items (M = .93, SD = .26) than low-density (M = .92, SD = .27). However, there was no evidence of an interaction with test session (pruned from final model; p = .951).

Cued meaning recall[5]

Recall of the word meanings was much lower in this experiment, as intended: participants scored an average of 0.39 (*SD* = 0.78) points per item at the first test, which did not change over time (*p*s > .7; Figure 2d). Whilst this level of performance was highly comparable to T1 for Experiment 2 (*M* = .36, *SD* = .76), recall of meanings was not affected by the semantic neighbourhood density of the concepts in adult participants (*p* = .704) as it had been for children.

Semantic categorisation

Accuracy. All fixed effects were retained in the model. Accuracy was generally very low (*M* = .57, *SD* = .50), and did not change across the course of the week (*p*s > .35). There was also no significant effect of neighbourhood density (*p* = .508).

Reaction time. At this low level of performance, 16 participants were excluded from RT analyses on the basis of chance-level performance. Only 44.98% of the data was retained after data trimming (as above), and so caution is needed in interpreting these data. Modelling was carried out on the log-transformed data, and showed only a decrease in reaction time across test sessions: participants were slowest at the first test (*M* = 1202 ms, *SD* = 495 ms; *delay1:* $\beta = -0.08$, *SE* = 0.01, $t = -6.53$, $p < .001$), and continued to improve between the day (*M* = 1017 ms, *SD* = 430 ms) and week (*M* = 911 ms, *SD* = 403 ms) memory tests (*delay2*: $\beta = -0.05$, *SE* = 0.02, $t = -2.74$, $p = .009$). There was no effect of neighbourhood density ($p = .330$).

3.5.3 Experiment 3 Summary

In Experiment 3, we sought to reduce adults' learning levels to aid in interpreting differences between Experiments 1 and 2. The performance of adults on the T1 explicit memory tasks indicates that this reduction in performance was successful: adults recalled a comparable proportion of the stimuli as children in Experiment 2 across the different tasks, although it should be noted that the overall information learned was still higher for adults as they were provided with more items (24 vs. 16). As with the previous experiments, memory for new word-forms improved across the week, whereas definition knowledge remained stable.

[5] One participant did not complete 2/3 definitions tests, and was excluded from this analysis.

At this lower level of performance, a semantic neighbourhood density effect emerged for adults. In contrast to children, adults only showed this effect in their memory for the new word forms – despite no explicit demands on accessing semantic knowledge in these tasks. The effects on recall were similar in direction to Experiment 2, showing a high-density disadvantage in recalling the word-forms, yet were also accompanied by marginal benefit for recognising high-density items. However, adults showed equivalent recall of meanings across both semantic density conditions.

3.6 General Discussion

We examined the influence of semantic neighbourhood density on adults' and children's language learning. Across all three experiments, participants' recall of word-forms improved across the course of the week, whereas recall of the associated definitions either remained stable or declined. Where influenced by semantic neighbourhood density, memory for low-density items showed a recall advantage over high-density items, consistent with Tamminen et al.'s (2013) interpretation of competition when training novel concepts into high-density networks. Interestingly, these neighbourhood effects were apparent in recall of word-forms for adults, but recall of definitions for children. However, there were no influences of semantic neighbourhood density in the speeded semantic categorisation task, suggesting that the novel items had not adopted the lexical processing characteristics of their neighbours during the time-span of these experiments.

In all three experiments, there were increases in recall and recognition of the new word-forms across the course of the week. These improvements are consistent with previous studies demonstrating benefits of offline consolidation for this aspect of word knowledge (e.g., Henderson et al., 2013c; Henderson et al., 2012; Storkel, 2001). Interestingly, children showed greater benefits of consolidation on their form recall than adults: children showed more substantial improvements at each test - even when adults started at a similar level of performance - and were more likely to show continued improvements between the day and week memory tests. These developmental differences are in line with a recent study that also showed greater benefits of offline consolidation for children versus adults (James, Gaskell, & Henderson, 2018), and are hypothesised to reflect enhanced levels of slow-wave sleep contributing to consolidation processes during development. However, it is also important to note that it is not possible to isolate influences of offline consolidation from retrieval practice and spaced exposures in the present experiments: participants had additional exposures to the new word-forms in the definition and semantic categorisation tasks at each test point. Thus whilst still revealing potentially interesting developmental differences, we can only speculate on the possible mechanisms.

Our primary research questions related to the ways in which new learning would be influenced by associated semantic knowledge, and there was some evidence that semantic neighbourhood density influenced the recall of meanings in children and word-forms in adults. We initially hypothesised that concepts classified primarily by feature norms might show a semantic density benefit in word learning, on the premise that feature similarity conveys more concrete and informative properties about the referent than the language-based norms used in previous studies (Tamminen et al., 2013). Although there was weak statistical evidence for a semantic density *benefit* for adults' form recognition in Experiment 3, this was very small (mean difference = 0.5%) and not related to our primary hypotheses for this final experiment. Overall, there was more evidence in line with a low-density benefit in the present study, supporting earlier findings that high-density neighbours compete during learning and/or retrieval (Storkel & Adlof, 2009; Tamminen et al., 2013). The contrast between tasks is somewhat puzzling, but differing influences of prior knowledge on recall and recognition have been demonstrated in other paradigms (e.g., Storkel, Armbrüster, & Hogan, 2006).

Key to this semantic interference may be that trained concepts will have remained very near in semantic space to their base concepts, differing only by a single feature. Mirman and Magnuson (2008) showed that an abundance of *near* neighbours – as defined by concepts sharing more than half of the target's features – slowed decisions in a semantic categorisation task, whilst it is distant neighbours that drive the overall facilitation seen in word recognition studies. Our novel concepts were all near-neighbours of their associated base concepts (with only a single feature differing), but base concepts from the high-density condition were likely to include more near-neighbours than those from the low-density condition. This influence of semantic distance may also account for the contrasting findings reviewed earlier: studies using stimuli with large numbers of overlapping features showed a negative impact of semantic density (e.g., Storkel & Adlof, 2009), whereas infant studies that assessed broader categorical knowledge related to the to-be-learned items showed facilitation (e.g., Perry et al., 2015). Perhaps this broader approach to semantic knowledge and measuring individual differences will be useful to extend to older children and adults, if we are to better understand potential benefits for related semantic knowledge in vocabulary learning.

In line with CLS models of vocabulary learning (Davis & Gaskell, 2009), we also predicted that effects of neighbourhood density would be more likely to emerge after opportunities for consolidation, following increased opportunities for the new lexical representations to engage with neocortical vocabulary. However, effects of semantic neighbourhood were consistent across all three test sessions for explicit recall tasks, and did not emerge for semantic categorisation at any test point (which we included as a marker for neocortical integration). Interestingly, the density effects emerged only for explicit tasks and in the context of low performance levels, suggesting that semantic density is perhaps most influential for these measures when memory traces are fragile. The influence of semantic density on these early stages of learning could better informed by trial-level analyses during learning, but is also supported by our lack of neighbourhood effect for the semantic categorisation task: this implicit task provided no evidence of neocortical integration. In contrast, Tamminen et al. (2013) showed no influence of semantic neighbours on explicit measures of memory, but an emergent density effect for the semantic categorisations task at the week test suggested integration of the novel concepts into existing vocabulary knowledge. Experiments 2 and 3 moved away from questions of semantic integration to better understand the early-emerging density effects seen in explicit the explicit memory measures. However, the transition between influences of semantic knowledge on early explicit knowledge of words to implicit markers of semantic integration clearly warrants further investigation.

Experiments 2 and 3 did indicate that influences of semantic neighbourhood density may differ across children and adults: children showed these effects in their explicit recall of definitions, whereas adults' recall of word-forms showed a semantic neighbour effect despite not requiring retrieval of associated semantic information. Such task differences were not anticipated and may be spurious (for example, adults showed a non-significant trend in the same direction as children for meaning recall), and thus we can only speculate on potential mechanisms. However, given evidence that semantic neighbours can influence very low-level processing of phonological forms (e.g., Buchanan et al., 2001), it may be that the mature semantic system activates this knowledge so automatically during learning that it affects the resources available to encode or retrieve associated word-forms. For children, this semantic activation may be less automatic, and only engaged during relevant tasks. Given that automatic activation of semantic knowledge was not seen in the semantic categorisation task for any experiment, it may be that semantic knowledge was triggered via presentation of the base concept during encoding - with lasting impact on its representation - rather than by the learned associations with the pseudoword not yet consolidated within the timeframe of the experiment. However, it is not possible to draw clear distinctions between encoding- and retrieval-related accounts within these experiments.

In summary, the present study showed that novel concepts with similar features to many known objects were more challenging to learn and/or recall than those with fewer neighbours. This finding corroborates those of Tamminen et al. (2013) whilst using a different conceptualisation of semantic space and across two different age groups. We demonstrated that influences of semantic neighbours on explicit recall of new information can emerge at the early stages of word learning if average performance is low, and persist across a week-long period of consolidation. These influences may be distinguishable from the later integration with existing knowledge seen in previous studies, leading us to conclude that close semantic neighbours can interfere with explicit knowledge of word-forms as well as for later consolidation into existing vocabulary knowledge. However, other studies have clearly demonstrated some *benefits* in related semantic knowledge for new vocabulary learning. In drawing parallels with word recognition research, we propose that these benefits of semantic knowledge for new learning might arise from more distant and varied connections than trained in the present study. Experiments that can capture these broader influences of semantic knowledge – and individual differences in them – will contribute to a more comprehensive understanding of how vocabulary learning might change as the semantic system develops.

Chapter 4. Manipulating Access to Word-Form Knowledge

Previously published as:

James, E., Gaskell, M.G., & Henderson, L.M. (2018). Offline consolidation supersedes prior knowledge benefits in children's (but not adults') word learning. Developmental Science, e12776.

All experiment pre-registrations, materials and data are available on the Open Science Framework: https://osf.io/s2628/

4.1 Abstract

Prior linguistic knowledge is proposed to support the acquisition and consolidation of new words. Adults typically have larger vocabularies to support word learning than children, but the developing brain shows enhanced neural processes that are associated with offline memory consolidation. This study investigated contributions of prior knowledge to initial word acquisition and consolidation at different points in development, by teaching children and adults novel words (e.g., *ballow*) that varied in the number of English word-form "neighbours" (e.g., wallow, bellow). Memory for the novel word-forms was tested immediately after training, the next day, and one week later, to assess the time-course of prior knowledge contributions. Children aged 7-9 years (Experiments 1, 3) and adults (Experiment 2) recalled words with neighbours better than words without neighbours when tested immediately after training. However, a period of offline consolidation improved overall recall and reduced the influence of word-form neighbours on longer-term memory. These offline consolidation benefits were larger in children than adults, supporting theories that children have a greater propensity for consolidating phonologically distinctive language information. Local knowledge of just a single word-form neighbour was enough to enhance learning, and this led to individual differences in word recall that were related to adults' global vocabulary ability. The results support the proposal that the relative contributions of different learning mechanisms change across the lifespan, and highlight the importance of testing theoretical models of word learning in the context of development.

4.2 Introduction

Word knowledge is essential for efficient language comprehension and has widespread ramifications for academic achievement (Spencer, Clegg, Stackhouse, & Rush, 2016), particularly literacy (e.g., Braze, Tabor, Shankweiler, & Mencl, 2007; Catts, Adlof, & Weismer, 2006). The ability to learn new words is highly variable across individuals: an 8-year-old child in the highest quartile of vocabulary ability already knows over 3000 more words than a child in the lowest quartile (Biemiller & Slonim, 2001), and this performance gap persists or even broadens over time (Biemiller, 2003; Cain & Oakhill, 2011). Yet, the mechanisms that underlie this broadening variability are poorly understood. Taking a developmental perspective, this study strives to better understand the mechanisms by which prior vocabulary knowledge may impact further word learning in children and adults.

4.2.1 Matthew effects in vocabulary acquisition

The importance of a child's existing vocabulary ability in contributing to further word learning has long been acknowledged: the well-cited *Matthew effect* (Stanovich, 1986) describes how the "rich" get "richer" in literacy skills. Stanovich proposed that this broadening skill gap is perpetuated by differences in literacy exposure: children with good language skills enjoy reading more, engage in more literacy activities, and encounter more new words in doing so. Indeed, comprehension skill *and* reading experience have been shown to predict vocabulary growth (Cain & Oakhill, 2011), and are argued to be fundamental to literacy development (Nation, 2017). From this perspective, accelerated rates in vocabulary acquisition for children with good vocabulary skills are due to their increasing engagement with texts.

However, Matthew effects in word learning have also been demonstrated in a number of experimental settings where exposure levels are controlled (e.g., Cain, Oakhill, & Lemmon, 2004; Wilkinson & Houston-Price, 2013). For example, Penno, Wilkinson, and Moore (2002) showed that children with better vocabulary ability learned more words from listening to stories than children of lower vocabulary, and these differences were sustained even in conditions that included direct word teaching. That is, even when children with lower vocabulary ability are given the same learning opportunities, they continue to show differences in new word acquisition. These findings implicate learning mechanisms or processes as a source of individual differences in word learning. If so, then what do children with better vocabulary bring differently to the task of word learning?

4.2.2 A Complementary Learning Systems approach to understanding Matthew effects

The present study set out to test one (not mutually exclusive) alternative to the literacy exposure hypothesis as an account of vocabulary Matthew effects. With reference to neurocognitive theories of memory, James et al. (2017) proposed that existing vocabulary knowledge might act as a "language schema" that speeds the acquisition and integration of new words. It was predicted that a child with good vocabulary knowledge to support these intrinsic processes would consolidate new words more rapidly than a child with poorer vocabulary, leading to a cumulative benefit in language development.

This account draws upon the Complementary Learning Systems (CLS) model of memory (McClelland et al., 1995) which Davis and Gaskell (2009) proposed as a useful framework for understanding lexical consolidation. In this context, the CLS model posits two interacting systems for learning new words. An encounter with an unfamiliar word forms a new distinct representation in memory that is initially dependent on hippocampal mechanisms (e.g., Warren & Duff, 2014). Over time, reactivation of this representation enables it to become gradually integrated with existing vocabulary knowledge in the neocortex, decreasing hippocampal dependence (Davis et al., 2009). This reactivation process can occur "offline", and a number of studies have demonstrated that sleep (versus wake) can strengthen and integrate a new word with existing knowledge in adults (Dumay & Gaskell, 2007) and children (Henderson et al., 2012). In both age groups, memory improvement is associated with slow-wave sleep (SWS) duration (Smith et al., 2017; Tamminen et al., 2010): the sleep stage characterised by slow neural oscillations which are argued to reflect systems communication in memory replay (Diekelmann & Born, 2010). Thus, different factors may support the initial encoding and longer-term storage of newly learned words, making it important to assess word recall immediately and after opportunities for offline consolidation in studies of vocabulary acquisition.

Recent domain-general CLS accounts have considered that prior knowledge may contribute to initial learning and/or consolidation (Kumaran et al., 2016; McClelland, 2013). In line with studies showing enhanced acquisition of schema-consistent information (e.g., Tse et al., 2007), it has been argued that new information consistent with existing knowledge can undergo faster consolidation. However, the underlying mechanisms are not well understood. One possibility is that schematic knowledge can advance neocortical learning of related information, reducing the need for hippocampal replay to occur offline (Kumaran et al., 2016). By this cortical learning account, individuals with more prior knowledge should benefit immediately when learning information that can capitalise upon it. Alternatively, the neural connections formed between new and existing memory representations during learning may facilitate offline consolidation itself: the information overlap to abstract (iOtA) model proposes that these shared connections cause co-activation of new and existing representations during sleep, enabling integration to happen more efficiently than when prior knowledge connections are more limited (Lewis & Durrant, 2011). By this account, individuals with more prior knowledge should benefit more from their richer connections during offline consolidation.

Therefore both the cortical learning and iOtA interpretations of the CLS account assume that related prior knowledge is helpful, but with one emphasising an advantage in initial encoding and the other proposing that the advantage is strongest during the consolidation process. Returning to questions of whether "language schema" might similarly facilitate word learning, we proceed to discuss two ways to conceptualise the relationship between a new word and prior lexical knowledge: the first emphasising the global properties of an individual (i.e., the size and richness of their vocabulary), and the second emphasising more local properties of the word (i.e., the similarity between a new word and existing word neighbours).

4.2.3 Global associations between vocabulary knowledge and word learning

Evidence for prior knowledge contributions to word learning comes from analyses of individual differences: several developmental studies have shown a positive correlation between vocabulary ability and memory for new words measured immediately after learning (e.g., Penno et al., 2002) and over a period of consolidation (e.g., Horváth et al., 2015b). Henderson et al. (2015) found that children with better expressive vocabulary ability showed greater overnight improvements in word-form recall than those with poorer vocabulary, even when controlling for differences in immediate performance. Consistent with the iOtA model (Lewis & Durrant, 2011), recent studies suggest that this prior vocabulary knowledge might be particularly important for supporting the offline integration of overlapping memory traces: for new words learned across multiple story contexts (Henderson & James, 2018), and for their integration with existing word knowledge (Henderson et al., 2015; James et al., 2017). Together, these studies suggest that prior global vocabulary knowledge offers support in consolidating new words.

4.2.4 Local associations between vocabulary knowledge and word learning

Studies that have examined the global associations between general vocabulary ability and word learning cannot elucidate causal mechanisms. Does the association between new word learning and existing vocabulary ability simply arise because good word learners have the skills that have built them a more extensive vocabulary, or does this existing vocabulary knowledge actively support new word acquisition? We address this question by manipulating the local word-form connections between particular new words and real words that may be present in an individual's language schema. If existing knowledge *actively* supports acquisition and consolidation then new words that overlap with real words should be better acquired and consolidated than words that do not.

Previous studies have manipulated this local overlap by training participants on pseudowords that varied in the number of existing word-form neighbours: real words that could be created by changing a single letter/phoneme. A number of these studies have shown that pseudowords with more phonological neighbours are recalled better in picture-naming tasks than those with fewer neighbours, for pre-schoolers (Hoover, Storkel, & Hogan, 2010) and adults (Storkel et al., 2006). This neighbour benefit also appears to be related to pre-schoolers' expressive vocabulary (e.g., Storkel & Hoover, 2011), supporting the utility of this paradigm for addressing individual differences in prior knowledge. In other words, the benefit of local neighbours to the acquisition process will only be obtained if those neighbours are known to the individual, and the likelihood of knowing the neighbours is predicted by global vocabulary measures.

4.2.5 Developmental differences in prior knowledge contributions to word learning

A final, broader approach to assessing prior knowledge contributions to word learning is to compare adults and children: whilst both groups *can* benefit from prior knowledge, adults will typically have a larger body of prior knowledge to support language acquisition. However, children may receive greater benefit from offline consolidation, which could facilitate language acquisition despite often receiving less global support from prior knowledge (Wilhelm et al., 2012). This proposal stems from evidence of sleep architectural changes across development: children show larger proportions of SWS (Ohayon et al., 2004; Wilhelm et al., 2013) that are tightly linked to ongoing neural reorganisation (Feinberg & Campbell, 2010). In comparing children and adults, two clear predictions can be made to isolate the contribution of prior knowledge: adults will show larger and/or more robust effects of local prior knowledge, given that they should know more local word neighbours; and, children will show larger overnight consolidation effects than adults under conditions of limited local knowledge connections (James et al., 2017; Wilhelm et al., 2012).

4.2.6 The current study

Our study manipulated the availability of local word-form neighbours in explicit word learning, extending existing findings in three important ways. First, we examined the longevity of neighbour effects: we taught children (Experiments 1, 3) and adults (Experiment 2) novel words that systematically varied in the number of word-form neighbours, and tested word recall immediately after training, the next day, and one week later. Very few studies have assessed the longer-term benefit of word-form neighbours (although see Hoover et al., 2010), and none to our knowledge have carried out a comprehensive assessment of *when* during the learning and consolidation process a novel word might place demands on connections to existing vocabulary. Studies by Storkel and colleagues suggest that this benefit might be apparent immediately, but studies relating overnight changes in word learning to global vocabulary suggest there may be a further benefit during consolidation. We continued to track memory performance a week later, given that knowledge-related differences can emerge with more prolonged periods of consolidation than a single night (e.g., Henderson et al., 2013c).

Second, by examining individual differences in the benefit of word-form neighbours for word learning, we aimed to further understand the relationship between individuals' global vocabulary knowledge and their ability to acquire and consolidate new words. Crucially, if this relationship is due to general differences in word-learning skill (i.e., good word learners *acquire* a better vocabulary), then we would expect to see this association between vocabulary ability and word learning performance regardless of a novel word's neighbours (note that we use "vocabulary ability" to refer to performance on standardised assessments of vocabulary). However, if existing vocabulary *actively* supports consolidation processes, in accordance with a CLS approach to Matthew effects (James et al., 2017), then we might expect participants with good vocabulary ability to show a stronger benefit for novel words with neighbours compared to novel words that do not have close neighbours, under the assumption that more of the neighbours will exist in their lexicon.

Third, in testing adults and children, we examined how the contributions of prior knowledge and offline consolidation might differ across development. Whilst consolidation effects are anticipated in both age groups, children's higher proportions of SWS compared to adults might lead us to expect greater improvements in novel word recall at subsequent time points in our experiments with children (Experiments 1, 3) regardless of our word neighbour manipulation. Adults, on the other hand, have superior linguistic knowledge than children and are proposed to more readily access this knowledge during learning. As such, adults may show larger and more persistent benefits of word-form neighbours on learning novel words.

4.3 Experiment 1

4.3.1 Experiment 1 Hypotheses

Three hypotheses were pre-registered at https://osf.io/fnu6c: 1) There will be a positive correlation between vocabulary ability and the overnight improvement in word memory; 2) Novel words with many word-form neighbours will be recalled more easily than novel words with no/few word-form neighbours, and this benefit could arise immediately and/or after opportunities for consolidation; and 3) Children with better vocabulary ability will experience a greater benefit from word-form neighbours than children with poorer vocabulary.

Experiments were approved by the University of York Psychology Ethics Committee.

4.3.2 Experiment 1 Methods

Participants

Ten Year 3 and 4 classes from three North Yorkshire schools participated. Two children were excluded for reported learning disabilities, and three had low levels of English that prevented participation. A further 22 datasets could not be analysed due to the child's absence during the vocabulary assessment. The resulting sample included 232 children (124 males) aged 7;03-9;03 years old (*M* = 8;03). This age group maximised comparability with previous studies showing overnight improvements in word learning and associations with vocabulary ability (Henderson et al., 2015). The large sample size allowed screening for poor comprehenders for a future study and was considered appropriate to compensate for the increased noise in the data while using whole-class testing procedures.

Design and procedure

Children participated in three 60-90 minute whole-class sessions, which incorporated word-learning measures and cognitive tests. On Day 1, children learned 16 fictitious words with no or many orthographic neighbours, and completed the memory tests for the new words (T1). Memory for the words was tested again the next day (T2) and one week later (T3). The timing of the sessions was constrained to the school day (9am-3pm), and each of the three sessions were scheduled at a similar time of day for each class.

During these sessions, children also took part in shortened versions of standardised tests adapted for whole-class administration: vocabulary ability, alongside spelling, nonverbal IQ and listening comprehension. The latter (unreported) measures were included for identifying poor comprehenders for a subsequent study.

Experimental stimuli

Sixteen pseudowords were selected from the English Lexicon Project (Balota et al., 2007) for having no orthographic neighbours (e.g., *peflin*, for which substituting any individual letter cannot create a known English word) or many orthographic neighbours (e.g., *ballow,* for which letters can be substituted to create multiple known words, including *bellow, wallow, ballot,* etc.; Appendix B1). Pseudowords with no orthographic neighbours also had significantly lower phonological neighbourhood density and phonotactic probability (Marian, Bartolotti, Chabal, & Shook, 2012). All pseudowords were bisyllabic, 5-6 letters, and began with a single consonant and vowel. The two lists were matched for number of phonemes and letters, as well as bigram probability (Table 3). We trained orthographic forms to enable testing for new word memory in a group setting, supported by spoken word presentations to reduce differences attributed to reading ability.

Although our primary research question related to word form learning, the purpose of new vocabulary is to convey meaning. As such, the pseudowords were paired with novel objects to provide a basic semantic component. Two sets of eight novel objects were selected from the NOUN database (Horst & Hout, 2016). The assignment of each set to the word neighbour condition was counterbalanced across classes.

Table 3. Properties of stimuli in each of the word-form neighbour conditions

	Orthographic neighbours[a]	Phonological neighbours[b]	Length (letters)	Length (phonemes)	Bigram frequency[a]	Biphone probability[b]
All experiments						
None	0	0.38	5.63	5	1369	0.003
Many	8.38	10.13	5.63	4.63	2054	0.006
Experiments 2&3						
One	1	0.88	5.63	5.25	1224	0.005
p-value	<.001	<.001	1	.130	.106	.035

Note. Mean values were computed from a) English Lexicon Project (Balota et al., 2007), and b) CLEARPOND (Marian et al., 2012). There were 8 items in each condition, although only a subset of 6 were used for Experiment 3. Patterns of significance were identical for the stimuli used in each experiment.

Novel word training

The training tasks were set in the context of a discovery adventure of two popular film characters. In a fixed random order, the class was presented with each word-form on screen and a recording of its pronunciation, and repeated the word aloud. After two rounds of form-only repetition, a further two rounds also presented the novel object on screen. As well as repeating the word aloud, children had to find each object in training booklets and write its name, enabling practice with the orthographic form.

The final rounds of training consisted of a multiple-choice quiz. In the first round, children were presented with an object and asked to select which of three words was its name. The second round presented a word with three objects to choose from. Children circled their answers in booklets, with the correct answer presented on screen afterwards.

Novel word tests

Cued form recall. A stem completion task assessed explicit recall of the new forms. Children were given the first consonant and vowel of each word in written form (e.g., *ba* for *ballow*), and heard the cue recorded by the same speaker as in training. Children were asked to write the rest of the word in their test booklets, and encouraged to attempt answers even if unsure. Test items were presented in a fixed random order, re-randomised for each time point. To minimise confounds of spelling ability, children could ask for help spelling words, and answers were scored correct if they were phonologically accurate (e.g., *balloe, balo*).

Recognition. A four-alternative-forced-choice task assessed familiarity with the word-forms and semantic mappings. Children were presented with each object, and asked to choose its name from: the correct answer, a phonological foil for the correct answer (vowel change; e.g., *ballew*), an incorrect learned novel word, plus its matched phonological foil. Children heard recordings of the four options, alongside the written form on screen, and circled answers in test booklets. Test items were re-randomised at each time point, but each item's answer options remained consistent across sessions.

Spelling. To identify children whose performance in cued-recall may reflect spelling difficulty rather than word learning, a spelling test for the novel words was administered at the end of T3. Each item was read aloud for spelling. Items that would have been scored as incorrect according to the cued-recall scoring principles were excluded from the cued-recall analysis on a by-participant basis (e.g., if a child spelled *ballow* as *blowe,* this item was treated as missing data). These items were excluded across all cued-recall test points regardless of performance, given that incorrect answers would have been impossible to interpret in the context of unreliable spelling (i.e., not remembered vs. not spelled correctly).

Vocabulary ability

Measures of expressive and receptive vocabulary were administered, but neither provided a stronger correlate of overall task performance (see Appendix B2). Therefore, we used the expressive task as a measure of global vocabulary ability during analysis for consistency with previous studies highlighting relationships between vocabulary and word learning (e.g., Henderson et al., 2015; Storkel & Hoover, 2011).

Expressive vocabulary. Children were asked to provide written definitions for a subset of 11 age-appropriate items from the British Ability Scales-II Word Definitions task (Elliot, Smith, & McCulloch, 1997). Children heard each item read aloud, and were asked to write down its meaning. An example was provided at the start. Because this method of administration could not prompt children for further detail (as is standard to oral administration), a bespoke scoring system was developed that enabled item scores of 0 (incorrect), 1 (borderline/vague), or 2 (correct), summed for an overall score.

Analyses

Analyses used R (R Core Team, 2015), with graphs using *yarrr* (Phillips, 2017) and *ggplot2* (Wickham, 2009). For each measure of word learning, we used *lme4* (Bates, Maechler, Bolker, & Walker, 2014) to fit a mixed-effects binomial regression model to the data with fixed effects of session, neighbourhood condition, vocabulary ability, and all interactions between them. Two orthogonal contrasts were set for the three-level factor of session: *delay1* contrasted words with or without the opportunity for consolidation (T1 vs. T2&T3), *delay2* contrasted performance at T2 vs. T3. Vocabulary score was scaled and centred before entering into the model.

For all experiments, we had pre-registered an initial attempt at maximal random effects structures, but these frequently suffered convergence issues and required model simplification. We therefore pruned higher-order interactions from fixed-effects where not contributing to the model to allow better-specified random-effects structures, using pairwise likelihood ratio tests to confirm that simplified models were not significantly poorer in their fit to the data. We then used a forward-best-path approach (Barr et al., 2013) to test between simple and progressively complex random-effects structures, retaining only random-slopes that improved model fit according to a liberal criterion, $p < .2$ (Barr et al., 2013; Bates et al., 2015a).

4.3.3 Experiment 1 Results[6]

Cued form recall

Five participants were excluded on the basis of unintelligible handwriting on the novel spellings test. A further 5.92% of the remaining data was excluded on a by-item basis for individual participants, where poor spelling during the novel spelling test rendered the data uninterpretable. There was no evidence for a three-way interaction between session, neighbour condition, and vocabulary ability, and this was pruned from the final model (Table 4) with no reduction in model fit ($\chi^2 = 0.198$, $p =$

[6] Note that the original pre-registration specified an additional subgroup analysis on children identified as poor comprehenders. However, our individual follow-up assessments with these children indicated that very few met traditional criteria for specific comprehension difficulty in this sample (*n* = 3/254). As such, it was deemed inappropriate to analyse and interpret these as a distinct group.

.91). Recall was significantly better at later sessions than T1 (*delay1* contrast), and also improved between T2 and T3 (*delay2*). The presence of word neighbours did not affect cued-recall performance overall (*neighb*), but did in interaction with test session (*delay1:neighb*): the negative coefficients show that the benefit of word neighbours was larger at T1 compared to subsequent tests (Figure 3). There was no further reduction in neighbour benefit between T2 and T3 (*delay2:neighb*) which, in the context of no overall neighbour effect, indicates that the neighbour benefit was only present at T1.

Vocabulary positively predicted cued-recall performance (*vocab*; Figure 4), but contrary to hypotheses did not interact with improvements over time or word-neighbours.

Table 4. Predictors of cued recall performance in Form Neighbour Experiment 1

Fixed effects	*b*	*SE*	*z*	*p*	
(Intercept)	-0.87	0.27	-3.20	.001	
delay1	**0.27**	**0.04**	**6.63**	**<.001**	
delay2	**0.42**	**0.04**	**11.13**	**<.001**	
neighb	0.25	0.26	0.96	.337	
vocab	**0.50**	**0.09**	**5.75**	**<.001**	
delay1:neighb	**-0.11**	**0.04**	**-2.82**	**.005**	
delay2:neighb	-0.03	0.04	-0.71	.480	
delay1:vocab	0.02	0.02	1.07	.283	
delay2:vocab	-0.04	0.03	-1.24	.215	
neighb:vocab	0.05	0.04	1.25	.211	
Random effects	**Variance**	***SD***	**Correlations**		
participant: (intercept)	1.47	1.21			
participant:(slope) neighb	0.11	0.33	0.23		
item: (intercept)	1.06	1.03			
item: (slope) delay1	0.02	0.14	0.36		
item: (slope) delay2	0.01	0.08	-0.28	0.31	
item: (slope) vocab	0.01	0.10	0.28	0.61	0.83

Note. Model formed from 10,135 observations, collected from 227 participants across 16 items. Orthogonal contrasts were used for the three-level factor of session: delay1 (Session 1 vs. Sessions 2&3), delay2 (Session 2 vs. Session 3).

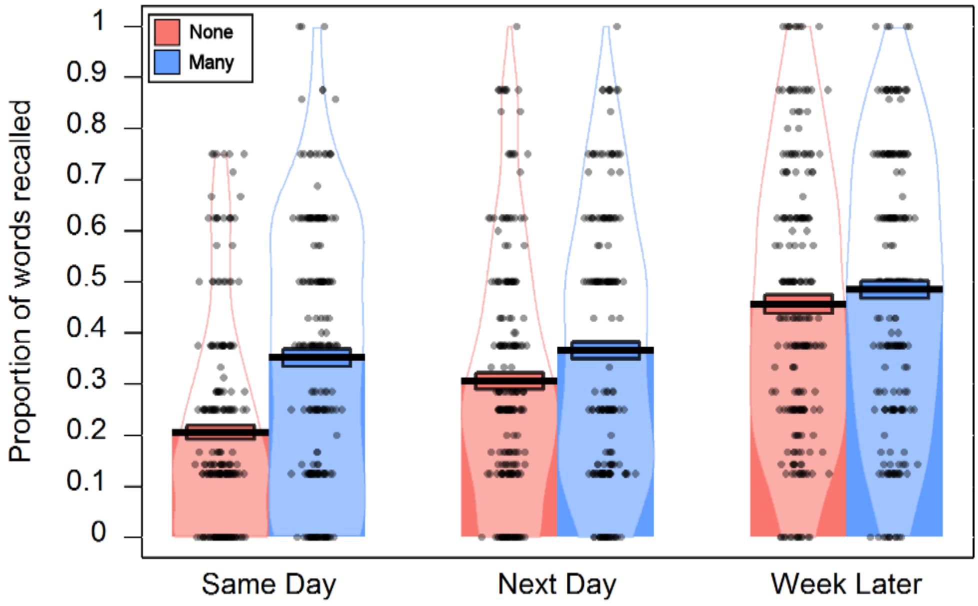


Figure 3. *RDI plot of children's cued form recall performance in Form Neighbour Experiment 1, plotted by neighbour condition and test session. The dark coloured bars can be interpreted as traditional bar charts, with the outlined areas representing smoothed distribution curves. Thick black horizontal bars represent the mean for each condition, and surrounding boxes mark +/-1 standard error of the mean. Black dots indicate by-participant means for each condition.*

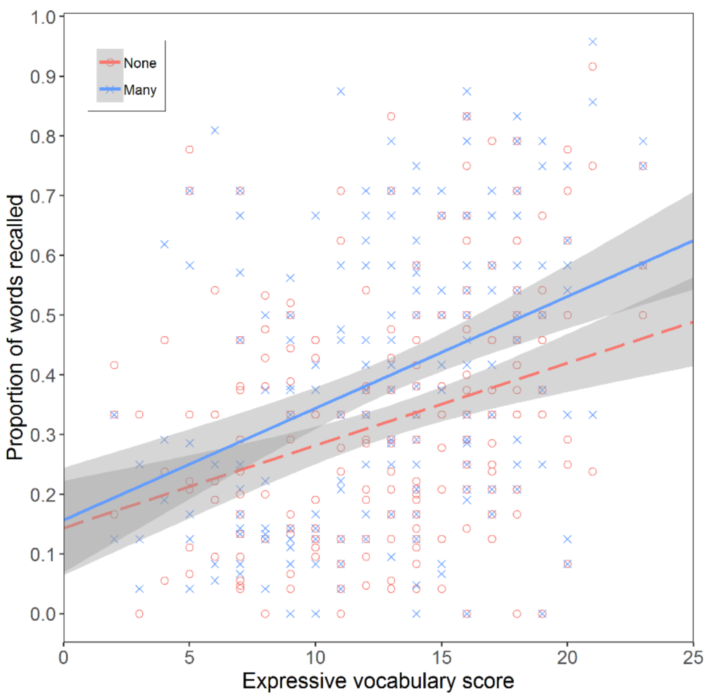


Figure 4. *Scatterplot of mean proportion of words recalled for each word neighbour condition in Experiment 1 (collapsed across test sessions), plotted against children's expressive vocabulary score. Grey shade areas represent 95% confidence intervals.*

Recognition

Performance was good across all conditions (*M* = .80, *SD* = .18), with a slight decline between T2 and T3 that was not significant (b = -0.08, Z = -1.73, p = .08). Only vocabulary emerged as a significant predictor (b = 0.79, Z = 8.54, $p < .001$). Of limited theoretical interest, the full model and figure are presented in Appendices B3-B4.

4.3.4 Experiment 1 Discussion

Children aged 7-9 years became familiar with the new words very quickly (with recognition at ~80%), but word neighbours and a period of consolidation facilitated the acquisition of higher quality lexical representations as reflected by superior production in the recall task. Children benefited from existing neighbours during novel word retrieval immediately after initial learning, consistent with our hypotheses and previous findings (Hoover et al., 2010; Storkel et al., 2006). However, this neighbour benefit diminished at the 24-hour test, leaving no overall benefit of word neighbours. We therefore propose that it is the new words without local connections to prior knowledge that subsequently show greater strengthening from offline consolidation processes (see van Kesteren et al., 2013, for a similar interpretation). As such, learning in the context of this paradigm is more in line with a cortical learning approach to prior knowledge, and does not support the iOtA model.

As with previous studies (e.g., Penno et al., 2002), global vocabulary ability predicted word learning across both tasks. However, the results did not support our predicted relationship between vocabulary ability and overnight improvements in performance, as has been found in numerous previous studies (e.g., Henderson et al., 2015). In many respects this finding is consistent with our theoretical approach: if prior knowledge benefits are apparent immediately and *weaken* during consolidation, then we would no longer predict existing vocabulary knowledge to support the overnight improvements. However, it still limits our ability to draw conclusions regarding the ways in which global vocabulary knowledge might support offline consolidation of new words in relation to earlier studies.

The results also failed to support the hypothesis that those with good vocabulary ability would show bigger benefits of word neighbours, which is somewhat puzzling considering the clear benefit for local neighbour connections themselves. One could interpret this lack of interaction as evidence in support of the general skill account: better word learners simply learn more words to acquire a better vocabulary. However, given the clear benefit for knowing some word neighbours, we consider a number of explanations for the lack of individual differences in this benefit. First, the group training and testing nature of Experiment 1 introduces a significant amount of noise into the data compared to previous studies. Second, whilst we took care to minimise the impact of spelling on recall performance (i.e., providing spelling help, removing problematic items during analysis), children's ability to produce the written words may have been constrained by their writing and spelling ability. These individual differences in orthographic knowledge may have made it more challenging to identify individual differences related specifically to vocabulary ability.

We also consider a third—more theoretically interesting—account of our lack of neighbourhood interaction with vocabulary ability: that the number of neighbours is crucial. Computational models of visual word recognition have suggested that one neighbour influences word processing, but that there is little impact of additional neighbours (Davis & Andrews, 1996, as cited in Bowers, Davis & Hanley, 2005). If this primary benefit for one neighbour is also true during learning, we may have maximised the potential for *all* children to have known and activated *at least one* neighbour during training by using stimuli that had many possible neighbours to benefit from. For example, one child might access *ballow*'s neighbour *bellow* whilst another might access *wallow*, but with nothing further to be gained from accessing both. Indeed, previous studies demonstrating a relationship between global vocabulary ability and overnight consolidation have trained novel words related to a single existing word in order to study lexical integration (e.g., *dolpheg* derived from *dolphin*, James et al., 2017). We therefore added a 'one-neighbour' condition to subsequent experiments to explore whether this condition is as beneficial as having many neighbours, first in adults (Experiment 2) and then children (Experiment 3). Importantly, we asked whether the one-neighbour condition would be more sensitive to individual differences in learning and consolidation.

4.4 Experiment 2

4.4.1 Experiment 2 Hypotheses

The pre-registered hypotheses (https://osf.io/tm538) were: 1) Vocabulary ability will be an overall predictor of word-learning ability in adults (as for children in Experiment 1); 2) Memory for new words will improve with opportunities for consolidation, and—consistent with children in Experiment 1—initial overnight improvement will be larger for words without neighbours; 3) Novel words with only one neighbour will benefit from this prior knowledge compared to words without neighbours; and 4) Existing vocabulary ability will most strongly predict performance in the one-neighbour condition, under the assumption that the most critical benefits arise from activating *at least one* neighbour and that this lower end of the scale will be more sensitive to individual differences in existing vocabulary. Importantly, Experiment 2 also provides the opportunity to draw developmental comparisons with Experiment 1, under the assumptions that adults have greater prior knowledge that might be more readily activated during learning, but that children benefit more from offline consolidation.

4.4.2 Experiment 2 Methods

Participants

Seventy-nine adults participated (15 male), aged 18–35 years (*M* = 20;02). This smaller sample size was appropriate given the reduced noise in this dataset: adults show better compliance during group training, have tighter phonological-orthographic mappings, and were tested individually. Participants were recruited via lecture advertisements or participant database, and were native monolingual English speakers, with normal/corrected-to-normal vision and hearing. Note that although the gender balance did not match Experiments 1 and 3, gender did not predict recall performance alone (p = .22), nor in interaction with time or neighbour (all ps > .1)

Design and procedure

Participants learned novel words in a 30-minute group training session in an IT suite (scheduled between 10am and 4pm). They then completed the three test sessions independently via an online web link, scheduled as before. Participants were asked to complete the tests at a similar time each day, but we retained data from all sessions completed on the correct day. Mean hour of test remained highly similar across all three sessions (T1: *M* = 2.12pm, *SD* = 3.42 hours; T2: *M* = 2.41pm, *SD* = 3.31 hours; T3: *M* = 1.53pm, *SD* = 4.39 hours). An additional online session (completed at any time over the week) collected background and vocabulary information.

Experimental stimuli

Twenty-four novel words were trained from three conditions. The no- and many-neighbour conditions were identical to Experiment 1, but a third set of eight words with only one orthographic/phonological neighbour was created, and matched to the other conditions on length and bigram frequency (Table 3).

A third set of novel objects was selected from the NOUN database (Horst & Hout, 2016). The assignment of each set to each word-neighbour condition was altered across two counterbalancing conditions, such that each set of objects appeared in two of the three conditions across participants.

Novel word training

As Experiment 1, with the exception that participants labelled items and submitted their multiple-choice answers via a web browser, consistent with the testing format.

Novel word tests

All three test sessions exploited an online survey platform (Qualtrics, Provo, UT). We retained the written test format, given that adults have much tighter spoken-written language mappings (Samuels & Flor, 1997), reducing variability in orthographic support.

Cued form recall. As Experiment 1, except participants were instructed to click a speaker to hear the cue presented through speakers/headphones (unrestricted), and provided typed responses. Item order was fully randomised.

Recognition. As Experiment 1. Item order was randomised, and participants clicked an icon to hear each item spoken aloud.

Vocabulary ability

Participants provided typed definitions for 13 age-appropriate written items selected from WASI-II Vocabulary (Wechsler, 2011), adapted for online administration (for the receptive vocabulary measure see Appendix B5). An example was provided. Answers were scored as 0, 1, or 2, according to manual guidelines.

Analyses

As Experiment 1. The additional neighbour condition enabled us to test the following orthogonal contrasts: *neighb1* was set to compare the presence versus absence of neighbours (no vs. one+many), and *neighb2* contrasted one versus many neighbours[7].

4.4.3 Experiment 2 Results

Cued form recall

All fixed-effects were retained in the final model (Table 5). As in Experiment 1, individuals with better vocabulary ability performed better on the cued recall task (*vocab*), and overall levels of performance improved after a period of consolidation (*delay1*). However, for adults there was no further improvement between T2 and T3 (*delay2*). The presence of one/more word neighbours did not significantly affect recall across the week (*neighb1*: $p = .053$), and nor did its influence change with test session (*delay1:neighb1*: $p = .06$; all other ps > .75). That is, there was weak statistical evidence for a benefit of word-neighbour connections, and for the prioritisation of no-neighbour items in consolidation (Figure 5).

Consistent with our prediction that only one neighbour is needed to support learning, there was no overall difference between recall performance in the one- and many-neighbour conditions (*neighb2*: $p = .52$). However, the inclusion of this manipulation enabled us to identify individual differences in neighbour benefit related to vocabulary ability (in support of an active role for prior knowledge in word learning): there was a significant interaction between neighbour condition and vocabulary ability (*neighb2:vocab*). As depicted by Figure 6, there was a stronger association between vocabulary ability and performance in the one-neighbour condition compared to the many-neighbour condition, with only participants with poorer vocabulary showing a difference between these conditions. However, as in Experiment 1, vocabulary ability did not predict differences in recall performance for words with/without neighbours overall (*neighb1:vocab*). Thus, the weaker association between vocabulary ability and performance in the many-neighbour than one-neighbour condition may result from the increased chance that all participants know at least one neighbour.

Vocabulary ability was related to the change in adults' recall performance between T2 and T3 (*delay2:vocab*: $p = .038$). Adults with good vocabulary ability

[7] By setting orthogonal contrasts, analyses were deemed to be highly comparable but more informative than the treatment contrasts initially pre-registered for Experiment 2, removing the need for follow-up comparisons to fully interpret the model.

showed a slight benefit in retention of their new word knowledge across the week compared to adults with poorer vocabulary. However, vocabulary ability did not interact with any neighbourhood effects on retention.

Table 5. Predictors of cued recall performance in Form Neighbour Experiment 2

Fixed effects	*b*	*SE*	*z*	*p*		
(Intercept)	-0.89	0.20	-4.36	<.001		
delay1	**0.29**	**0.04**	**7.22**	**<.001**		
delay2	0.08	0.04	1.77	.077		
neighb1	0.22	0.11	1.93	.053		
neighb2	0.12	0.19	0.64	.522		
vocab	**0.40**	**0.14**	**2.80**	**.005**		
delay1:neighb1	-0.04	0.02	-1.88	.060		
delay2:neighb1	0.00	0.03	-0.07	.947		
delay1:neighb2	-0.01	0.03	-0.31	.759		
delay2:neighb2	-0.01	0.05	-0.17	.866		
delay1:vocab	0.01	0.04	0.31	.760		
delay2:vocab	**0.09**	**0.04**	**2.07**	**.038**		
neighb1:vocab	-0.02	0.04	-0.42	.677		
neighb2:vocab	**-0.12**	**0.06**	**-2.16**	**.031**		
delay1:neighb1:vocab	0.01	0.02	0.77	.444		
delay2:neighb1:vocab	0.02	0.03	0.52	.605		
delay1:neighb2:vocab	-0.02	0.03	-0.72	.474		
delay2:neighb2:vocab	0.07	0.05	1.42	.157		
Random effects	Variance	*SD*	Correlations			
participant: (intercept)	1.41	1.19				
participant: (slope) delay1	0.07	0.26	0.09			
participant: (slope) delay2	0.02	0.14	0.02	0.98		
participant: (slope) neighb1	0.06	0.24	-0.29	0.16	0.02	
participant: (slope) neighb2	0.07	0.26	0.25	-0.13	-0.11	-0.04
item: (intercept)	0.54	0.73				
item: (slope) vocab	0.01	0.09	0.83			

Note. Model formed from 5,640 observations, collected from 79 participants across 24 items. Orthogonal contrasts were used for three-level factors: delay1 (Session 1 vs. Sessions 2&3), delay2 (Session 2 vs. Session 3), neighb1 (no vs. one&many), neighb2 (one vs. many).

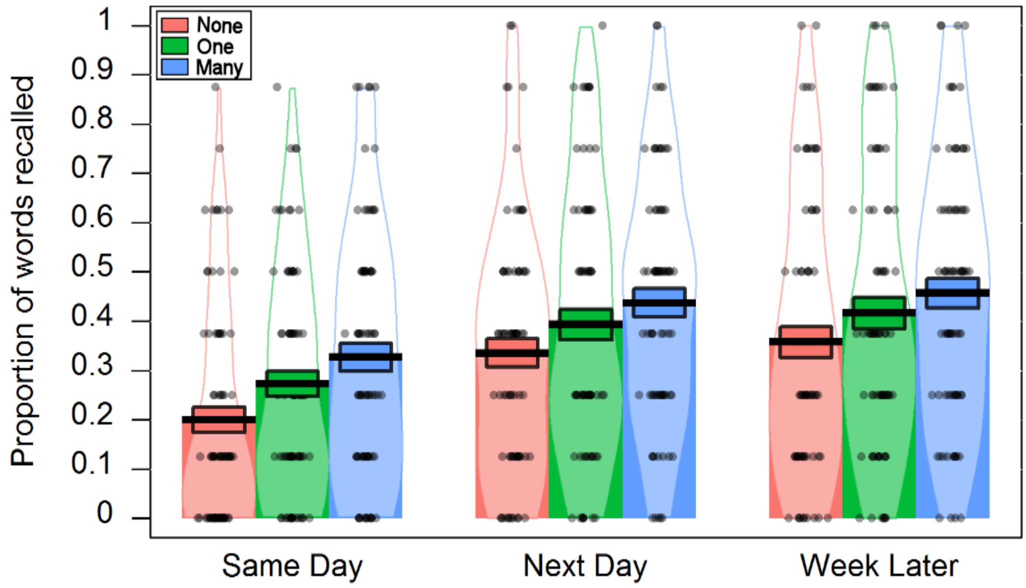


Figure 5. *RDI plot of adults' cued form recall performance in Form Neighbour Experiment 2, plotted by neighbour condition and test session. Thick horizontal bars represent the mean for each condition, and surrounding boxes mark +/- standard error of the mean.*

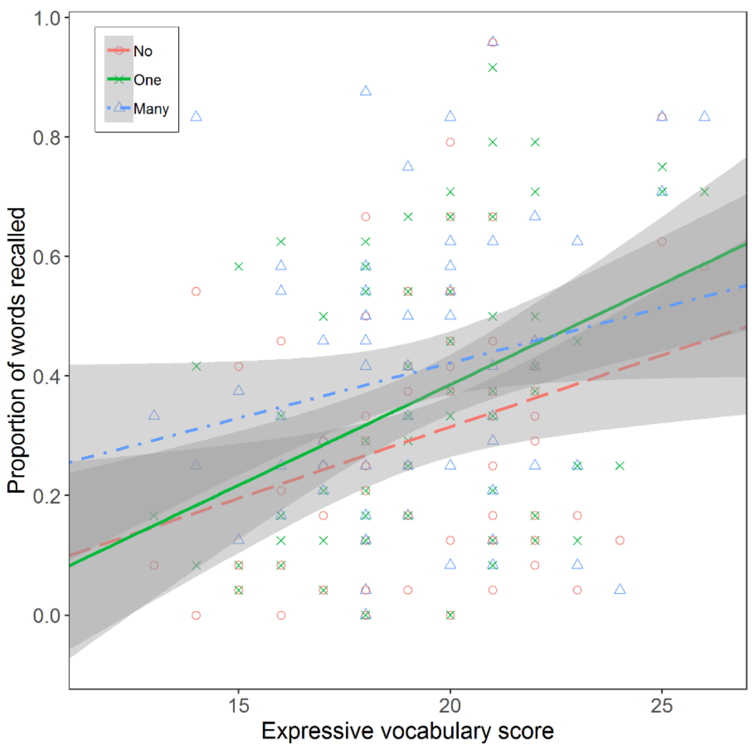


Figure 6. *Scatterplot of mean proportion of words recalled for each word neighbour condition in Experiment 2 (collapsed across test sessions), plotted against adults' expressive vocabulary score. Grey shaded areas represent 95% confidence intervals.*

Recognition

Recognition performance was very high (M = .90, SD = .1) and, as with Experiment 1, only vocabulary ability significantly predicted performance (b = 0.37, Z = 2.52, p = .012) (Appendix B6-B7).

4.4.4 Experiment 2 Discussion

Adults, like children, improved in their explicit recall of new words after opportunities for consolidation, and global vocabulary ability was a strong predictor of word learning overall. Crucially, by including stimuli with only one neighbour, we demonstrated that global vocabulary can *actively* support new word acquisition: adults with good vocabulary showed a comparable benefit for words with one and many neighbours, whereas adults with poorer vocabulary showed poorer performance for the words with more limited local overlap compared to many-neighbour words. As before, support from word neighbours, although statistically weak, was apparent immediately.

Unlike Experiment 1, the interaction between word neighbour condition and time did not reach significance. This weaker consolidation of no-neighbour novel items relative to the findings from children in Experiment 1 may result from introducing a third condition (with an increase of items from 16 to 24). However, this finding might reflect genuine developmental differences in the mechanisms supporting consolidation: whilst children have superior sleep-associated mechanisms to support the consolidation of novel information, adults are argued to retain greater dependence on prior knowledge across the course of consolidation (James et al., 2017; Wilhelm et al., 2012).

Experiment 3 sought to replicate the superior consolidation for no-neighbour items in children found in Experiment 1, alongside the introduction of the one-neighbour condition. We taught only spoken word-forms to remove the possibility that orthographic knowledge was constraining the identification of a relationship between vocabulary ability and overnight change in performance in Experiment 1.

4.5 Experiment 3

4.5.1 Experiment 3 Hypotheses

The hypotheses were pre-registered at https://osf.io/4abw3 as follows: 1) Again, vocabulary ability will be an overall predictor of word learning performance; 2) Memory for the new words will improve after opportunities for consolidation, and the improvement for no-neighbour words will be larger than for words with many neighbours; 3) Novel words with only one neighbour would benefit from this local prior knowledge in word learning compared to words with no neighbours, but this could either be apparent immediately or require consolidation to emerge; and 4) Vocabulary ability will show the strongest relationship with learning in the one-neighbour condition, emerging either immediately (as Experiment 2) or after opportunities for consolidation (as with previous developmental studies, e.g., Henderson et al., 2015).

4.5.2 Experiment 3 Methods

Participants

Four classes of Year 3/4 children from one school took part (adopting eligibility criteria from Experiment 2). 78 participants met these criteria, but a further six were excluded due to self-withdrawal (n = 1), inattention (n = 2), technical errors (n = 2) and teacher-reported speech and language difficulties that made testing unfeasible (n = 1). The final sample comprised 72 children (38 male) aged 7;06–10;05 (M = 8;08).

Design and procedure

Children participated in whole-class training in the morning, but completed the memory tests individually. All sessions took place within the school day (9am-3pm). After the memory tests or in a separate session, children completed a standardised assessment of expressive vocabulary. Assessments of nonverbal IQ, reading efficiency and reading comprehension were also administered for another study.

Experimental stimuli

A subset of 18/24 items were selected from Experiment 2, allowing six words in each neighbour condition. Given that only spoken word-forms were trained, these words were selected to ensure that the strict neighbour criteria withheld across phonological as well as orthographic neighbours. The assignment of each set of novel objects to each word-neighbour condition was altered across two counterbalancing conditions.

Novel word training

As Experiment 1, except that there was no written presentation or writing practice at any point. For multiple-choice tasks, children circled numbers corresponding to spoken answer options.

Novel word tests

Cued form recall. Children heard the cue through headphones, and spoke the remainder of the pseudoword. Item presentation was randomised, and the experimenter recorded responses on an answer sheet.

Recognition. The previous recognition tasks had four response options per trial, testing form and semantic specificity together. However, with the removal of orthographic support, this was deemed to be too demanding on working memory, and thus response options were reduced to two per trial. Because it was important to maintain the level of between-test exposure across experiments, two tests were created to test form and semantic recognition separately. Two practice trials with real words and pseudowords were administered, and trials timed out after 7 seconds.

Form-recognition. Children heard each item and its phonological foil through the headphones, and used a key press to respond whether they had learned the first or second item presented.

Form-picture recognition. Children heard two learned pseudowords through headphones, and used a key press to indicate which word was the name of the presented object.

Vocabulary ability

Only expressive vocabulary ability (Vocabulary subtest from the WASI-II; Wechsler, 2011) was measured, given well-established relationships with cued-recall in Experiments 1 and 2.

Analyses

As Experiment 2.

4.5.3 Experiment 3 Results

Cued form recall

The initial model provided no evidence of a three-way interaction between delay, neighbour condition, and vocabulary ability, which was therefore pruned with no reduction in model fit ($\chi^2 = 1.70$, $p = .79$). The final model is presented in Table 6.

There was a clear improvement in recall performance across all three test sessions (*delay1*; *delay2*). There was no overall benefit of word neighbours on recall performance across the week (*neighb1*), but there was an initial benefit that changed with consolidation (*delay2:neighb1*). This interaction was consistent with—although later to emerge than in—Experiment 1: there was a larger benefit of having at least one word neighbour (vs. no neighbours) at earlier test points (T2 benefit = .09), which diminished by T3 (benefit = .01, Figure 7).

Consistent with Experiment 2, only one neighbour mattered in influencing performance: there were no significant differences in recall between the one- and many-neighbour conditions overall (*neighb2*) or in interaction with the test session (*delay1:neighb2; delay2:neighb2*). However, in contrast to the adult study, this one-

Table 6. Predictors of cued recall performance in Form Neighbour Experiment 3.

Fixed effects	*b*	*SE*	*z*	*p*
(Intercept)	-2.36	0.29	-8.17	<.001
delay1	**0.55**	**0.07**	**8.00**	**<.001**
delay2	**0.60**	**0.08**	**7.61**	**<.001**
neighb1	0.32	0.19	1.71	.087
neighb2	0.07	0.32	0.22	.822
vocab	**0.36**	**0.13**	**2.69**	**.007**
delay1:neighb1	-0.08	0.05	-1.71	.087
delay2:neighb1	**-0.16**	**0.05**	**-3.00**	**.003**
delay1:neighb2	-0.08	0.07	-1.10	.271
delay2:neighb2	0.04	0.08	0.45	.653
delay1:vocab	0.05	0.04	1.18	.237
delay2:vocab	-0.07	0.06	-1.28	.202
neighb1:vocab	-0.04	0.05	-0.89	.372
neighb2:vocab	-0.05	0.08	-0.72	.473
Random effects	Variance	*SD*	Correlations	
participant: (intercept)	0.99	0.99		
participant: (slope) neighb1	0.03	0.17	-0.61	
participant: (slope) neighb2	0.13	0.37	-0.19	0.42
item: (intercept)	1.13	1.06		
item: (slope) delay1	0.03	0.16	-0.34	
item: (slope) delay2	0.03	0.18	-0.01	0.94

Note. Model formed from 3852 observations, collected from 72 participants across 18 items. Orthogonal contrasts were used for three-level factors: delay1 (Session 1 vs. Sessions 2&3), delay2 (Session 2 vs. Session 3), neighb1 (no vs. one&many), neighb2 (one vs. many).

neighbour condition was not more sensitive to individual differences relating to children's vocabulary ability (*neighb2:vocab*). Vocabulary ability was a significant predictor of cued-recall performance overall (*vocab*), but did not interact with neighbour benefit or changes across the week (Figure 8).

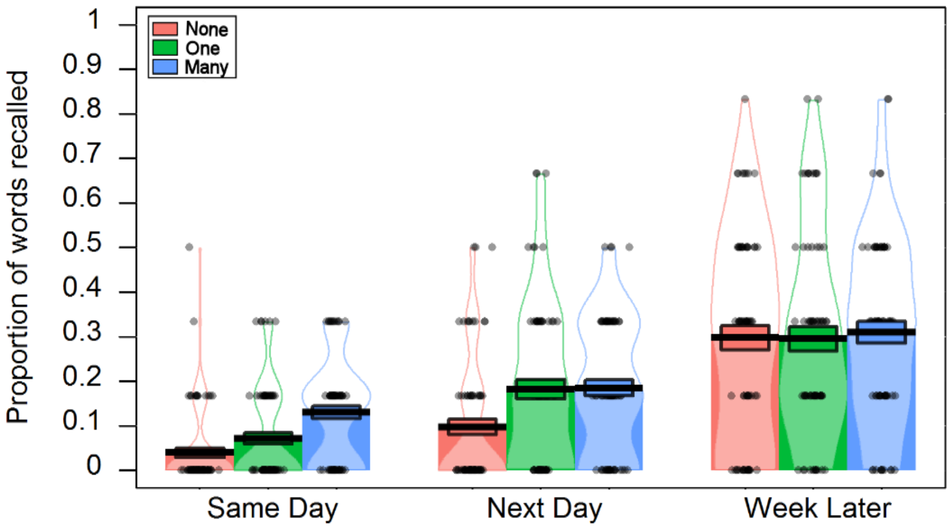


Figure 7. RDI plot of children's cued form recall performance in Form Neighbour Experiment 3, plotted by neighbour condition and test session. Thick black horizontal bars represent the mean for each condition, and surrounding boxes mark +/-1 standard error of the mean.

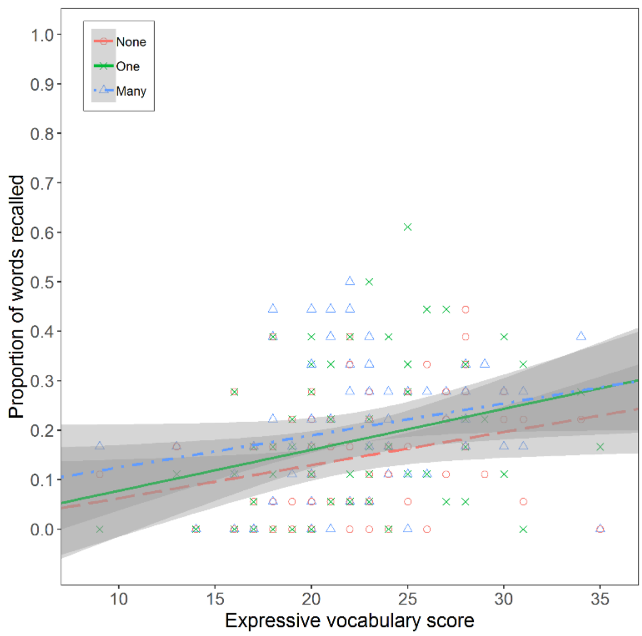


Figure 8. Scatterplot of mean proportion of words recalled for each word neighbour condition in Experiment 3 (collapsed across test session), plotted against children's expressive vocabulary score. Grey shaded areas represent 95% confidence intervals.

Recognition

Form-recognition. Recognition performance was very high (M = .88, SD = .32) and showed significant improvements across the week: performance was lower at T1 than subsequent tests (b = 0.28, Z = 4.94, p < .001), and continued to improve between T2 and T3 (b = 0.39, Z = 3.48, p < .001). The improvement after T1 was also related to changes in neighbour benefit, consistent with the cued-recall data: there was a larger difference between no and one/many neighbours (b = -0.06, Z = -2.18, p = .029) and between one and many neighbours (b = -0.11, Z = -2.13, p = .034) at T1 compared to the subsequent sessions (Figure 9). As with previous experiments, vocabulary ability was a significant predictor of performance (b = 0.53, Z = 3.89, p < .001), but not in interaction with any other variable. No other factors/interactions predicted performance (Appendix B8).

Form-picture recognition. Performance was slightly worse in picture-recognition than form-recognition (M = .77, SD = 0.42), but remained stable over time. As with recognition tasks in previous experiments, only vocabulary ability was a significant predictor of performance (b = 0.34, Z = 3.05, p = .002; Appendix B9-B10).

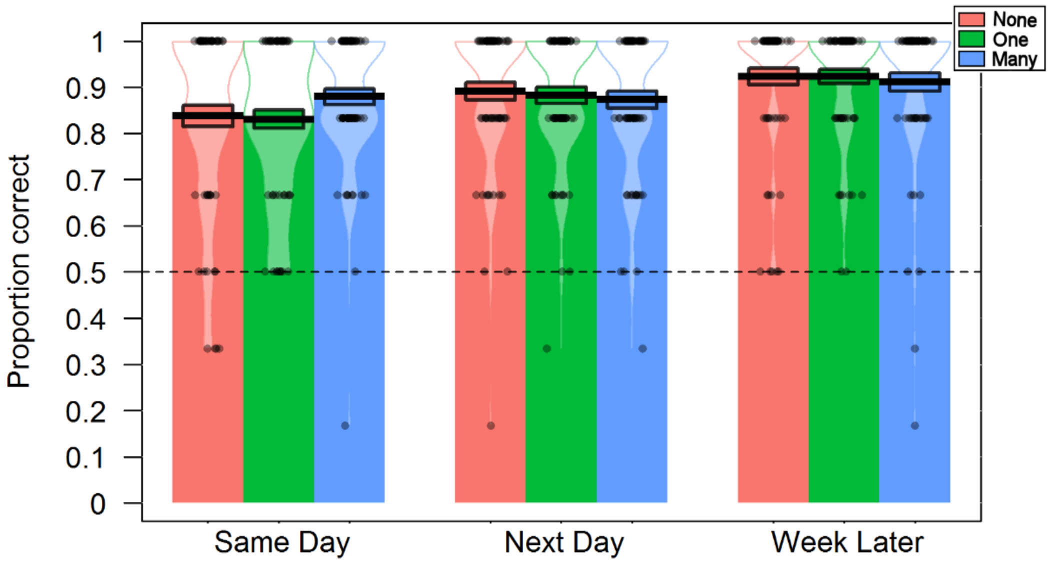


Figure 9. RDI plot of children's form recognition performance in Form Neighbour Experiment 3, plotted by neighbour condition and test session. Thick black horizontal bars represent the mean for each condition, and surrounding boxes mark +/-1 standard error of the mean. The dashed line indicates chance level performance.

4.5.4 Experiment 3 Discussion

Experiment 3 provided further evidence that children benefit from word neighbours in initial word acquisition, but that this benefit is short-lived: consolidation processes can facilitate memory for highly distinctive information in this age group (note we use the term "distinctive" to refer to pseudowords that are phonologically dissimilar to real English words). These no-neighbour words were slower to "catch up" with many-neighbour words in Experiment 3 than in Experiment 1: this may have resulted from the additional condition competing for consolidation processes, and/or overall weaker lexical representations that might be more demanding on consolidation processes to strengthen (Drosopoulos, Schulze, Fischer, & Born, 2007). Importantly though, the presence of this neighbour-by-delay interaction remained even using a different testing format (spoken versus written) and at a lower level of average recall performance than in Experiment 1, and was apparent even in recognition data.

Vocabulary ability remained a clear predictor of overall performance in both the recall and recognition tasks, lending support to global prior knowledge contributions to word learning. Contrary to our hypotheses, this global vocabulary ability again did not interact with neighbour benefit and/or test session in predicting performance. This is somewhat surprising considering the interactions seen in Experiment 2 with the single neighbour condition, and alongside previous studies showing associations between standardised measures of vocabulary and overnight consolidation of novel words that overlap with one real word (e.g., Henderson et al., 2015). However, as noted above, recall performance was substantially lower in this experiment (*M* = 3.20/18 words, *SD* = 6.88), leaving less variability in performance to distinguish individual differences in experimental manipulations than in Experiment 2 (*M* = 8.4/24 words, *SD* = 11.46).

4.6 Comparison between children and adults

Experiments 1 and 3 showed a clear pattern with children: an initial benefit of word neighbours that declined after opportunities for consolidation, leaving no overall benefit for local knowledge connections on memory. For adults however (Experiment 2), this pattern was less clear, featuring weaker evidence of a decline in neighbour benefit. This developmental difference is consistent with the hypothesis that children have a greater propensity for offline consolidation of distinctive information (James et al., 2017; Wilhelm et al., 2012). To further explore this possibility, we carried out an additional unregistered cross-experiment analysis on cued recall data. We analysed just the no- and many-neighbour conditions (as the one-neighbour manipulation was absent in Experiment 1). Two orthogonal contrasts compared performance across experiments: the group contrast compared adults (Experiment 2) versus children (Experiments 1 & 3), whereas the modality contrast compared the two child experiments that differed in the inclusion of orthography (Experiment 1) versus spoken language only (Experiment 3). Contrasts were set for test session as in previous analyses.

4.6.1 Cross-experiment results

The full model is presented in Table 7. We were predominantly interested in differences in consolidation between children and adults. Children continued to improve to a greater extent than adults later in the week (T2 to T3; *group:delay2*). Most importantly, there was a significant three-way interaction between participant group, test time (T1 vs. T2&T3), and the neighbour effect (*group:delay1:neighb*). The negative coefficient shows that children experienced a larger reduction in neighbour effect at later time points than adults (Figure 10).

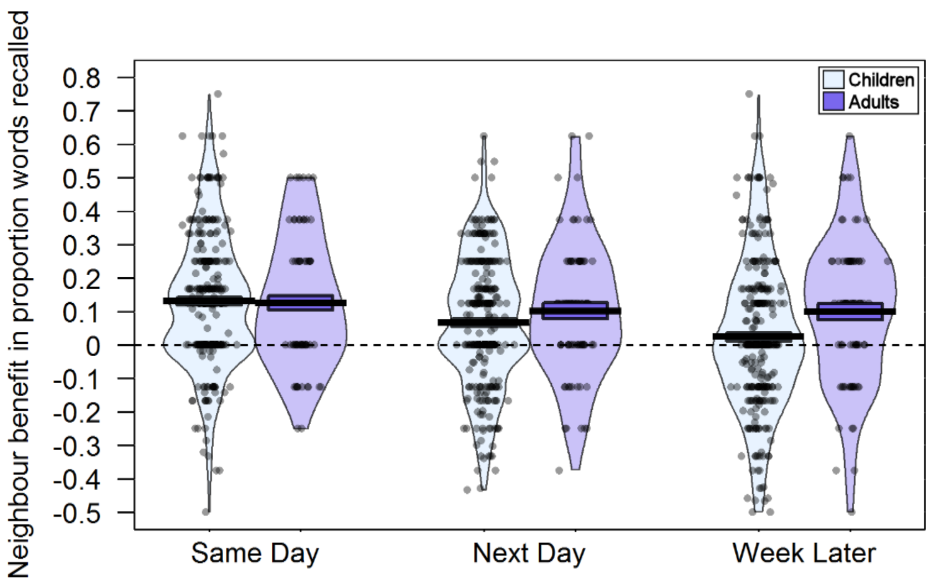


Figure 10. RDI plot of the neighbour benefit at each time point for child and adult participants across experiments. Thick black horizontal bars represent the mean difference in performance for many and no neighbour words, and surrounding boxes mark +/-1 standard error of the mean. The dashed line indicates no difference in performance across neighbour conditions, such that positive values mean better performance in the many-neighbour condition.

Table 7. Comparing cued recall performance across all three form neighbour experiments (no vs. many neighbours)

Fixed effects	*b*	*SE*	*z*	*p*
(Intercept)	-1.44	0.25	-5.77	<.001
group	**-0.27**	**0.09**	**-3.11**	**.002**
modality	**-0.80**	**0.13**	**-6.39**	**<.001**
delay1	**0.35**	**0.04**	**9.72**	**<.001**
delay2	**0.40**	**0.04**	**11.09**	**<.001**
neighb	0.26	0.24	1.07	.283
group:delay1	0.03	0.02	1.91	.057
modality:delay1	**0.11**	**0.03**	**3.45**	**.001**
group:delay2	**0.17**	**0.02**	**7.45**	**<.001**
modality:delay2	**0.15**	**0.04**	**3.53**	**<.001**
group:neighb	-0.05	0.07	-0.71	.477
modality:neighb	-0.05	0.10	-0.49	.627
delay1:neighb	**-0.11**	**0.03**	**-3.33**	**.001**
delay2:neighb	**-0.08**	**0.03**	**-2.47**	**.013**
group:delay1:neighb	**-0.03**	**0.01**	**-2.14**	**.033**
modality:delay1:neighb	-0.03	0.03	-0.89	.374
group:delay2:neighb	-0.04	0.02	-1.77	.077
modality:delay2:neighb	**-0.09**	**0.04**	**-2.16**	**.031**
Random effects	Variance	*SD*		
participant: (intercept)	1.64	1.28		
participant: (slope) neighb	0.11	0.34		
participant: (slope) delay1	0.02	0.13		
participant: (slope) delay2	0.01	0.12		
item: (intercept)	0.88	0.94		
item: (slope) group	0.06	0.24		
item: (slope) modality	0.09	0.29		
item: (slope) delay1	0.01	0.11		
item: (slope) delay2	0.00	0.05		

Note. Model formed from 16463 observations, collected from 378 participants across 16 items. Orthogonal contrasts were used for three-level factors: delay1 (Session 1 vs. Sessions 2&3), delay2 (Session 2 vs. Session 3), group (exp2 vs. exp1&3), modality (exp 1 vs. exp3)

4.7 General Discussion

This study examined the ways in which prior linguistic knowledge supports new word learning, and how this might differ from childhood to adulthood. First, we manipulated pseudowords' local connections to prior knowledge, using word-form neighbourhoods, and demonstrated that a pseudoword's similarity to existing English word-forms was advantageous for its immediate recall. Contrary to our initial hypotheses, our findings suggested that these neighbour benefits may be relatively short-lived: a period of offline consolidation reduced the influence of word neighbours on longer-term memory, notably more so for children than for adults. Second, we assessed more globally the prior linguistic knowledge that individuals bring to the task, and showed that existing vocabulary ability was a strong predictor of performance in all measures of pseudoword learning in both children and adults. However, in relating this to our word-neighbour manipulation, our adult data suggest that having one related word-form in vocabulary may be sufficient to facilitate recall of a new word. This supports an active role for prior knowledge in word learning, albeit more constrained than initially hypothesised.

4.7.1 The influence of local neighbourhood in learning and consolidating new words

Consistent with previous experiments using a similar word-neighbour paradigm (e.g., Hoover et al., 2010; Storkel et al., 2006), our experiments showed an initial learning benefit for pseudowords with existing neighbours. We interpret this result as demonstrating that local connections with existing knowledge can facilitate initial acquisition and/or immediate recall of new words, consistent with accounts of memory processing that highlight benefits of schematic knowledge in learning new information (e.g., van Kesteren, Ruiter, Fernández, & Henson, 2012). One plausible mechanism for this facilitation is that it is the word forms themselves provide the schematic structure for supporting learning. Alternatively, neighbouring word forms may provide access to alternative semantic information that can implicitly or strategically facilitate learning (Dumay et al., 2004). Subjective reports collected from Experiment 2 supported this latter proposal, and we consider ways to further elucidate the causal mechanisms in Section 4.7.4.

The present study set out to specifically address *when* prior knowledge connections might support word acquisition and consolidation. Based on the iOtA model, one possibility was that the overlap between novel words and their neighbours would provide further support during offline consolidation, leading to a larger benefit for novel items with multiple neighbours. However, our data were not in line with these predictions, and instead showed the opposite pattern: words with no local connections to existing knowledge showed greater improvements with a period of offline consolidation, reducing the benefit of word neighbours over time. We consider two possible interpretations for this finding. First, in accordance with the cortical learning interpretation of the CLS model, we proposed that these no-neighbour words are more reliant on hippocampal mechanisms during initial acquisition and thus undergo the biggest changes during subsequent consolidation. By this account, the prior knowledge available to support the learning of words with neighbours is proposed to speed neocortical encoding (Kumaran et al., 2016; McClelland, 2013), without the need for this integration to happen offline (van Kesteren et al., 2013). This neocortical learning account would similarly explain why existing vocabulary knowledge had no further role in consolidation in the present studies. Alternatively, we could also consider that the reduced offline improvements for items with neighbours can be attributed to increased interference when integrating the words with existing vocabulary knowledge (Storkel, Bontempo, & Pak, 2014), not present for words without neighbours. Future studies will benefit from using behavioural or neuroimaging markers of lexical integration to distinguish these processes of initial acquisition and consolidation.

It is important to acknowledge that the effects we attribute to offline consolidation may be partly a consequence of repeated retrieval practice (e.g., Tamminen & Gaskell, 2013). However, it has been demonstrated that overnight recall improvements occur in the absence of repeated retrieval practice (Henderson et al., 2013c). Furthermore, Havas et al. (2017) found similar changes in schema benefit to occur over a 12-hour period containing sleep and not over an equivalent period of wake. This suggests that reductions in schema benefit can be at least partially attributed to offline consolidation processes, given that the wake group will have had identical amounts of retrieval practice. We anticipate that offline consolidation processes contribute to the changes seen at T2 alongside other sources of variability, and similarly at T3 despite other influences being greater. Our key questions remain of theoretical interest in these contexts: the differences acquiring and consolidating words that can/cannot benefit from connections to prior knowledge, and differences attributable to the learner's existing vocabulary ability.

4.7.2 Developmental differences in the influence of word neighbours

The reduction of neighbour influence after opportunities for offline consolidation was most striking in the two child experiments: both showed an initial neighbour benefit on the same day as learning that had disappeared by the following week, despite differences in test format and levels of performance. However, the reduction in neighbour influence over the week of the experiment did not reach significance when modelling the adult data alone (ps $\geq .06$). Our cross-experiment analysis showed that children receive greater benefit from offline consolidation than adults overall, and that this supported a larger reduction in neighbour influence overnight. As such, the data suggest that children have a greater propensity for consolidating schema-unrelated information than adults.

We speculate that this dissociation may relate to changing neural mechanisms that support learning and consolidation across development. As reviewed in James et al. (2017), children typically show a higher proportion of consolidation-relevant processes (e.g., slow neural oscillations) during sleep than adults (Feinberg & Campbell, 2010; Ohayon et al., 2004), which may support their enhanced consolidation of new words (see also Gómez & Edgin, 2015, for a review of sleep and memory changes earlier in childhood). Adults have a greater amount of prior knowledge to support consolidation, which may compensate for reduced levels of sleep-associated consolidation processes in many tasks (Wilhelm et al., 2008; Wilhelm et al., 2012). Interestingly, using a motor sequence task that could not benefit from prior knowledge, Wilhelm et al. (2013), children showed greater gains in recall performance than adults over sleep, which could be linked to their higher levels of slow-wave sleep activity. Our present findings are consistent with this pattern, but a valuable future direction will be to measure brain activity during sleep to discover whether differences can indeed be attributed to sleep-associated processes in this domain.

4.7.3 Relating global vocabulary knowledge to the influence of word neighbours

Given that local connections to prior knowledge appeared to speed word learning, we asked whether individuals with good global vocabulary knowledge could capitalise upon their superior lexical knowledge when learning words that could benefit from such connections. Experiment 1 did not offer support for this hypothesis: although children with good vocabulary ability emerged as better word learners, they did not show a superior benefit of word neighbours than children with poorer vocabularies. However, it became apparent in Experiment 2 that adults' global vocabulary ability was a better predictor of performance in the one-neighbour condition, suggesting that those with good vocabulary may actively benefit compared to those with poorer vocabulary when learning such items. Although there is some evidence of a linear relationship between the number of word neighbours and learning performance in preschool children (Storkel, Bontempo, Aschenbrenner, Maekawa, & Lee, 2013), these effects have been very small, and not investigated at the lower end of the scale (≥ 4 neighbours). Our data suggest that the most critical difference appears to be in having one versus no neighbours activated in learning (Bowers et al., 2005), and thus that learning words with many neighbours is less sensitive to vocabulary-related differences because most participants will access at least one neighbour.

Despite the clear influence of existing vocabulary ability on learning one-neighbour words for adults, we did not observe this finding in our subsequent child study (Experiment 3). One explanation is that the neighbours of our selected stimuli may not have been readily accessed during learning by children of this age. Interestingly, whilst the neighbours spanned age-of-acquisition ratings aimed to maximise differences between individual children, we instead observed these individual differences only in adults. This leads us to suggest that the support offered by accessing a word neighbour may be driven by the quality of the individual's lexical representation (Perfetti, 2007) and their experiences with the word (Vitevitch, Storkel, Francisco, Evans, & Goldstein, 2014), rather than simply word familiarity. Lexical quality is likely lower in children than adults, and also in adults with lower expressive vocabulary ability – a measure that arguably probes well-specified and rich lexical representations. These differences could account for the variability in benefiting from a single neighbour in word learning in two ways: by affecting the likelihood of activating the neighbour, or by whether this activated representation is rich enough to provide support. Perhaps one neighbour is sufficient if this representation is of high quality, but that cumulative activations are necessary for those with weaker lexical representations.

However, there was no significant difference between performance in the one and many neighbour conditions for children overall, suggesting that children were still receiving some support from existing knowledge in this single neighbour condition. One possibility is that this support could be being driven by a different mechanism than in adults. Alternatively, we think it likely that the significant reduction in overall performance in Experiment 3 may have reduced the variability present to detect individual differences in neighbour benefit between children. Whilst the overall pattern of neighbour influence in relation to vocabulary ability looked very similar in Figure 6 and Figure 8, the lower overall levels of cued-recall makes clear the lack of variability across participants. Future studies should thus closely examine conditions under which individual differences in prior knowledge benefit emerge, by manipulating the levels of learning performance and the accessibility of word neighbours.

It is also interesting to consider that the ways in which local and global prior knowledge contribute (and interact in contributing) to word learning performance might vary depending on the retrieval conditions. Whilst we saw influences of test session and local neighbour manipulations in tasks primarily assessing knowledge of the new word forms (cued recall, Experiment 3 form recognition), only global vocabulary knowledge remained a significant predictor in tasks that also presented the novel objects as a cue (recognition in Experiments 1, 2; picture-form recognition in Experiment 3). There are many differences between these tasks beyond the presence of semantic information: the latter tasks primarily tested familiarity rather than holistic retrieval of all elements, they require only a button press response from presented options, and performance levels are much higher. Given that previous studies have demonstrated recall benefits for many-neighbour words when cued with pictures rather than stem completion tasks, we speculate that the performance differences seen are largely driven by differences in retrieval demands. Nevertheless, it remains an interesting avenue to explore the kinds of learning and memory that are most influenced by prior knowledge.

4.7.4 Evidence for prior knowledge contributions to Matthew effects

If vocabulary-related differences in the neighbour benefit only emerge when a new word has one existing neighbour, it begs the question of whether prior knowledge really makes an active contribution to Matthew effects in word learning in light of other explanations: do we readily encounter new words with only one neighbour that enable some individuals to benefit more than others? Whilst vocabulary ability is undoubtedly related to word learning performance in ways beyond prior knowledge support (as suggested by its strong association with performance across all experimental tasks regardless of other manipulations), three arguments make the prior knowledge account worthy of further investigation. First, we do learn a substantial number of words with very small neighbourhoods: 9.94% of the 40,481 entries in the English Lexicon Project have only one orthographic and phonological neighbour, making it plausible that some individuals may be able to excel in learning this new vocabulary quicker than others. Second, individual variability in knowing at least one word-neighbour is greater when the lexicon is smaller, suggesting that there could be a greater contribution of prior knowledge to Matthew effects in younger children.

Third, the lack of individual differences in the many-neighbour condition may have partly resulted from the training paradigm used here, in which learners could make use of strategies that actively involved neighbouring words. As noted above, adults reported using explicit strategies to link the novel words to known neighbours, and develop semantic connections between the two. In encountering a new word within more naturalistic contexts, the activation of a word-neighbour may be less reliable and greater influenced by the quality of the learner's existing lexical representations. Furthermore, without opportunities to make explicit linguistic connections during training, benefits of prior knowledge may be more likely to emerge offline during consolidation. Indeed, when Henderson et al. (2015) presented novel words within a story context, existing vocabulary ability correlated *only* with overnight improvements in cued-recall, and not with immediate performance. This difference in learning context may be a key factor in explaining why we did not find a relationship between existing vocabulary knowledge and consolidation of new words (see also Henderson & James, 2018): there may be no further benefit for prior linguistic knowledge during consolidation when explicit strategies can be used during learning, whereas more implicit connections to existing knowledge may be strengthened by consolidation (in line with the iOtA model). We therefore speculate that prior knowledge-related differences in naturalistic word learning may be understated by the present experiments, and suggest that future studies should consider using alternative training paradigms.

4.7.5 Conclusions

This study revealed that children and adults benefit from local connections to prior linguistic knowledge during word learning: novel words with one or more neighbours were recalled better at the initial test points, suggesting they could be acquired more quickly than words with no neighbours in the English language. This immediate benefit for prior knowledge favours accounts of memory consolidation that permit early neocortical learning for schema-related information. However, children's but not adults' memory for no-neighbour words reached equivalent performance levels by the end of the week, supporting proposals that ongoing neural development in children may provide increased support for consolidating large amounts of new information during this developmental period. These data demonstrate that the CLS model of learning could be further informed by taking developmental approaches that seek to contrast the contributions of different mechanisms to learning and consolidation. Furthermore, understanding how the relative reliance on prior knowledge changes across the lifespan may also be important for understanding why early language difficulties can persist to adulthood, and stresses the importance of targeting difficulties whilst learning mechanisms are most able to overcome such constraints.

Chapter 5. Accessing Prior Linguistic Knowledge when Learning Words from Stories

All experiment pre-registrations, materials, data, and analyses are available on the Open Science Framework: https://osf.io/stx6q

5.1 Abstract

Children and adults show advantages for learning pseudowords with versus without phonological neighbours under explicit training conditions, supporting the proposal that new vocabulary can benefit from connections to prior knowledge. In the present study, we examine the extent to which such neighbour benefits persist when pseudowords are encountered incidentally in stories, with more limited opportunities for strategically engaging prior knowledge. Children (Experiment 1) and adults (Experiment 2) were exposed to pseudowords with zero, one, or many neighbours via a spoken story with illustrations. After listening to the story, participants completed a stem completion task to assess recall of the new word-forms, and a recognition task to assess familiarity with the forms and their meanings. The memory tasks were repeated one day and one week later to assess changes in memory after opportunities for offline consolidation. Children and adults both improved in their ability to recall the word-forms at the later tests, but only adults were influenced by the pseudowords' phonological neighbours. Like in previous studies using explicit training, adults' benefit of phonological neighbours was apparent immediately and persisted across the week. In contrast, children were less able to benefit from neighbours after encountering pseudowords in stories, and actually performed more poorly in their ability to recognise pseudowords with versus without neighbours. We consider how vocabulary learning strategies may be differently engaged for children and adults across learning contexts, and discuss alternative mechanisms of prior knowledge support in consolidating new words learned from stories.

5.2 Introduction

Both children and adults face the task of acquiring new vocabulary from a multitude of situations. Many words are taught explicitly and intently, both via early language learning experiences and formal vocabulary instruction in school, yet the majority are encountered incidentally and acquired without strategic effort. Understanding the factors that influence word learning in these incidental learning contexts may thus be key to understanding why some children acquire vocabulary at a slower rate than others across development. In explicit teaching contexts, studies suggest that children and adults can bootstrap new words to similar word-forms in their existing vocabulary to speed new word acquisition (James et al., 2018). Here, we ask whether this prior linguistic knowledge can also support incidental vocabulary learning from stories.

5.2.1 Learning and consolidating new vocabulary

Models of learning distinguish between processes that help us to quickly acquire new words from the environment and those that enable these newly formed representations to become consolidated in long-term vocabulary. Davis and Gaskell (2009) applied the Complementary Learning Systems model (McClelland et al., 1995) to vocabulary learning, describing two neural systems involved in new word acquisition. The hippocampal learning system enables rapid learning about a new word: its phonological form, its meaning, and syntactic properties. The neocortical learning system represents longer-term memory, whereby the distributed nature of lexical storage allows for speeded processing of linguistic information. Integration of new words into this existing vocabulary system is a slower process requiring a prolonged period of consolidation. Recent evidence supports that consolidation is facilitated by processes that happen “offline” during sleep, and a number of studies support that sleep (versus wake) can improve explicit knowledge of new words as well as their interaction with existing vocabulary (e.g., Dumay & Gaskell, 2007; Henderson et al., 2012). To understand the ways in which prior knowledge might support new word acquisition, it is therefore important to assess new word memory before and after opportunities for consolidation.

5.2.2 The role of global prior knowledge in learning and consolidating new vocabulary

Recent models of word learning have considered factors that might support this slower consolidation process for new words. Drawing on schema-based accounts of learning and memory (e.g., McClelland, 2013), James et al. (2017) proposed that prior linguistic knowledge may facilitate more rapid integration of new words into existing vocabulary. A number of studies support that individual differences in vocabulary knowledge are predictive of improvements in memory for new words during offline consolidation (e.g., Henderson et al., 2015; Horváth et al., 2015b; Sénéchal, Thomas, & Monker, 1995). For example, Henderson and James (2018) presented children aged 10-11 years with novel words (e.g., *crocodol*) embedded in stories. When tested with a stem completion task ("Which word began with *cro*-?"), children with higher scores on a standardised vocabulary assessment improved more overnight than children with poorer vocabulary. This benefit was specific to learning new words in varied story contexts that enabled children to make multiple connections to their richer vocabulary. Studies such as this one suggest that "global" prior knowledge – i.e., an individual's existing lexicon - may facilitate the offline consolidation of new vocabulary.

5.2.3 Local manipulations of prior knowledge in word learning

However, correlational studies between global vocabulary knowledge and consolidation of new word-forms cannot address the nature and specificity of this lexical support. To better understand the mechanisms that might underlie this relationship, one approach has been to manipulate potential prior knowledge connections at the word level – hereafter, "local" prior knowledge. These local prior knowledge manipulations have used phonological neighbours to quantify similarity to potential word knowledge. For example, studies by Storkel and colleagues (Hoover et al., 2010; Storkel, 2009; Storkel et al., 2006) taught participants pseudowords that varied in the number of real words that could be created by substituting a single phoneme. Thus, words with few phonological neighbours have more limited potential connections to existing knowledge than words with many phonological neighbours. This paradigm has consistently demonstrated that phonological neighbours facilitate word learning across a wide range of ages (Storkel, 2009; Storkel et al., 2006) and languages (van der Kleij, Rispens, & Scheper, 2016). There is also some – albeit limited - evidence that the size of an individual's benefit from word neighbours is positively correlated with their expressive vocabulary knowledge, in pre-school children (Storkel & Hoover, 2011) and in adults (Experiment 2, James et al., 2018). This relationship follows the logic that those with good existing vocabulary will likely know more of a word's local neighbours, and supports that global vocabulary knowledge may plausibly offer some support to new learning at the local word level.

5.2.4 Local manipulations of prior knowledge during consolidation

Whilst studies manipulating phonological neighbours have consistently demonstrated a benefit for local connections to prior knowledge in learning new words, few studies have assessed the longevity of these benefits, and those that have present somewhat conflicting findings. Based on studies of global vocabulary and consolidation (James et al., 2017), one possibility is that prior knowledge benefits may be exacerbated during consolidation, following increased opportunities for the new words to engage with existing vocabulary knowledge. Storkel and Lee (2011) found results that were consistent with this pattern: 4-year-old children exhibited a neighbourhood density benefit only at a one-week retention test, and not when tested immediately after learning.

In contrast, a number of studies have indicated the opposite pattern: a benefit of prior knowledge immediately that *diminishes* after opportunities for offline consolidation. For example, in two experiments by James et al. (2018), 7-to-9 year-old children showed a significant benefit of phonological neighbours in recalling pseudowords immediately after learning but not at the one-week retention test. One explanation for this is that the pseudowords with strong connections to prior knowledge might engage with the neocortical system immediately. Pseudowords that do not benefit from local prior knowledge are instead proposed to be more hippocampally dependent at encoding, and therefore receive greater benefit from offline consolidation processes (see Havas et al., 2017; Mirković & Gaskell, 2016, for similar interpretations). This pattern is challenging to reconcile with evidence that global prior knowledge predicts overnight improvements in new word knowledge, although such a relationship was not found in either experiment when using local manipulations.

In sum, it is clear that local connections to prior knowledge facilitate new word learning in children and in adults, but that the time course of this benefit and its longevity are not well understood. One key difference in the reviewed studies is whether there was a prior knowledge benefit immediately after learning, leading us to propose that the strength of prior knowledge activation during encoding determines whether prior knowledge benefits emerge immediately after learning or later during consolidation. In the context of the CLS model, it may be that pseudowords with strong local connections to prior knowledge have reduced need for neocortical connections to be strengthened offline. Conversely, weaker activation of prior knowledge during encoding may leave potential prior knowledge connections susceptible to further strengthening during later consolidation.

5.2.5 Contexts of activating prior linguistic knowledge

What might affect the engagement of prior linguistic knowledge during learning? One possibility is the use of explicit strategies: in directing attention towards and actively trying to learn the word-form and its meaning, similarities to known word-forms may become part of the learning process. Indeed, adults in James et al. (2018) reported making intentional comparisons to the words they knew, and using those similarities to make semantic connections with the novel objects being learned (from subjective reports, data unpublished). This early, proactive engagement with prior knowledge may bypass the need for strengthening offline.

On the contrary, the studies that have demonstrated relationships between global prior knowledge and later overnight consolidation of new words have presented the new words in spoken stories (Henderson et al., 2015; Henderson & James, 2018; Sénéchal et al., 1995) – minimising opportunities for individuals to draw comparisons between new and known words in a strategic manner. In the present study, we sought to bring these two approaches to understanding prior knowledge in vocabulary learning in line. Whilst previous studies of phonological neighbours in word learning have sometimes presented the to-be-learned items in stories (e.g., Hoover et al., 2010; Storkel et al., 2006), participants have still been made aware of the learning nature of the task and test trials have been interleaved with story exposures. Here, we test whether *incidental* learning of novel words via stories leads to later-emerging benefits of local prior knowledge connections.

5.2.6 The present study

In the present study, we tested whether embedding pseudowords in stories (i.e., incidental word learning) results in the later emergence of local prior knowledge benefits compared to previous studies in which pseudowords were learned intentionally. We used a subset of 15 pseudowords from James et al. (2018), and embedded them in a story based on Henderson et al. (2015). Importantly, the pseudowords came from one of three phonological neighbour conditions (none, one, many), enabling us to directly assess how these potential connections to existing knowledge might benefit new word acquisition for children (Experiment 1) and adults (Experiment 2) of varying global vocabulary ability. To assess the time course of these prior knowledge benefits, we used recall and recognition tasks to test memory for the new word-forms immediately after learning, the next day, and one week later. In doing so, we aimed to better understand early versus late influences of prior knowledge in vocabulary learning initially in children, and later in adults for whom vocabulary learning from stories is less challenging (e.g., Henderson et al., 2015). We predicted that the benefit of phonological neighbours would be initially weaker after learning pseudowords from stories compared to previous studies that used direct teaching (James et al., 2018). Under such circumstances, we predicted that these more fragile connections to prior knowledge would be *strengthened* over a period of offline consolidation, as a result of increased opportunities for the new lexical representations to engage with existing global vocabulary knowledge. This strengthening could plausibly lead to later-emerging benefits of prior knowledge than previous studies.

Our primary research questions were: 1) Do individuals show benefits of offline consolidation after learning pseudowords from stories? 2) Do individuals show benefits of local prior knowledge when learning pseudowords from stories? 3) Is the time course of prior knowledge benefits different when individuals learn novel words from stories, compared to previous studies using explicit teaching paradigms? And 4) Are benefits of prior knowledge related to individuals' existing expressive vocabulary knowledge?

5.3 Experiment 1

5.3.1 Experiment 1 Methods

Participants

Six Year 4 and 5 classes were recruited from across four primary schools in the North Yorkshire area. Consent was gained from each school's headteacher, alongside parental consent on an opt-out basis. The study was approved by the Psychology Research Ethics Committee at the University of York. From the initial

123 children, two withdrew their participation. Twenty children were excluded from analyses due to their knowledge of another language (c.f., Meade, Midgley, Dijkstra, & Holcomb, 2018), and a further four children could not be entered into analyses because of absence at the time of the vocabulary assessment. These additional datasets are available online. The final analyses presented incorporate data from 97 children (51 male), aged between 8;06 – 10;09 years (*M* = 9;07). This age range was selected to overlap with previous studies (James et al., 2018), but was slightly older to reduce risk of floor effects when learning from stories. Three of the children were absent on the second day, and thus contribute only two sessions of data to the final analyses.

Design and procedure

Each child participated in three test sessions, administered individually in a quiet area of the school. On Day 1, children completed the learning phase of the experiment, in which they were exposed to 15 pseudowords presented in a spoken story. They then participated in memory tests for the new words immediately afterwards (T1), the next day (T2), and one week later (T3). All experimental tasks were programmed in OpenSesame (materials available on the OSF) and administered on a laptop with headphones.

During the second and third test sessions, each child's vocabulary and nonverbal ability was assessed using the Wechsler Abbreviated Scale of Intelligence (WASI-II; Wechsler, 2011). The administration of a short home language questionnaire identified those children who were not native monolingual English speakers, enabling us to exclude them from the present analyses.

Experimental stimuli

A subset of fifteen pseudowords were selected from the stimuli used in James et al. (2018). These pseudowords had been selected from the English Lexicon Project (Balota et al., 2007) for having no, one, or many orthographic neighbours, which also aligned with the number of phonological neighbours (CLEARPOND; Marian et al., 2012). We selected five words from each condition (reduced from six) due to the lower levels of performance seen in comparable story learning paradigms (Henderson et al., 2015). The three lists remained matched for phoneme and letter length, and bigram probability. All words were bisyllabic, and began with a single consonant and vowel for purposes of cueing in the recall task.

Learning phase

Children were exposed to the novel words embedded in a spoken story, presented audio-visually. The story was based on that created by Henderson et al. (2015) - *Trouble at the Intergalactic Zoo* - replacing the novel words with those described above. The original story contained 12 pseudowords, but we increased this number to 15 to increase the power within each word-neighbour condition (extending the story where necessary). There were five exposures to each word within the story, occurring across 3-4 different paragraphs.

To facilitate engagement with the story, a number of accompanying images were created using PowerPoint, and their presentation programmed along with the audio recording using OpenSesame. Each pseudoword had its corresponding object feature in three of the picture scenes. Three versions of the story were recorded in order to counterbalance whether any novel object was assigned a name from the no-, one-, or many-neighbour condition.

During the learning phase, children were warned that they may not know all the words in the story, but that they should keep listening until the end of the story without asking questions. The story lasted approximately 7 minutes.

Test phase

Cued form recall. Recall of the new word-forms was assessed using a stem completion task. Children were cued with the first consonant and vowel sound from each novel word, and were asked to speak the remainder of the word that they heard during the story. Partial attempts were encouraged even if children were not certain of their responses. Items were presented in a randomised order using OpenSesame, and the experimenter transcribed the responses for scoring offline.

Form recognition. Children heard each new pseudoword paired with a phonological foil (incorporating a vowel change), and were asked to select which word they heard during the story. They responded using keys assigned to the first or second option, and completed two practice trials (known words and foils) with feedback to adjust to the response mappings.

Form-picture recognition. To assess learning of the semantic mappings, children were presented with each of the novel objects in the story and were asked to select its name from two pseudoword options using a key press. The incorrect answer for each trial was always another pseudoword heard during the story, and remained consistent across test sessions.

Analyses

Analysis plans were pre-registered on the Open Science Framework prior to the completion of data collection (http://osf.io/t5fmd). Analyses were conducted in R, using *lme4* (Bates et al., 2015b) to fit mixed effects models and *ggplot2* (Wickham, 2016) for graphs. A mixed effects binomial regression model was used to analyse each of the dependent variables, with fixed effects of session, neighbourhood condition, vocabulary ability, and all corresponding interactions. Orthogonal contrasts were used for each of the factorial predictors. For the fixed effect of session: *delay1* contrasted responses before and after opportunities for offline consolidation (T1 vs. T2&T3), and *delay2* assessed continued changes T2 vs. T3. For the fixed effect of neighbours: *neighb1* contrasted words without vs. with neighbours (no vs. one&many), and *neighb2* contrasted words with one vs. many neighbours. We used raw vocabulary scores for analyses, which were scaled and centered before entering into the model.

For each analysis, we first computed a random-intercepts model with all fixed effects and interactions. If there was no indication of a three-way interaction in the model (all *p*s > .2), this was pruned to enable a more parsimonious model with better-specified random effects. We then incorporated random slopes into the model using a forward best-path approach (Barr et al., 2013), progressively adding slopes into the model and retaining only those random effects justified by the data under a liberal α-criterion ($p < .2$). In the text, we report statistics in full for only significant predictors of performance. The final model details and all statistics are presented in Appendix C (C1-C3).

5.3.2 Experiment 1 Results

Cued form recall

The proportion of pseudowords successfully recalled after listening to the story was very low (T1: $M = 0.03$, $SD = 0.17$), but significantly improved at later test points ($\beta = 0.68$, $SE = 0.07$, $Z = 9.18$, $p < .001$). Recall performance also continued to improve substantially between T2 ($M = 0.07$, $SD = 0.26$) and T3 ($M = 0.21$, $SD = 0.41$; $\beta = 0.84$, $SE = 0.08$, $Z = 10.63$, $p < .001$), supporting the hypothesis that opportunities for offline consolidation would improve recall for the pseudowords.

Vocabulary ability was a significant predictor of a child's recall performance ($\beta = 0.69$, $SE = 0.16$, $Z = 4.24$, $p < .001$), suggesting that global prior knowledge could plausibly support new word learning. However, more local connections to prior knowledge did not facilitate memory for the pseudowords: there was no benefit of word neighbours overall or in interaction with any other variable. No interactions reached statistical significance.

Form recognition

One participant's recognition data did not save properly at T1, and thus is missing from the recognition analyses. Immediately after story exposure, children could recognise the pseudowords over their phonological foils at above chance performance (T1: $M = .65$, $SD = .48$; $t(95) = 10.21$, $p < .001$). Performance improved at later test points ($\beta = 0.23$, $SE = 0.03$, $Z = 8.96$, $p < .001$), and continued to improve from T2 ($M = .75$, $SD = 0.43$) to T3 ($M = .80$, $SD = 0.40$; $\beta = 0.17$, $SE = 0.05$, $Z = 3.58$, $p < .001$). There was a significant effect of word neighbours on performance but - in contrast to our hypothesis - pseudowords with one ($M = .72$, $SD = .45$) or more ($M = .71$, $SD = .46$) neighbours were recognised more poorly than those without neighbours ($M = .78$, $SD = .42$; $\beta = -0.12$, $SE = 0.06$, $Z = -2.10$, $p = .036$).

Vocabulary ability was again a significant predictor of performance ($\beta = 0.30$, $SE = 0.08$, $Z = 3.88$, $p < .001$). There was a trend towards an interaction with neighbour condition, suggesting that children with good vocabulary performed slightly better for pseudowords without neighbours. However, this did not reach our threshold for statistical significance ($\beta = -0.07$, $SE = 0.04$, $Z = -1.82$, $p = .069$), nor did any other interaction in the model.

Picture-form recognition

Five participants were administered the incorrect version of this task during one session (according to their counterbalancing condition), and are excluded from this analysis. At the first test point, children could successfully select the correct name for the objects at above chance performance (T1: $M = 0.66$, $SD = 0.47$; $t(90) = 10.81$, $p < .001$). Memory for these picture-form mappings improved overnight (T2: $M = 0.72$, $SD = 0.45$; $\beta = 0.10$, $SE = 0.03$, $Z = 4.08$, $p < .001$), but there was no further improvement across the week (T3: $M = 0.72$, $SD = 0.45$; $p = .862$). As with the other two tasks, vocabulary ability was a significant predictor of performance overall ($\beta = 0.35$, $SE = 0.09$, $Z = 3.96$, $p < .001$). Vocabulary ability was also a significant predictor of consolidation in this task: children with higher vocabulary scores improved more from T1 to T2 and T3 than children with poorer vocabulary (β = 0.09, *SE* = 0.03, *Z* = 3.53, *p* <.001; Figure 11).

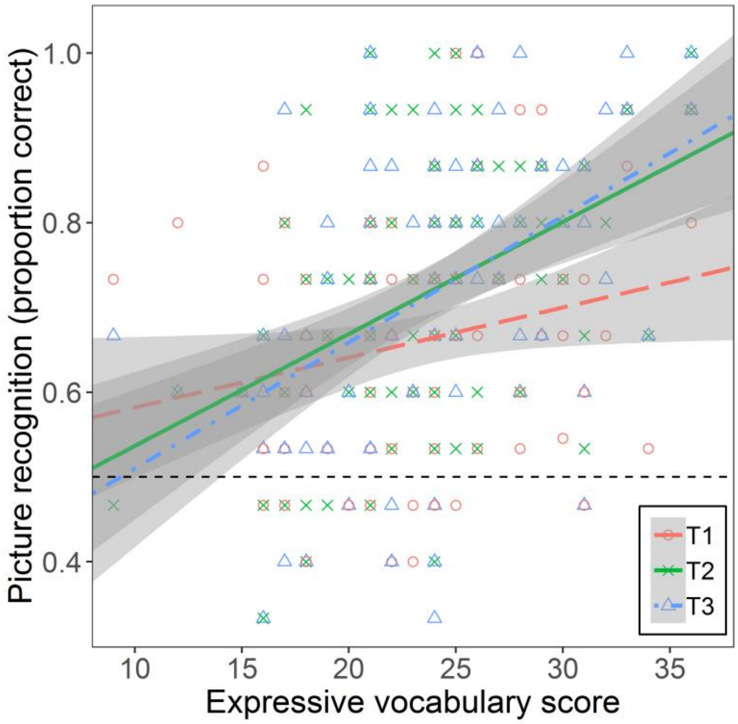


Figure 11. Mean picture-form recognition performance for immediate (T1), next day (T2) and week (T3) tests, plotted for each participant against their vocabulary ability. The dashed horizontal line marks chance performance.

5.3.3 Experiment 1 Summary

Children became familiar with the pseudowords after learning from the story at above-chance levels of recognition, but performance across all tasks was much lower than in previous experiments (James et al., 2018). This highlights the increased challenge of learning new words without explicit instruction, and the more incremental nature of learning new words incidentally. Despite this change in format, performance in all tasks improved with opportunities for offline consolidation.

The influence of phonological neighbours was not as robust in this experiment as in previous studies - likely due to the much lower levels of performance – and we found only a significant effect of neighbour condition in the form-recognition task. Interestingly, however, the effect appeared to be in the opposite direction to the one predicted: children were significantly *poorer* at recognising recently encountered pseudowords from their phonological foils if the pseudoword had phonological neighbours in the English language. This result differs from form-recognition performance in an earlier explicit teaching experiment, whereby children received an initial memory benefit from phonological neighbours with these stimuli (Experiment 3, James et al., 2018).

For the first time when using these pseudoword stimuli, there was some evidence that children with good vocabulary improved more with offline consolidation than those with poorer vocabulary. This relationship was present only for the task that tested semantic knowledge of the new items, which might perhaps indicate that the benefit of global vocabulary knowledge during consolidation is for strengthening connections with semantic knowledge. We should be cautious in drawing strong conclusions given the very low levels of performance in this experiment, but return to speculate on potential mechanisms in the General Discussion (Section 5.5.5).

The poor levels of recall in Experiment 1 present a significant challenge to understanding whether prior knowledge interacts differently with word learning that occurs incidentally through stories, compared to word learning that occurs in explicit training studies. We therefore conducted a second experiment with adults, whom we anticipated would show higher levels of recall. Although there is evidence that adults show different influences of prior knowledge during word learning than children – or at least a different time course of these effects – we conducted this experiment primarily as a comparison to previous adult studies to gain further insight into potential differences across contexts. Adults may be able to provide better insight into this question under the present experimental design, given that their superior language comprehension skill may leave more cognitive resources available to learn new vocabulary from story contexts.

5.4 Experiment 2

5.4.1 Experiment 2 Hypotheses

We pre-registered four hypotheses on the Open Science Framework (http://osf.io/cdyrw): 1) Memory for the novel words will be different after opportunities for consolidation at the day and week follow-up tests compared to when tested immediately after learning. We predicted that recall in the stem completion task would improve at later test points. 2) Memory for the novel words will be affected by their number of phonological neighbours. We predicted that words with one/more neighbours would be better recalled than words without neighbours in the stem completion task, but did not predict a direction for this hypothesis in the recognition task. 3) The influence of phonological neighbours on memory for the novel words will change after opportunities for consolidation. 4) Expressive vocabulary scores will be positively associated with overall memory performance for the new words, and that this association would be strongest for words with only one phonological neighbour.

5.4.2 Experiment 2 Methods

Participants

Experiment 2 was an online experiment. 130 adults were included in the analysis, and were recruited via Prolific Academic according to the following criteria: aged 18-35 years old, native monolingual British English speakers residing in the UK, with no reported visual, hearing, or literacy difficulties. In line with Prolific's recommendations for longitudinal studies, we restricted recruitment to individuals who had participated in at least ten studies on the platform with a minimum 95% approval rate. The study was approved by the Psychology Research Ethics Committee at the University of York. Participants received £5 for completion of all three sessions, and an additional £1 bonus if they completed each session within the same four-hour time window.

An additional 41 participants started the study but did not complete all three test sessions, and one participant failed an attention screener after listening to the story. A further 20 participants completed all three test sessions but were excluded for one/more of the following reasons: underage (*n*=1), self-report of external strategy use (*n*=3) or task misunderstanding (*n*=1), little evidence of learning (*n*=1), failure to complete the sessions by 9pm (*n*=3), or failure to complete the vocabulary task properly (*n*=13). The majority of vocabulary exclusions were due to participants not following the instructions (retyping the word or attempting to provide one of the learned pseudowords), and one participant was a clear outlier.

Design and procedure

Participants completed three test sessions online, programmed and hosted on the Gorilla Experiment Platform (www.gorilla.sc; Anwyl-Irvine et al., 2018). The first session took approximately 20 minutes, including a sound check, providing basic background information, reading along with a story, and memory tests for the new words encountered in the story. As in Experiment 1, each participant was exposed to and tested on 15 novel words, 5 from each of three word neighbour conditions (none,

one, many). They were informed that the experiment was testing how comprehension was affected by the inclusion of different numbers of nonsense words, as are frequently encountered in children's stories.

The second session (~5 minutes) was completed the day after the first, and involved completing the same memory tasks as in the first session. The third session was completed one week after the first session, and lasted approximately 10-15 minutes. Participants completed the memory tests for a third time, completed an assessment of their existing vocabulary knowledge, and filled out a questionnaire regarding strategy use.

Experimental stimuli

As Experiment 1.

Learning phase

As Experiment 1. However, we also added written text below each picture, and instructed participants to read along with the story. Our reasons for doing this were threefold: 1) to enhance performance; 2) to bring encoding procedure in line with the written test format for this study; and 3) to make the study more comparable to the explicit learning paradigm in James et al. (2018) for this age group.

Test phase

Cued form recall. Recall of the new word-forms was tested using a stem completion task, as for Experiment 1. However, they were also provided with the written cue (first consonant and vowel) alongside the spoken cue, and were required to type their responses. Answers were scored as accurate if they read as phonologically correct.

Recognition. We administered only a single recognition task, as in James et al. (2018). Participants were provided with each picture and were asked to choose which of four options its name was. The options consisted of the correct answer, a phonological foil for the correct answer, an incorrect learned answer, and the phonological foil for the incorrect learned answer. Participants could hear each option spoken by clicking a speaker.

Analyses

As Experiment 1. Model tables are in Appendix C (C4-C5).

5.4.3 Experiment 2 Results

Cued form recall

Recall performance was higher for this experiment compared to Experiment 1: adults successfully recalled a mean proportion of .20 of the pseudowords (SD = .40) in the first session, and showed small improvements at later test sessions (T2: M = .21, SD =.41; T3: M = .24, SD = .43). These improvements in performance were statistically significant after the first day (β = 0.06, SE = 0.03, Z = 2.31, p = .021), and were more substantial between the day and week tests (β = 0.14, SE = 0.05, Z = 2.97, p = .003).

For adults, there were benefits of both global and local prior knowledge. Vocabulary ability was a significant predictor of recall performance (β = 0.48, SE = 0.13, Z = 3.58, p < .001): adults with better vocabulary were better at recalling the new words. Pseudowords that had many neighbours in the English language were better recalled (M = .31, SD = .46) than words with only one neighbour (M = .19, SD = .39; β = 0.47, SE = 0.22, Z = 2.16, p = .031). However, the contrast between words with and without neighbours (no vs. one&many) was not significant (p = .19), suggesting that words with only one neighbour did not benefit from these more limited connections compared to words without neighbours (M = .16, SD = .37). There was also no interaction between vocabulary ability and neighbour benefit (p = .18), suggesting that all participants benefited from local connections to prior knowledge, and no evidence of a three-way interaction (pruned from model; p = .70).

Recognition

Recognition performance was highest immediately after story exposure (M = .74, SD = .44), with performance clearly above chance for adults (t(129) = 28.74, p < .001). Performance significantly declined by the later tests (β = -0.11, SE = 0.02, Z = -4.66, p < .001), but the decrease in performance between the day (T2: M = .70, SD = .46) and week (T3: M = .68, SD = .47) tests was not statistically significant. Vocabulary ability was again a positive predictor of performance (β = 0.35 SE = 0.10, Z = 3.54, p < .001), but there was no effect of word neighbours or any further interactions.

5.4.4 Experiment 2 Summary

Adults successfully learned more words from the story than children did in Experiment 1, as seen in the higher levels of performance in both the recall and recognition tasks. As in Experiment 1, adults improved in their recall performance after opportunities for consolidation, although their recognition performance showed a slight decline.

In line with previous studies using these pseudoword stimuli, adults were better at recalling words with many phonological neighbours compared to only one phonological neighbour. That is, even when encountering words in a story context, participants were still benefiting from local prior knowledge connections. We can consider that this activation may be slightly weaker than in explicit teaching contexts, given that there was no evidence of a benefit for words in the one-neighbour condition compared to the no-neighbour condition (i.e., with more limited connections to potential knowledge; c.f. Experiment 2, James et al., 2018). However, in contrast to our hypothesis, these weaker connections did not receive any greater benefit from offline consolidation, and the influence of phonological neighbours on memory remained stable across the week. This was the case for all participants regardless of vocabulary ability: participants with good vocabulary learned more words overall, but they were no different in their ability to consolidate new words or benefit from word neighbours.

5.5 General Discussion

Previous studies showed that children and adults could bootstrap new word-forms to existing knowledge to facilitate word learning in explicit teaching paradigms (e.g., James et al., 2018; Storkel & Lee, 2011). We tested whether similar facilitation would occur when pseudowords were encountered incidentally when listening to stories – during which opportunities for strategic comparisons to existing knowledge would be reduced – and what the time course of such effects would be. Learning new vocabulary from stories was weaker than in related experiments using explicit training regimes, in line with previous studies that have highlight the challenges of learning from stories without additional instruction (Wilkinson & Houston-Price, 2013). However, participants could successfully recognise some of the new word-forms immediately after learning, and recall strengthened with opportunities for consolidation for both children and adults. Adults showed a robust and persistent benefit for phonological neighbours in recall of the new words. However, children did not show benefits for local prior knowledge in this experiment, and in fact showed interference from phonological neighbours in recognising the new word-forms. We review each of these findings in turn, and consider the possible mechanisms underlying differences across experiments.

5.5.1 Benefits of offline consolidation for new vocabulary

Both children and adults showed improvements in their recall of the word-forms across the course of the week. These findings are consistent with proposed benefits for offline consolidation in strengthening knowledge of new word-forms, which have elsewhere been attributed to processes during sleep (e.g., Henderson et al., 2012). It is likely that processes of consolidation also benefited from repeat testing in the present experiments (Antony, Ferreira, Norman, & Wimber, 2017), either alongside or in interaction with sleep-associated benefits. However, previous studies have demonstrated that offline benefits can emerge without repeated tests in children (Henderson et al., 2013c) and that sleep-associated improvements outweigh repeat testing benefits (e.g., Dumay & Gaskell, 2007; Henderson et al., 2012), highlighting that memory consolidation mechanisms likely contribute to the observed improvements. The improvements in recall seen for children were larger across the week (mean improvement in proportion recalled of .19) than they were for adults (mean improvement of .04) in the present experiments. Whilst the different levels of initial learning performance prevent strong interpretations of this finding, it is interesting to note that similar child-adult differences were seen in previous studies (Chapter 3, Chapter 4) when initial learning was more comparable. Together, these studies offer support to the hypothesis that children benefit more from a period of offline consolidation than adults (James et al., 2018; Wilhelm et al., 2012).

5.5.2 The influence of local prior knowledge on learning new vocabulary

Adults showed superior recall performance for pseudowords with many phonological neighbours, compared to words with either no neighbours or one neighbour. This finding extends those of previous studies (e.g., Storkel et al., 2006) to show that adults benefit from local prior knowledge connections even without explicit instruction to learn the new words. Interestingly, adults did not appear to benefit from only a single neighbour in this experiment, indicating that access to local prior knowledge may have been slightly weaker than for intentional learning conditions (James et al., 2018). However the lack of benefit for one neighbour may also have been driven by the lower overall levels of learning in this experiment, leaving less variability to distinguish between as many experimental manipulations. Regardless, adults were still benefiting from local prior knowledge connections during the recall task.

For children however, there was no benefit of phonological neighbours for recall of the pseudowords encountered in stories. This perhaps suggests that children are less automatic in activating their related prior knowledge than adults and only engage prior knowledge when explicitly directing attention to vocabulary learning – as was the case in previous studies. The near-floor levels of performance prevent us from drawing strong conclusions in this regard, but this finding clearly warrants replication at more comparable levels of learning. Interestingly, children did show an effect of phonological neighbours in their recognition of the new word-forms, but this was in the opposite direction to previous findings: children showed poorer recognition performance for pseudowords with one/more neighbours than without. In many respects this finding reflects the broader word recognition and production literature: real words are recognised more quickly if they have few competing neighbours (e.g., Metsala, 1997) but are produced more accurately with they have many neighbours (e.g., German & Newman, 2004). Despite this, the most comparable experiment to this one using explicit teaching (James et al., 2018; Experiment 3) still showed a facilitation effect for neighbours in recognising the new word-forms in an identical task. Perhaps then the differences seen in the broader literature relate more to the learning context than the memory tasks for new pseudowords: drawing explicit attention to neighbours can enable individuals to benefit in forming a new representation, whereas implicit activation of word-form similarities otherwise causes interfering activation in memory.

5.5.3 The influence of local prior knowledge across consolidation

Previous explicit training studies showed an immediate benefit for phonological neighbours in children's memory for pseudowords, but this benefit diminished with opportunities for offline consolidation (James et al., 2018). That is, no-neighbour words appeared to be preferentially strengthened by offline consolidation processes, enabling children to recall them with comparable accuracy to many-neighbour pseudowords at memory tests one week later. The present study set out to test whether this would also be the case for pseudowords learned incidentally through stories, or whether weaker access to prior knowledge during encoding might lead to later-emerging benefits of word neighbours - after increased opportunities for engaging with existing vocabulary. However, we saw no changes in neighbour benefit across consolidation in either experiment. For adults, the benefit of local prior knowledge was present early, and did not change with consolidation. This finding was consistent with the previous adult explicit training study which did not show a statistically significant change in neighbour influence across the week. For children, there was no benefit of phonological neighbours in recall at any test point, and no changes in the effect for recognition. This might suggest that previous findings are limited to explicit training contexts. However, it remains an open question whether a similar pattern would emerge if children could be brought to similar levels of performance when learning vocabulary from stories.

5.5.4 Interactions between global and local prior knowledge

Existing vocabulary ability was a strong predictor of overall learning performance – across recall and recognition tasks, and for both children and adults. However, individuals with good global vocabulary knowledge showed no evidence of a superior benefit from phonological neighbours than those with poorer vocabulary, as one might predict if they are more likely to know more of the neighbours. We previously suggested that knowing a single phonological neighbour might be enough to support new learning, and hence that using stimuli with many possible neighbours might enable all individuals to access at least one during learning. However, unlike earlier experiments, there was no benefit for one neighbour overall or in interaction with vocabulary knowledge. This may suggest that those with good vocabulary were previously benefiting from single neighbours in a strategic manner, and/or that benefits for more limited connections to existing knowledge only emerge at higher levels of performance. These higher levels of performance could drive differences in two ways: by enabling increased variability to distinguish between conditions, but also by freeing the cognitive resources to engage with more limited prior knowledge connections.

5.5.5 What is the role for global vocabulary knowledge in lexical consolidation?

These experiments were designed to better understand the relationship between global vocabulary knowledge and offline improvements in the recall of new words (e.g., Henderson et al., 2015; James et al., 2017; Sénéchal et al., 1995). In general, we have failed to find evidence of this relationship when using paradigms that manipulated local prior knowledge of the stimuli: in Experiment 2 here and in all three of the previous explicit teaching experiments (James et al., 2018). However, Experiment 1 did show a similar pattern in one of the analyses: children with good global vocabulary improved more in the picture-form recognition task across the course of the week than children with poorer vocabulary. This improvement was seen across test points despite no further opportunities to learn the correct mappings between the items and their forms. Although we are cautious not to overstate this one finding, the clear lack of relationship across other experiments leads us to reconsider what the role of vocabulary ability in supporting consolidation may be.

The picture-form recognition task was the only measure that assessed semantic knowledge of the new words, suggesting that prior vocabulary knowledge may offer more support in consolidating the semantic mappings of the pseudowords. Whilst we used a picture-form recognition task in the previous explicit learning studies, the previous stimuli used unusual objects selected from a database (Horst & Hout, 2016) with limited opportunities for developing connections to existing semantic knowledge. The concepts in the present study related to known objects (e.g., a cactus-flavoured drink, a car for driving around space, space currency), and the story context perhaps provided enhanced opportunities to develop rich semantic connections with existing knowledge. These opportunities for developing rich semantic connections have previously been shown to better benefit children with good vocabulary knowledge (Henderson & James, 2018). Perhaps then – whilst related phonological knowledge clearly facilitates new word acquisition across a range of learning contexts – it is the opportunity for establishing rich connections with semantic knowledge that drives individual differences in consolidation. Future studies should make direct comparisons using the same referents across tasks and manipulate contextual variability to test hypotheses in this regard.

5.5.6 Challenges of assessing incidental vocabulary learning

The purpose of embedding pseudowords into story contexts was to attempt to assess activation of prior knowledge during learning without participants' use of explicit learning strategies. That is, do individuals still benefit from pseudoword *rafar*'s similarity to *radar* without explicitly making semantic connections between the novel and related words' meanings? In the present experiments, participants were informed they were taking part in a comprehension study, and were instructed to keep listening to the story even if they heard words they did not know. However, feedback data collected from adults at the end of the experiment revealed that 48 percent of participants still thought their task was to learn the new words (with a further 34 percent reporting being suspicious), and 18 percent reported using strategies to remember them during the story. These issues clearly limit our ability to infer more incidental vocabulary learning in the present experiments, but there are two ways of considering this issue in context. First, adults recruited into studies are acutely aware that they are participating in an experiment, and may engage strategies accordingly. In this instance, we suggest that similar studies could be improved by using fewer words, and in more challenging literacy contexts in which adults would ordinarily expect to be unfamiliar with some of the words. Our participants were also fully aware of the subsequent test sessions from the onset of the study – primarily to minimise drop-out under time and cost pressures – but avoiding this and simply re-inviting participants for further sessions would be another way to avoid use of memory strategies in between sessions (e.g., Hulme, Barsky, & Rodd, 2018).

However, a second – and not mutually exclusive – possibility is that adults are always more strategic in encountering new words in texts or discourse regardless of experimental context, and that their linguistic and cognitive expertise enable them to do so more than children regardless of the difficulty of the material. These strategic differences were likely exacerbated by our presentation of the written material alongside the spoken story for adults, giving participants the freedom to revisit unfamiliar words and attempt to match them with the pictures. Given that most adults will read silently at a pace faster than the spoken recording, they would have had the opportunity to engage these strategies without necessarily disrupting comprehension. In contrast, children had only the spoken story to listen to, which may have been more successful in minimising strategy use. Future studies should address differences in learning from stories with and without the accompanying text to better understand the use of strategies for adults. Alternatively, differences across adults and children could be addressed by making the vocabulary-learning nature of the task explicit to the children, and test whether the neighbour benefit re-emerges with this additional instruction.

5.5.7 Conclusions

In conclusion, the present study showed that adults still benefit from local prior knowledge connections when learning vocabulary via stories instead of via direct teaching. It may therefore be that adults activate and benefit from connections to prior knowledge automatically and implicitly, or that adults have good linguistic skills that enable them to approach vocabulary-learning strategically regardless of other task demands. Children, on the other hand, did not access prior knowledge benefits when encountering pseudowords in stories – suggesting on the converse that their benefits of prior knowledge are less automatic, and/or that increased linguistic demands in story contexts make vocabulary learning more challenging for this age group. Local prior knowledge benefits were not related to adults' global vocabulary ability, suggesting once again that – whilst local knowledge benefits are strong and robust – there is little variability in the support that they offer across individuals. Instead, children with good global vocabulary knowledge were better at consolidating semantic mappings for the newly learned words, suggesting that benefits of global prior knowledge may relate more to building semantic representations than to specific word connections. Understanding how these sources of variability contribute in different learning contexts will have important implications for best supporting word learning in those with weaker vocabulary ability.

Chapter 6. Identifying Children with Comprehension Difficulty

6.1 Introduction

Vocabulary knowledge is closely and causally related to reading comprehension ability (Beck, Perfetti, & McKeown, 1982; Clarke, Snowling, Truelove, & Hulme, 2010; Ouellette, 2006; Ricketts, Nation, & Bishop, 2007). It is clear from longitudinal studies that this relationship is bidirectional: having good vocabulary knowledge will facilitate comprehension of texts, and good reading comprehension will enable the learning and development of rich lexical representations (e.g., Verhoeven, van Leeuwe, & Vermeer, 2011). Unsurprisingly then, children identified as having reading problems specific to comprehension – rather than decoding – often have poorly developed semantic representations (e.g., Henderson, Snowling, & Clarke, 2013a). In Chapter 7, we present a final study examining the learning and consolidation of new vocabulary in poor comprehenders. The present chapter summarises the methods used to identify children with comprehension weaknesses in the context of good reading accuracy ability, and the challenges faced in doing so.

6.1.1 Who are poor comprehenders?

According to the Simple View of Reading (Hoover & Gough, 1990), successful reading is the product of two sets of skills: those which support successful decoding of written words from the page, and those which support linguistic comprehension. Whilst difficulties in the decoding domain are widely recognised – with 5-10% children being recognised as dyslexic in the UK – there is little formal support for individuals with comprehension difficulties *despite* good decoding skills (Hulme & Snowling, 2011). Specific comprehension difficulties also have an estimated prevalence of 5-10%, and constitute over half of the reading deficits that emerge later in the junior school years (Catts et al., 2012). Although both dyslexia and reading comprehension difficulty can be classed as “Specific Learning Disorder with impairment in reading” in the DSM-5, there are no procedures in place to identify those with comprehension difficulty within school settings, and little agreement over what such diagnostic criteria would be.

Poor comprehenders have been a group of interest to many researchers interested in causes and consequences of comprehension break-down, yet the lack of a coherent approach has proved problematic here also: substantial differences in the assessment types and selection criteria used present a challenge to drawing conclusions across studies. This chapter describes the issues we faced with assessment types and the severity and specificity of comprehension problems. To inform discussion and highlight the inconsistencies across previous research, we also reviewed the sampling approaches of published studies that categorised groups of poor comprehenders[8]. The approaches described throughout are based on those drawn from 84 journal articles (incorporating 88 samples of poor comprehenders).

6.1.2 The present datasets

To identify poor comprehenders and typically developing controls to take part in the vocabulary learning study, comprehension data were collected from eight different North and East Yorkshire schools across 2016-2018. These are presented as seven different datasets (summarised in Table 8) which each took a different screening approach at different times. Across all datasets, we screened 809 children from school years 3-7 (aged 7-12), with standardised comprehension assessment scores available for 626. All but Dataset E made use of opt-out consent procedures. Datasets B, C, F and G assessed the vast majority of children in recruited classes, and can thus be used for approximate prevalence estimates. Dataset A consisted of a subset of children selected from group-based screening on 261 children (detailed below), and so does not necessarily reflect broader reading profiles. Datasets D and E were collected on secondary school children, and were reliant on pupil motivation and organisation to attend the allocated sessions.

6.2 Standardised assessments for identifying poor comprehenders

To be identified as a poor comprehender, a child must not only demonstrate weak comprehension on a standardised assessment of reading comprehension, but also age-appropriate decoding skills to ensure that their poor comprehension is not a result of reading accuracy errors. For the purposes of our word learning experiment (Chapter

[8] Journal articles were sourced via a Web of Science search, using the terms “poor comprehenders” and “specific comprehension deficit”. We then manually selected papers identifying poor comprehenders with English as their first language, assessing children with poor comprehension in the context of adequate decoding and without other diagnosed learning/physical disabilities. The initial search was supplemented using Google Scholar searches.

Table 8. Summary of screening datasets incorporating relevant standardised measures of reading ability.

Dataset	School years	TOWRE valid *n*	TOWRE Mean SWE	TOWRE Mean PDE	YARC valid *n*	YARC Mean Comp	WASI Matrices valid *n*	WASI Matrices Mean nonverbal	NGRT valid *n*	NGRT Mean NGRT
A	3, 4	76	104.12	108.16	72	103.94	64	47.50	0	-
B	4	27	93.00	98.11	26	96.77	27	42.63	0	-
C	5	0	-	-	68	99.19	0	-	0	-
D	7	141	100.49	105.28	129	106.82	98	53.48	204	111
E	7	100	99.70	104.41	88	104.15	56	49.29	0	-
F	6	42	101.69	106.61	42	104.67	0	-	40	110
G	4, 5	130	101.26	104.50	126	102.46	130	43.58	0	-
Total		**516**	**100.72**	**105.01**	**551**	**103.45**	**375**	**47.62**	**244**	**111**

7), we restricted our search to native monolingual English speakers without reported learning difficulties or developmental disorders. Although children who speak English as an additional language often display a poor comprehender profile (e.g., Burgoyne, Kelly, Whiteley, & Spooner, 2009; Hulme & Snowling, 2011), it is not clear that their difficulties would be similarly reflected in learning and consolidating novel vocabulary (Kaushanskaya & Marian, 2009). In this section, we describe the standardised assessments we used to measure decoding and comprehension skills, before later discussing the appropriate thresholds for identifying comprehension impairment.

6.2.1 Reading comprehension

To assess reading comprehension, we used the York Assessment for Reading Comprehension (YARC) Passage Reading (Snowling et al., 2009) and its corresponding assessment for Secondary school children (Stothard, Hulme, Clarke, Barnby, & Snowling, 2010). The YARC is the most recently developed individual comprehension assessment in the UK, providing the most appropriate norms for our sample of children. For the primary edition, children's scores are based on two ability-appropriate passages. The child reads each passage aloud, and the experimenter corrects the child for any accuracy errors in order to help maintain comprehension. Providing the passage is read with sufficient accuracy (< 20 errors for Levels 3 and above), eight open-ended comprehension questions are asked upon completion of the passage. Children are able to look back at the text when answering the questions – minimising demands on memory – and experimenters are instructed to probe ambiguous responses. This process is repeated with a second passage determined by the child's accuracy and comprehension ability. Average scores are computed across both passages for reading accuracy (number of errors), reading rate (time taken to read the passage), and reading comprehension (questions correct), and standardised by age.

YARC Secondary scores are also based on two reading passages, but with two key differences. First, children may read the passages silently, which is more naturalistic for children of this age. This means that it does not provide a measure of text-based reading accuracy, and that alternative measures are required to ensure that the child has sufficient decoding skills to attempt the passage level. Second, there are more comprehension questions per passage (13 vs. 8), likely contributing to the assessment's higher reliability than the primary edition (Cronbach's alpha for relevant passages = .90; versus .48-.77 for primary version).

Whilst the YARC was the most appropriate comprehension assessment for our sample, it should be noted that no published studies have attempted to identify poor comprehenders on the basis of this measure in the UK; all previous studies used the Neale Analysis of Reading Ability (NARA; Neale, 1997). We return to this issue and its implications for ongoing research at the end of the chapter (Section 6.4.2).

6.2.2 Decoding ability

By definition, a poor comprehender's weaknesses are specific to their comprehension skills, and are not attributable to difficulties decoding words from the page. However, there lies discrepancies in the literature here too: many studies have used reading accuracy measured by the same comprehension test as reading, whereas others have used separate timed or untimed measures of word or nonword reading accuracy. These differences are not trivial, given evidence that different types of assessment vary in their independence from other linguistic skills (Keenan et al., 2014; Nation & Snowling, 1997). Our reading assessments incorporated two measures of decoding skill: text-based accuracy from the YARC, and the Test of Word Reading Efficiency (TOWRE-2; Torgesen, Wagner, & Rashotte, 2012). The TOWRE consists of two subtests that assess how many real words (Sight Word Efficiency, SWE) and nonwords (Phonemic Decoding Efficiency, PDE) a child can accurately read aloud in 45 seconds. Here, we summarise our decision to rely on the TOWRE for poor comprehender identification.

In reviewing the previous studies, 28% used text accuracy as a favoured measure for decoding skill (e.g., Cain & Oakhill, 1999; Stothard & Hulme, 1995), computed from children's errors when reading the texts aloud. Using the accuracy measure from the comprehension assessment reduces the testing demands of administered a separate assessment, and also follows the logic that impairments in reading comprehension cannot be attributed to errors in reading the text. However, the main focus of these earlier studies were on children in the early-to-mid junior school ages (ages 7-9 years) who still typically engage in shared reading, whereas reading aloud becomes unnatural in skilled readers and would likely interfere with other estimates of reading ability. A reliance on reading aloud makes it challenging to extend the same text-based accuracy approach across development. Indeed, our sample extended to secondary school children (up to age 12), and the YARC Secondary edition permits children to read the passages silently.

Furthermore, it can be argued that reading accuracy in context is confounded with comprehension ability: children with poor comprehension are less able to predict upcoming words to support their decoding online, and likely bring poorer vocabulary knowledge known to support reading of irregular words (Ricketts et al., 2007). To test this assumption in our dataset, we conducted Pearson's correlations between YARC Comprehension scores and each of the three decoding measures: text-based (YARC Accuracy), real word reading (TOWRE SWE), and nonword reading (TOWRE PDE). All measures of decoding were significantly correlated with performance on the reading comprehension task (

Table 9), and a Fisher *r-z* transformation was carried out to enable statistical comparisons between the correlations. As expected, the strongest correlate of reading comprehension ability was the text-based measure of reading accuracy, and this correlation was significantly stronger than those between comprehension and the SWE

Table 9. Correlations between measures of decoding and comprehension ability in subset of children with all three accuracy measures (n = 264)

	YARC Comprehension	YARC Accuracy	TOWRE SWE	TOWRE PDE
YARC Comprehension	-	.53	.38	.44
YARC Accuracy	-	-	.64	.77
TOWRE SWE	-	-	-	.78

($t = 3.36$, $p < .001$) and PDE ($t = 2.53$, $p = .01$) tests. This stronger correlation supports that - even for younger children for whom a text-based reading aloud measure is appropriate - it may be more difficult to identify children with discrepant accuracy and comprehension using these measures.

In our data, there did not appear to be a difference between the real word and nonword relationships with comprehension ability, in the primary ($t = 1.63$, $p = .1$) or secondary school ($t = 1.64$, $p = .1$) data. This differs from a previous study by Nation and Snowling (1997), which showed that measures of nonword reading were least dependent on linguistic comprehension. Many studies of poor comprehenders thus began to use nonword reading accuracy as a "purer" measure of decoding ability – an approach common to 25% of the reviewed papers, with a further 9% using nonword reading alongside text-based accuracy measures. In support of this approach, Keenan et al. (2014) showed that poor comprehenders with good nonword reading were more prevalent (8%) of than those with good real-word reading (5%) within a large sample of 1522 children. On the basis of these earlier studies, we decided to use the Phonemic Decoding Efficiency subtest of the TOWRE as our measure of decoding skill.

6.3 Categorisation and cut-offs

Across the reviewed studies, there was substantial variability in the severity of comprehension problems identified. At the most extreme, individuals identified as poor comprehenders had reading comprehension scores at or below the 5th percentile, corresponding to a standardised score of 75 (Spencer, Quinn, Wagner, & Practice, 2014; Wagner & Ridgewell, 2009). At their most accommodating, groups of poor comprehenders were simply weaker than the comparison group (e.g., Landi & Perfetti, 2007; Oakhill, Hartt, & Samols, 2005) and/or had comprehension skills that were unexpectedly low in relation to their accuracy ability (e.g., Tong, Deacon, Kirby, Cain, & Parrila, 2011; Yuill, 2009). These differences will necessarily affect the number of poor comprehenders identified within any given sample, and size of between group differences in comprehension ability. Our primary goal was to choose criteria that were comparable to previous groups of interest, but that also enabled a reasonable sample size.

Nearly half of the reviewed studies recruited poor comprehenders using 16th-25th percentile cut-offs (corresponding to standardised scores of 85-90), and over a third specified a minimum discrepancy for comprehension ability to fall below accuracy (e.g., one year, or one standard deviation). Our approach was similar: we initially set out to identify poor comprehenders with a standardised comprehension score below 90, and at least 10 standard score points below their accuracy ability (e.g., Nation, Cocksey, Taylor, & Bishop, 2010). To ensure that our sample of poor comprehenders were otherwise good readers – and thus that their primary deficit was likely regarding vocabulary and semantic knowledge - we recruited only children that had a TOWRE PDE standardised score ≥ 95 (e.g., Cutting et al., 2013). All children were required to have nonverbal ability within the average range, as measured by the Matrix Reasoning subtest of the Wechsler Abbreviated Scale of Intelligence (WASI-II; Wechsler, 2011).

However, due to the challenges described below, the maximum comprehension threshold was later relaxed to 100, recruiting poor comprehenders who scored below the test mean (as well as substantially below their accuracy ability). We also had to relax the nonverbal ability threshold, as will be discussed in Section 6.4.4.

6.4 Challenges in identifying poor comprehenders

6.4.1 Pre-screening to identify children at risk

Given that reading comprehension assessments are relatively lengthy to administer, one approach to recruiting poor comprehenders for research purposes has been to first use a group-administered comprehension assessment. This permits identification of children with weak comprehension skills to follow up with individual reading assessments. This pre-screening approach to poor comprehender identification was taken for two datasets: Dataset A used a listening comprehension adaptation of the NARA-II with primary school children (as in Clarke et al., 2010), and Dataset D used the New Group Reading Test (NGRT; Burge et al., 2010) with secondary school children.

Listening comprehension

Children in Dataset A originally took part in a large-scale word learning experiment (Chapter 3) that incorporated a number of shortened standardised measures adapted for whole-class administration. These included a multiple choice version of the NARA-II (Neale, 1997) adapted for listening comprehension and a subset of age-appropriate items from the British Ability Scales spelling subtest. The procedure and resulting number of poor comprehenders is summarised in Figure 12a. Considering that a 5-10% prevalence estimate (e.g., Nation & Snowling, 1997;

Stothard & Hulme, 1995) predicts 13-26 within the present sample, finding only two poor comprehenders falls strikingly short of anticipated rates. This low rate might be due to the mismatch in tests: whilst Clarke et al. (2010) used alternate forms of the NARA-II for group-based and individual screening, we proceeded to use the YARC for individual assessments. Indeed, comprehension is a complex and multifaceted skill, and correlations between comprehension assessments can be as low as .31 (Keenan, Betjemann, & Olson, 2008). The correlation between scores on the listening comprehension adaptation of the NARA and subsequent YARC comprehension scores in our sample was reasonably high ($r(70) = .6$, $p < .001$) and in line with correlations reported between the full standardised versions ($r = .62$; Snowling et al., 2009). However, there remains substantial variability that may underlie poor correspondence between our group-based measure and individual identification of comprehension difficulties. It may also be the case that our particular sample happened to have strong comprehension skills as a whole, such that the relatively poor children in the sample did not measure up as particularly weak on the standardised assessment. However, we chose not to use this method of pre-screening again.

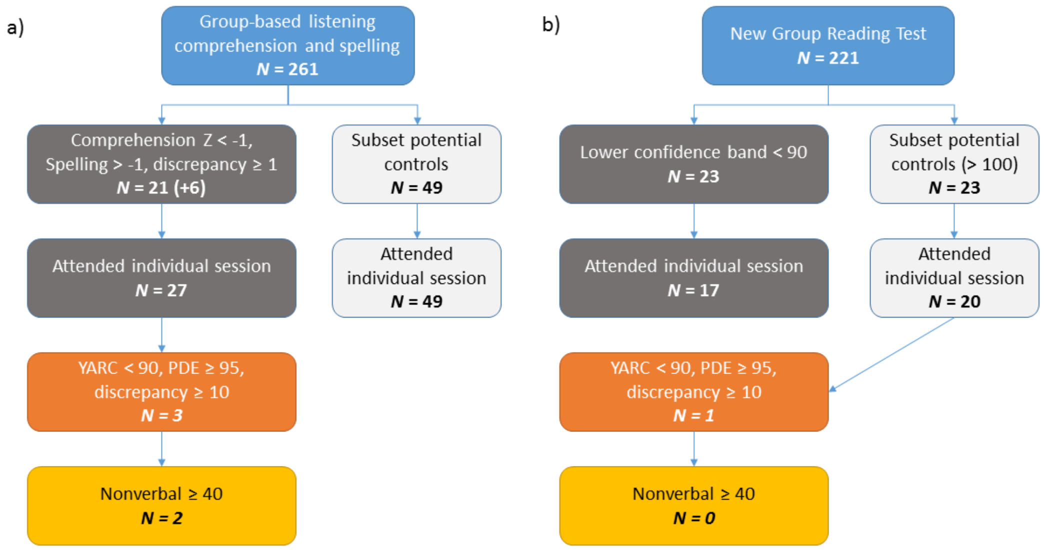


Figure 12. *Summary of group-based screening approaches for a) Dataset A using an adaption of the NARA-II for listening comprehension; and b) Dataset D using the New Group Reading Test. YARC: York Assessment for Reading Comprehension. PDE = Phonological Decoding Efficiency.*

New Group Reading Test

Dataset D included Year 7 children who each completed the NGRT through the school, and the data were released to us on a parental opt-out basis. The NGRT is a group-administered reading assessment standardised in the UK, which incorporates a sentence completion measure designed to measure decoding, and a passage comprehension task in which children answer multiple choice questions on increasingly difficult passages. The pre-screening procedure and results are summarised in Figure 12b, again highlighting limited success in identifying poor comprehenders. In the individual assessments, it became apparent that weak performance on the NGRT seemed to be more tightly related to weak decoding skills, and cases for discrepant reading comprehension below reading accuracy under our more liberal comprehension threshold (< 100) were also more commonly found in the potential control children (n = 3 of 5). In this small sample, it appeared that performance on the NGRT was as highly correlated with decoding ability ($r(35) = .57$, $p <.001$) as it was with comprehension ability measured by the YARC ($r(32) = .5$, $p = .003$).

Given that the NGRT did not appear to work well as a pre-screening measure, we returned to conduct individual assessments with the rest of the sample. There were 221 children in the year group, and we collected valid YARC measures on 128. The remaining children either did not want to participate, or could not complete the YARC due to weak reading skills and/or time constraints. The mean NGRT scores of children who attended the session were significantly higher (M = 113.91, SD = 12.54) than those who withdrew (M = 109.58, SD = 10.59), suggesting that those with weaker literacy skills were less likely to attend the individual session ($t(135) = 2.46$, $p = .015$). As such, we only identified one additional poor comprehender based on the original language criteria.

To better understand the relationship between performance on the NGRT and the YARC, we conducted additional screening with two Year 6 classes using both measures (Dataset F). In this broader sample of all children with both measures (n = 168), performance on the NGRT was more tightly related to comprehension on the YARC ($r(166) = .6$, $p < .001$) than it was to phonemic decoding efficiency ($r(166) = .41$, $p < .001$; $t = 2.49$, $p = .01$). However, it is clear that performance on the NGRT relies more heavily on decoding skills than the YARC, which shows much weaker correlations with the PDE at secondary school level ($r(126) = .21$, $p = .016$). Whilst in principle this should not be a concern for poor comprehender screening – the children in question have very good decoding skills – it does mean that the lower end of the NGRT standardisation sample likely consisted of generally poor readers. Furthermore, poor comprehenders have good decoding skills by definition, which may be sufficient for them to perform well on the NGRT.

6.4.2 YARC-ing up the wrong tree?

Regardless of whether we conducted pre-screening, we found very few poor comprehenders using the YARC. Nearly all UK studies of poor comprehenders have used the NARA for screening purposes, estimating a prevalence of 5-10 per cent when using this measure (e.g., Nation & Snowling, 1997; Stothard & Hulme, 1995). However, despite remaining an active research field elsewhere (e.g., Groen, Veenendaal, & Verhoeven, 2018; Ryherd & Landi, 2019), it is striking that no experimental studies of poor comprehenders in the UK have been published using the YARC. Although the two tests are similar in format, it appears that the YARC may be substantially less sensitive to detecting poor comprehenders than the NARA.

A handful of studies can speak to this question more directly. For example, Nation et al. (2010) administered a pre-publication version of the YARC to poor comprehenders selected on the NARA, which showed that - although the mean comprehension scores remained highly similar across the two tests - the variability in YARC comprehension scores was much higher than for the NARA. Although these differences may partly result from regression to the mean, the enhanced variability perhaps indicates that fewer of those individuals would lie below the threshold for comprehension impairment. In a recent study, Colenbrander, Nickels, and Kohnen (2017) indeed found that the rate of identifying poor comprehenders using the Australian edition of YARC was approximately one third of that when using the NARA within the same sample of children. Similarly, in analysing the UK YARC standardisation data, Hulme and Snowling (2011) showed that only 2.38% of children had poor comprehension in the context of good text-based reading accuracy (excluding children with EAL). These estimates are substantially less than previous prevalence estimates of 5-10% using the NARA. Whilst it is not appropriate to conclude which test is over- or under-sensitive from these data, poor comprehender identification certainly appears *different* across the two tests, and the resulting samples may be fundamentally different than in previous research. Our available sample cannot be used to provide precise prevalence estimates, but it is strikingly clear that poor comprehenders are no longer being identified at similar rates.

6.4.3 Decline in detection or prevalence?

It is important to stress that this challenge to identify poor comprehenders could be viewed positively. Whilst comparisons between the NARA and the YARC do imply test differences in detecting poor comprehenders, the reduced identification rates may also be in the context of a genuine decline in prevalence from earlier studies being conducted in the 1990s and early 2000s. The introduction of compulsory phonics teaching following the Rose Review (2006) has led to improvements in decoding skills, which may set a solid foundation for broader literacy development. Of course the converse could also be true: enhanced focus on phonics may leave other areas of literacy neglected and lead to *more* comprehension problems. However, recent reviews have stressed the importance of decoding skills within the context of the Simple View of Reading (Rose, 2009), and initiatives such as *Bookstart* promote broader oral language skills from a young age. Large-scale analyses are needed to determine the impact of these changes on the prevalence and/or nature of comprehension difficulties in modern-day classrooms.

6.4.4 How specific is a specific comprehension deficit?

Across 481 datasets that had both valid TOWRE and valid YARC measures, only 16 children met our original reading-based criteria for poor comprehenders (comprehension < 90, PDE ≥ 95, 10-point discrepancy; Table 10). Whether or not this prevalence is drawn from a representative sample, it is interesting to note that over a third of these children failed to meet the criteria for average-range nonverbal ability, as measured by the Matrix Reasoning subtest from the WASI-II (Wechsler, 2011). Many of the reviewed studies (20%) included average-range nonverbal ability as a recruitment criterion, to ensure that any group differences result from comprehension skill rather than more general cognitive difficulties (e.g., Nation, Clarke, Marshall, & Durand, 2004). We too had adopted this criterion in order to make our sample more comparable to previous studies of interest (Nation et al., 2007; Ricketts et al., 2008), and on the basis that previous datasets showed nonverbal ability to correlate with initial word learning performance. We later relaxed this threshold due to recruitment difficulty, including two poor comprehenders with below-average nonverbal ability, but did so similarly for the good comprehender group to ensure the groups were at least similarly matched (in line with 12% reviewed studies).

Our challenge in recruiting sufficient numbers of poor comprehenders with average-range nonverbal IQ is also reflected in a number of previous studies (e.g., Barnes, Stuebing, Fletcher, Barth, & Francis, 2016; Catts et al., 2006; Nation, Clarke, & Snowling, 2002; Nation et al., 2010). It seems unlikely that this association between weak nonverbal ability and comprehension difficulty is a coincidence, and may relate to the verbal strategies employed in matrix reasoning tasks. For example, aphasic patients show deficits for matrix reasoning items dependent on relational reasoning but not visual pattern matching (Baldo, Bunge, Wilson, & Dronkers, 2010). One might predict, therefore, that children with language weaknesses perform worse on complex analogical reasoning items compared to those requiring pattern completion. If true, these weaknesses could be considered an associated difficulty of poor comprehenders rather raising concerns over alternative causes to comprehension problems. Our matrix reasoning data were not entered on an item level to be able to address this question at present, but the language demands of different reasoning tasks remains an important avenue for determining specificity of impairments across measures.

Table 10. Number of children identified as poor comprehenders using different criteria.

Dataset	YARC & PDE *n*	YARC < 90, PDE ≥ 95, discrep. ≥ 10	YARC < 90, PDE ≥ 95, discrep. ≥ 10 nonverbal ≥ 40	YARC < 100, PDE ≥ 95, discrep. ≥ 10	YARC < 100, PDE ≥ 95, discrep. ≥ 10 nonverbal ≥ 40
A	71	2	1	11	7-8
B	26	4	3	7	5
C	-	-	-	-	-
D	127	2	0	12	8-10
E	88	5	2-4	12	5-8
F	42	0	0	4	0-4
G	127	3	2	12	9
	481	**16**	**8-10**	**58**	**34-44**

Notes. The number of children identified as poor comprehenders in each dataset, according to different selection criteria. Discrep. refers to YARC comprehension score being 10 or more points below TOWRE PDE score. Ranges are presented if nonverbal ability data is missing from some of the individuals.

6.5 Summary

We were interested in poor comprehenders as a group who – although clearly heterogeneous – typically show weaknesses in semantic aspects of language. As such, they present an interesting way of addressing individual differences in consolidation, indicated by previous vocabulary learning experiments with this group (Nation et al., 2007; Ricketts et al., 2008). Although using thresholds to classify children in this way is somewhat arbitrary, group contrasts can be useful in isolating variables of interest in small testing-intensive studies, and we hoped to make our sample comparable to those of earlier experiments to further our understanding of vocabulary acquisition in children with comprehension weaknesses. Unfortunately, this was not possible: despite an initial sample of over 800 children, we could not identify sufficient numbers of poor comprehenders using our chosen measures and we had to make compromises for both the severity and specificity of comprehension problems in our sample.

Chapter 7. Word Learning and Consolidation in Poor Comprehenders

All experiment pre-registrations, materials, data, and analyses are available on the Open Science Framework: https://osf.io/zqp8r

7.1 Abstract

Children with reading difficulties specific to comprehension often have vocabulary impairments, with pronounced difficulties on tasks that tax access to semantic knowledge. Extant evidence suggests that "poor comprehenders" can show initial word learning that is comparable to typically developing peers, but that relative impairments emerge at later follow-up tests. This retention difficulty is consistent with theories that propose weaker lexical consolidation in the context of impoverished semantic knowledge. To test this hypothesis directly, we tracked new word memory across wake and sleep to isolate processes of learning and consolidation in 8-to-12-year-old poor and good comprehenders. Each child took part in two encoding conditions in which they were taught 12 new words at the start (AM-encoding) or end (PM-encoding) of the day, alongside training on a nonverbal declarative memory task. Memory was assessed immediately, 12-, and 24- hours later, including stem completion, picture naming, and definition tasks to probe different aspects of new word knowledge. Sleep strengthened memory for the new word-forms, with improvements seen in stem completion and picture naming over sleep and post-sleep wake. Poor comprehenders were weaker than good comprehenders across all memory measures and – counter to our hypotheses - these relative deficits were apparent at encoding and persisted across consolidation. However, both comprehension groups were highly variable in existing vocabulary knowledge, and these individual differences predicted sleep-associated consolidation after an intervening day awake (and not if sleep could follow soon after learning). The results extend our understanding of poor comprehenders' word learning deficits to highlight persistent impairments across all aspects of new word knowledge, and provide new insights into the ways in which learning can be better timed to support those with poor vocabulary knowledge.

7.2 Introduction

Good vocabulary knowledge is a key contributor to comprehension success (Perfetti, 2007) and – in turn – successful comprehension permits the acquisition and development of new word knowledge (e.g., Verhoeven et al., 2011). Yet even in explicit vocabulary instruction, there lies considerable variability in the ease at which children learn new vocabulary. In attempting to understand differences in vocabulary acquisition, we must consider not only the factors that enable an individual to form a new word representation in memory, but those which enable this lexical representation to become consolidated as part of longer-term vocabulary. Understanding individual differences in both of these processes is critical to better targeting robust and long-lasting vocabulary instruction. One possible source of variation is in children's existing semantic knowledge, proposed to bolster the consolidation of new words (James et al., 2017). In the present study, we sought to understand these processes by comparing the learning and consolidation of new spoken vocabulary in children with good versus poor reading comprehension, who typically differ in lexical-semantic knowledge.

7.2.1 Learning and consolidating new vocabulary

Stages of vocabulary learning and consolidation are captured by the Complementary Learning Systems (CLS) account of new word acquisition (Davis & Gaskell, 2009). According to this model, two neural systems are engaged in the process of acquiring new vocabulary. The first of these systems is dependent on hippocampal and medial temporal lobe structures, required in forming an initial representation of a new word in memory. This new word representation is established by binding together a spoken form with associated semantic and syntactic information into a distinct episodic trace. Importantly, this initial memory formation can occur very rapidly as we encounter previously unknown words, without interfering with existing linguistic knowledge.

However, words in a proficient lexicon are proposed to be stored in a distributed and highly integrated fashion, enabling the rapid processing of incoming linguistic information (Davis & Gaskell, 2009; Gaskell & Marslen-Wilson, 1997). For a new word to become part of this second, neocortex-based system, a slower learning process must occur to carefully strengthen relevant vocabulary connections without disrupting existing knowledge. The CLS account proposes that this consolidation process happens as the hippocampus replays memory traces to the neocortex, gradually reducing hippocampal involvement in retrieving the new words (Davis et al., 2009).

Recent findings demonstrate that a large part of lexical consolidation may happen "off-line", during sleep. Indeed, numerous experiments across adults and children demonstrate greater improvements in explicit word knowledge after sleep compared to wake, as well as increased engagement of the new words with existing vocabulary knowledge (Dumay & Gaskell, 2007; Henderson et al., 2012; Tham et al., 2015). The magnitude of these improvements is associated with low-frequency neural oscillations occurring overnight (e.g., Smith et al., 2018; Tamminen et al., 2010), proposed to facilitate communication between hippocampal and neocortical memory systems (Staresina et al., 2015). Studies with adults have also used auditory cues to target replay of selected lexical associations during SWS, supporting causal interpretations of sleep-replay in consolidating new vocabulary (Schreiner & Rasch, 2016). Converging evidence thus suggests that sleep plays an "active" role in supporting consolidation of new vocabulary, and that comparing memory changes over sleep versus wake can inform us about consolidation processes.

7.2.2 Semantic knowledge in vocabulary acquisition

Whilst a role for sleep in vocabulary consolidation has been relatively well-established, less is known about factors that might influence this process. One factor that is proposed to support the learning and consolidation of new word-forms is the abundance of associated semantic information, forming an enriched lexical representation with many potential connections to existing knowledge. For example, McKague, Pratt, and Johnston (2001) presented 6-to-7-year-old children with pseudowords either in isolation (no semantics) or embedded in a story with semantic information. When tested using a free recall task immediately after learning, children recalled more pseudowords from the semantic training condition than form-only condition. In adults, Taylor, Plunkett, and Nation (2011) similarly showed that providing participants with definitions (versus phonological forms) facilitated their learning of new orthographic representations, and Havas et al. (2017) found that *familiar* objects (versus unfamiliar objects and no pictures) enabled quicker acquisition of new words. Together these studies suggest that semantic information facilitates even form-based components of word learning.

In assessing the longer-term advantages of semantic knowledge, Havas et al. (2017) showed only a marginally significant benefit for semantic knowledge on the overnight consolidation of new word-forms, and this was specific to word-forms that were less similar to the participants' native language. However, there is some suggestion that these semantic benefits in consolidating vocabulary are slower to emerge. Henderson et al. (2013c) taught two groups of children new science vocabulary (e.g., *hippocampus*), with one of the groups focusing only on the word form during training ("*hippocampus has three ps in it"*) and one receiving semantic information ("*hippocampus is a part of the brain that helps you to remember").* The two groups performed similarly in recalling the new word-forms immediately after training and 24-hours later, but the group who received semantic training continued to improve more across the week, outperforming the form-only group by the week follow-up assessment.

Given that Havas et al. (2017) found a benefit for only familiar and not unfamiliar objects, and that Henderson et al. (2013c) provided participants with equal linguistic material, it appears that it is not only the richness of information provided by semantic training paradigms that facilitates consolidation but the connections that this trained knowledge has to the learner's existing knowledge (James et al., 2017). To test this proposal, Henderson and James (2018) presented novel words to children across either two different stories or in the same story repeated twice, thus manipulating opportunities to draw upon connections with existing knowledge. The findings suggested that more variability in semantic information could lead to bigger improvements in form recall across consolidation, but only for those children who have more extensive vocabulary knowledge to capitalise upon. This highlights that the variability of semantic support in learning new vocabulary can also come from the learner as well as the learning environment (James et al., 2017). In the present study, we investigate vocabulary consolidation differences in a group of children who typically bring weaker semantic knowledge to language tasks: children described as *poor comprehenders*.

7.2.3 Vocabulary ability of poor comprehenders

Poor comprehenders are children who have at least age-appropriate phonological skills and reading accuracy abilities, but show relative weaknesses in their ability to access meaning from language (Nation & Snowling, 1998). Studies from the 1990s and 2000s estimated that approximately 5-10 per cent of children show such difficulties (e.g., Nation & Snowling, 1997; Stothard & Hulme, 1995), and that these comprehension problems frequently co-occur with poor vocabulary ability (e.g., Catts et al., 2006; Nation et al., 2010). Although good vocabulary knowledge is not sufficient for good comprehension and there are many possible reasons for comprehension to break down (Colenbrander, Kohnen, Smith-Lock, & Nickels, 2016; Oakhill et al., 2005), a wealth of evidence supports that the majority of poor comprehenders have weaker vocabulary than their typically developing peers, and that this performance gap widens throughout development (Cain & Oakhill, 2011). Furthermore, these weakness have been consistently demonstrated as specific to lexical-semantic rather than phonological components of word knowledge (see Landi & Ryherd, 2017, for a review).

One might anticipate that poor comprehenders' emergent vocabulary difficulties might result from their reduced engagement with literacy activities over time, in line with the well-cited "Matthew effect" (Stanovich, 1986). However, retrospective longitudinal studies reveal vocabulary weaknesses for poor comprehenders prior to the onset of literacy (e.g., Justice, Mashburn, & Petscher, 2013; Nation et al., 2010), suggesting that vocabulary might be a key and causal contributor to comprehension problems. Whilst there is some evidence that print exposure might also contribute to vocabulary difficulties for poor comprehenders (Cain & Oakhill, 2011), they also emerge as relatively impaired at vocabulary learning in experimental studies that control exposure to the new words. Some of these studies have supported that poor comprehenders are weaker at inferring new word meanings from context (Cain et al., 2004), although this finding has not been consistent (Ricketts et al., 2008). Importantly, poor comprehenders' weaknesses in vocabulary learning are not *restricted* to learning from texts, with evidence of poor vocabulary learning also arising from direct teaching paradigms (Cain et al., 2004; Nation et al., 2007).

7.2.4 Vocabulary consolidation in poor comprehenders

Given that poor comprehenders tend to have relatively poor vocabulary and that they show specific difficulties in accessing semantic knowledge (e.g., Nation & Snowling, 1998), one might anticipate that their word learning difficulties may intensify over a period of offline consolidation: if forming connections with neocortically based semantic knowledge facilitates consolidation of new words, then those with impoverished knowledge and/or weaker access to it will receive weaker benefits of offline consolidation than their peers. Two studies support this hypothesis. Nation et al. (2007) and Ricketts et al. (2008) trained poor comprehenders on new vocabulary, and assessed memory performance using tasks that required them to map the new word to its corresponding picture or definition. In both studies, poor comprehenders performed as well as the typically developing controls (matched on age, decoding skill, and nonverbal ability) when tested on the same day as learning, but relative weaknesses emerged by the week follow-up test. This pattern of performance is consistent with the proposal that poor semantic knowledge may constrain the broader consolidation of new vocabulary. Indeed, the children in Nation et al. (2007) were already performing more poorly in a more semantically demanding definitions task immediately after learning, whereas recall of the associated word-forms were only weaker at the delayed follow-up.

Two other approaches to understanding poor comprehenders' vocabulary weaknesses have also produced findings in line with proposed difficulties consolidating new vocabulary in long-term memory. Henderson et al. (2013a) examined poor comprehenders' access to the subordinate meanings of homonyms in a semantic priming task. Despite having explicit knowledge of the less frequent word meanings (e.g., *bank – river* versus *bank - money*), they did not access these meanings in speeded semantic tasks. This weaker semantic activation is consistent with a hypothesis that poor comprehenders' word knowledge is poorly integrated into the neocortical vocabulary system. Furthermore, in a neuroimaging study by Cutting et al. (2013), adolescent poor comprehenders showed abnormal engagement of hippocampal mechanisms during a simple lexical decision task. The authors suggested that one explanation for this finding could be that poor comprehenders have difficulty with consolidating new word representations into cortical structures, as would be predicted if existing semantic knowledge facilitates systems consolidation within the CLS model (McClelland, 2013). We take the first step in examining this hypothesis using a behavioural experiment of learning and sleep-associated consolidation processes.

7.2.5 The present study

The aim of this study was to investigate both the initial word learning and sleep-associated consolidation processes of poor comprehenders relative to good comprehenders, as a means of understanding contributions of semantic knowledge to vocabulary acquisition. Previous studies have not addressed consolidation over sleep and wake in this population, and have not always used tasks that probe in-depth semantic knowledge of new vocabulary. We taught children new spoken words in the morning or the evening, and tested their memory immediately, 12- and 24-hours later, enabling us to isolate memory changes in relation to sleep-associated consolidation processes. Memory was assessed using three tasks designed to probe different aspects of word knowledge: a stem completion task to assess memory of the new forms, a picture naming task to assess the form-meaning mapping, and definitions task to probe the richness of newly acquired semantic knowledge. These tasks enable us to assess the extent to which poor comprehenders' vocabulary impairments are specific to the semantic aspects of new word knowledge before and after opportunities for consolidation. By comparing memory across periods of wake and sleep, we aimed to assess whether poor comprehenders' vocabulary difficulties arise at the stage of consolidating new words into existing vocabulary, or whether they deteriorate before opportunities to do so. A nonverbal declarative memory task also enabled us to assess the specificity of any learning and consolidation difficulties to linguistic information. More broadly then, this study contributes to a growing literature on the importance of sleep for learning in development, allowing us to examine sleep-associated benefits for a multitude of tasks. Importantly, the sleep-wake design also allows us to directly compare how these benefits are influenced by a post-learning delay before opportunities to consolidate offline (i.e., when training commences in the morning relative to the evening).

7.3 Method

7.3.1 Hypotheses

Four hypotheses were pre-registered on the Open Science Framework (https://osf.io/4frxd/): 1) Poor comprehenders will display poorer semantic learning performance than control children (matched for age, reading accuracy, and nonverbal ability), as measured by a definitions task for new words learned; 2) Poor comprehenders will show broader word memory impairments (e.g., in stem completion, picture naming) after a period of consolidation; 3) If poor comprehenders' impairments are associated with problems during sleep-associated consolidation, then bigger differences in performance (vs. typically developing control children) should emerge after a period of sleep has occurred than over a period of wake; and 4) If learning and consolidation impairments are specific to language, then performance on a nonverbal declarative memory task should be equivalent across the two groups of children.

7.3.2 Participants

15 poor comprehenders and 15 good comprehenders were included in the study. Participants were 8-12 years old, and all were native English speakers with no reported learning, neurological, or sleep disorders. Participants were recruited following comprehension assessments conducted across eight different schools (809 children), with individual comprehension measures administered to 551. The study was approved by the University of York Psychology Ethics Committee, and children received a gift voucher to thank them for their participation.

All children invited to the study had average-good nonword reading (≥ 95), as measured by the Phonemic Decoding Efficiency subtest of the Test of Word Reading Efficiency (TOWRE-2; Torgesen et al., 2012). This criterion was important for ensuring that the poor comprehenders' reading difficulties could not be attributed to weak decoding skills. Poor comprehenders had a reading comprehension score that was below the test mean (< 100) on the York Assessment for Reading Comprehension (YARC; Snowling et al., 2009; Stothard et al., 2010), and at least 10 standard score points below the child's nonword reading score. Although we had initially aimed for a more stringent comprehension threshold (< 90), the threshold was relaxed due to recruitment difficulty, making our sample more akin to an "unexpected" poor comprehender approach (e.g., Tong et al., 2011). We were able to invite 59 children who met these criteria for the poor comprehender group. To ensure that group differences in comprehension were maximised after relaxing our poor comprehender criteria, we included good comprehenders whose comprehension score was above 100 and at or above their decoding ability. Note that we collected data from an additional 5 children for this group: 4 did not meet the at/above criteria (data available online), and one was excluded due to a scheduling issue.

It is also important to highlight that we relaxed our initial threshold for average-range nonverbal ability, as measured by the Matrix Reasoning subtest of the Wechsler Abbreviated Scale of Intelligence (WASI-II; Wechsler, 2011). The final sample included two poor and two good comprehenders with below-average matrix reasoning scores. Group profiles are presented in Table 11.

Table 11. Selection and background measures summarised by comprehension group

	Poor Comprehenders *(7m, 8f)*		Good Comprehenders *(8m, 7f)*		*t*	*p*
	M	*SD*	*M*	*SD*		
Age (years; months)	10; 04	1; 07	11; 0	1; 09	1.13	.269
TOWRE – Sight word[1]	102.20	10.37	106.40	8.77	1.20	.242
TOWRE – Phonemic decoding[1]	102.87	26.16	107.60	6.65	0.68	.507
YARC Accuracy[1, 3]	104.82	9.53	113.17	4.88	2.39	.030
YARC Rate[1]	106.07	13.28	113.13	7.12	1.82	.083
YARC Comprehension[1]	92.67	5.70	114.13	5.37	10.61	<.001
WASI Matrix Reasoning[2]	47.33	9.71	53.20	8.23	1.79	.085
WASI Vocabulary[2, 4]	49.93	10.15	61.43	5.56	3.82	<.001

Notes: [1]Standardised score (M = 100, SD = 15); [2]T-score (M = 50, SD = 10); [3]Only relevant for Primary edition, data from 6 GCs and 11 PCs only; [4]Data missing from one good comprehender due to time constraints.

7.3.3 Stimuli

Two lists of 12 words were created, each consisting of rare or unfamiliar living things that were unlikely to be known to the children (Appendix D1). Each list contained three exemplars from four different categories (e.g., three types of bird, three types of tree, etc.). This was designed to encourage in-depth semantic learning, as children were required to learn distinctive features to distinguish between other known and new exemplars. The two lists were matched on the number of syllables (List 1 M = 2.5; List 2 M = 2.67; p = .43), number of phonemes (List 1 M = 5.92; List 2 M = 5.42; p = .23) and biphone probability (List 1 M = .008; List 2 M = .003; p=.26; computed using CLEARPOND, Marian, Bartolotti, Chabal, & Shook, 2012).

We sourced an illustration for each item using a web-based image search, and presented each image centred on a plain white background during training. We also sourced three photographs per item for the picture naming task, and collected ratings from adults on the similarity of each photograph to the training image. These ratings enabled us to create three photo lists matched on rated similarity to the training image, for use at each test point. A fourth photo list was created for a delayed follow-up test.

7.3.4 Design and procedure

Each child took part in two sets of learning and test sessions, separated by at least one week. For each set, they completed an initial encoding session lasting approximately 45 minutes, either at the beginning (AM) or end of the day (PM). These were followed by three sets of memory tasks administered immediately, ~12- and ~24-hours later, enabling us to track memory changes across wake and sleep (Figure 13). AM sessions were administered as early as possible (range: 08:23-10:04), and usually took place at school. PM sessions were scheduled for as close to the child's bedtime as was practical (range: 16:08-21.26), and were usually administered in the child's home. Actiwatch data showed that the PM training session was started a mean of 2.80 hours before sleep onset (range 0.76-4.40 hours); compared to an average of 12.82 hours for the AM condition (range 11.78-15.23 hours). As is typical for children of this age (e.g., Henderson et al., 2012), the time elapsed overnight (M = 13.39; SD = 1 hours) was significantly longer than across the day (M = 10.03, SD = 1.30 hours). The poor comprehenders had slightly more time between overnight tests (M = 13.8, SD = 0.66 hours) than good comprehenders (M = 13.1, SD = 0.96 hours), but this time gap did not correlate with performance change in either learning condition for any task.

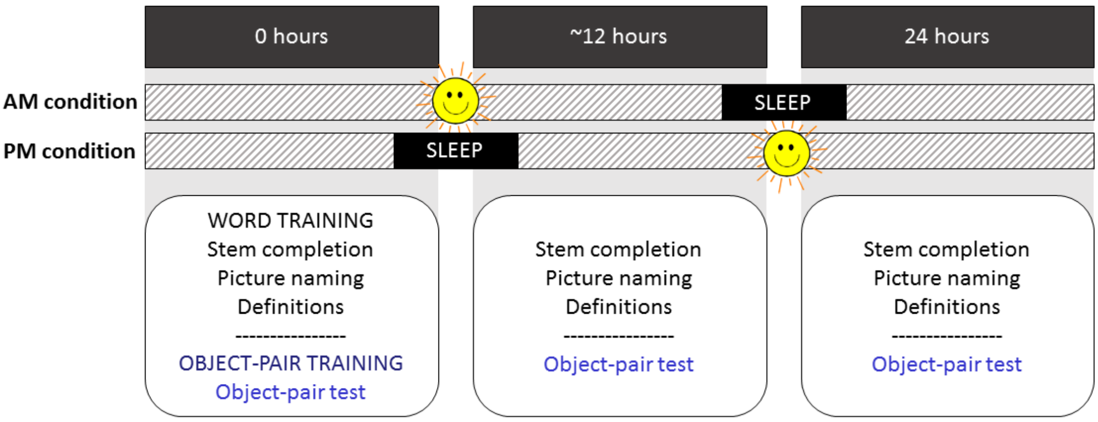


Figure 13. Schematic of overall experimental design. All children sat both an AM and PM learning condition, separated by at least one week.

The learning condition administered first (AM/PM encoding) and word lists assigned to each condition were counterbalanced across participants, and there was no difference in learning across word lists or weeks of the study (all *p*s > .17). The two conditions were separated by a period of at least one week (median = 7.41 days; range = 6.4–21.43 days), during which the child kept a sleep diary and had their sleep/wake times monitored using a Motionlogger Actigraph (Ambulatory Monitoring, Inc.).

All learning and test tasks were programmed using OpenSesame (version 3.1.9; Mathôt, Schreij, & Theeuwes, 2012), and administered on a laptop. A headset was used for audio presentation of items and recording of vocal responses.

7.3.5 Word exposure phase

Children were instructed that they were going to learn 12 new words, and that they must try their best to learn them. We trained and tested only spoken word learning; no orthography was presented at any stage. Children heard each of the new word-forms 19 times (13 alongside the corresponding image) across five training tasks, administered in the order below. Within each task, the order of item presentation was randomised.

Familiarity check. Children listened to each of the new words and were asked to say whether they had heard the word before and, if so, what its meaning was. If the child provided a relevant definition, this item was removed from analyses on a by-participant basis (*n* = 9). No child knew more than one word per list.

Form repetition. Children listened to each of the new words again, and repeated the word aloud.

Picture naming. Children heard each word alongside its illustration, and repeated the word aloud to name the picture themselves. They repeated this task a second time.

Multiple choice tasks. A picture round involved children listening to each word and selecting which of two pictures it matched using a key press response. An audio round involved choosing which of two words heard matched the picture on screen. For the first round of each, the incorrect option was a learned item from a different semantic category. For the second round of each, the incorrect option was a learned item from the same semantic category, aimed at promoting deeper semantic learning. Feedback was presented on all trials, providing the correct name of each of the presented images for the picture trials, and the correct name for the single image on the audio trials.

Delayed picture naming. Children heard each of the words, and were instructed to try to remember the correct picture. The correct picture appeared after 2.5 seconds, and children repeated the word form aloud to name the picture.

7.3.6 Word test phase

The same three test tasks were administered at each test point. Children provided a rating of sleepiness (1-10) at the start of each session, and completed the tasks in the order below. Within each task, order of items was randomised. There were two sessions of missing data: one child's vocal responses did not record for one session (missing data for the stem completion and picture naming tasks only), and another child was absent for one session.

Stem completion. A stem completion task was used to assess memory for the new word-forms. Children were presented with the starting sound of each word (incorporating the first consonant and vowel sound), and were asked to try to remember the remainder of the new word they had learned, and say the word aloud. Partial attempts were encouraged. Each response was voice-recorded and scored offline for accuracy (1, 0) using Check Vocal (Protopapas, 2007).

Picture naming. To assess memory for the form-meaning mapping, children were presented with a picture of each item and were asked to name the picture aloud as quickly as possible. An initial round used previously unseen photographs (with photo list assignment counterbalanced across participants), designed to probe generalisation of new knowledge and minimise repeat testing influences. A second round used the same images encountered at training. Each response was voice-recorded and again scored offline for accuracy (1, 0), as well as response time (ms).

Definitions. To probe for more explicit and rich semantic knowledge of the new items, children heard each of the words presented through the headphones, and were asked to tell the experimenter about the living thing they had learned about. Responses were transcribed by the experimenter, and later scored by an independent scorer (blind to condition) for semantic category and distinctive feature (maximum of 2 points per item). Where only one of these was provided, or the feature was generic to more than one item, the experimenter probed once for further information.

7.3.7 Object-pair location task

To compare the learning and consolidation of vocabulary to nonverbal declarative memory, a 2D object-pair location task was also administered (Wilhelm et al., 2008). For this task, 10 pairs of objects were presented across two locations on a 4x5 grid, and children had to remember the locations of each pair. We recreated the task from Henderson et al. (2012) and also developed a second version, which was again counterbalanced across order and AM/PM encoding condition. The stimuli were colour illustrations of easily nameable animals and objects, each with monosyllabic high frequency names (e.g., *drum, sheep*).

Learning phase. For the first round of training, children were instructed to watch the pairs on the grid and try to remember the location of each pair. For each of the 10 pairs, the first picture emerged at a grid location, and was followed by its matching picture 1000 ms later. Both pictures remained on the grid for 3000 ms. A 3000 ms inter-trial interval followed before presentation of the next pair. After all pairs had been viewed once, a second learning block involved testing with feedback. For each object-pair, one object would appear at its location on the grid, and the child used the mouse to click on the square where they thought the matching picture was (no timeout). Following their response, a sound was played to indicate whether their response was correct or not, and the correct pair location was displayed for 1000 ms. The inter-trial interval was 1000 ms.

Test phase. The test phase was identical to the second learning block, except that children received no feedback as to whether their response was correct. After selecting their answer, a sound played to register their response, and the task moved onto the next trial after 2000 ms. Note that this procedure diverges from Henderson et al. (2012), in which feedback was presented during the test trials (as in the second block of the learning phase). This decision was made in attempt to isolate processes of consolidation across sleep and wake in a way that was more comparable to the vocabulary tasks, minimising further learning opportunities.

7.3.8 Delayed follow-up

To test longer-term retention of the new information – particularly given that previous studies showed later-emerging semantic influences on vocabulary learning over longer time periods (e.g., after a week in Henderson et al., 2013) – we also administered a delayed follow-up session for all memory tasks approximately 1-2 months later. Given the challenges of scheduling around school holidays, there was substantial variability in the delay for each child (range: 4.09–10.77 weeks). However, the difference in delay was not statistically significant between comprehension groups. Furthermore, there was no correlation between the length of delay and change in performance for any dependent variable.

7.3.9 Analyses

Data from each task were analysed using mixed effects models, fitted using *lme4* (Bates et al., 2015b) and *ordinal* (Christensen, 2015). For the main analyses, we entered comprehension group (poor comprehender vs. control), encoding time (AM/PM) and test session (0-, 12-, 24-hour) as fixed effects, alongside all interactions between them. For the picture naming task, we also included picture type (novel, trained) as an additional fixed effect. All fixed effects were deviance coded to enable interpretation of the model according to the overall mean. The three-level factor of test session was coded to contrast 0-12 hour and 12-24 hour tests, enabling direct interpretation of interactions with encoding time.

All fixed effects were entered into an intercepts-only model in the first instance, and higher-order interactions that did not contribute to model fit ($p > .2$) were pruned to enable a more parsimonious model (Bates et al., 2015). We then incorporated random slopes using a forward best path approach (Barr et al., 2013), retaining only random slopes justified by the data under a liberal threshold ($p < .2$). Full model details are included in Appendix D (D2-D11), and details of the modelling process available on the OSF (https://osf.io/nyat5).

The delayed follow-up data were analysed in separate models, using the same principles described above. For these models, test session contrasted the 0-hour test point with the delayed follow-up scores.

7.4 Results

7.4.1 Definitions

In support of the hypothesis that poor comprehenders would show weaker semantic learning than good comprehenders, poor comprehenders averaged significantly lower definition point scores per item (M = 0.98, SD = 0.83) across test sessions than good comprehenders (M = 1.29, SD = 0.79; β = 0.48, SE = 0.2, Z = 2.42, p = .016). There were no changes in performance across test sessions for either AM or

PM encoding times (all *p*s > .15), and no consolidation-related differences between comprehension groups (pruned from model, $p = .679$).

The comprehension group difference in performance was maintained at the delayed follow-up test ($\beta = 0.45$, $SE = 0.21$, $Z = 2.11$, $p = .034$), with good comprehenders scoring better ($M = 1.02$, $SD = 0.88$) than poor comprehenders ($M = 0.78$, $SD = 0.84$) at the delayed test. There was a decline in performance between initial training ($M = 1.09$, $SD = 0.82$) and the delayed follow up ($M = 0.90$, $SD = 0.87$; $\beta = -0.30$, $SE = 0.10$, $Z = -2.85$, $p = .004$), but the size of this decline did not differ between comprehension groups.

7.4.2 Picture naming

Accuracy

For the picture naming task, picture type (novel photograph vs. trained illustration) was also entered into analyses, and a summary of all predictors is presented in Table 12. As with the definitions task, poor comprehenders were less accurate ($M = .25$, $SD = .43$) than good comprehenders ($M = .38$, $SD = .49$) at naming the pictures overall. The training images were also named more accurately ($M = .34$, $SD = .47$) than the novel photographs ($M = .29$, $SD = .45$), and this difference was consistent across comprehension groups and test sessions. There were significant improvements across all three test sessions and, importantly, an interaction between encoding time and the 0-12-hour change. In line with the hypothesis that sleep is beneficial for offline consolidation, there was a larger improvement for the PM-encoded items that featured sleep between the first and second test than for the AM-encoded items (Figure 14). Sleep later improved picture naming for the AM encoding time at the 12-24 hour test sessions, and equivalent improvements were seen during this time for the PM-encoded items. There were no significant three- or four-way interactions (pruned from the model, $p = .931$).

Picture naming accuracy was well-maintained by both groups at the delayed follow-up test, retaining an overall comprehension group difference ($\beta = 0.56$, $SE = 0.24$, $Z = 2.30$, $p = .021$) with no significant change in accuracy across the two test sessions (0-hour: $M = .25$, $SD = .43$; delayed: $M = .26$, $SD = .24$; $p = .969$). As with the main analyses, training images were named more accurately ($M = .28$, $SD = .45$) than novel photographs ($M = .24$, $SD = .43$; $\beta = 0.17$, $SE = 0.05$, $Z = 3.21$, $p = .001$). There was a significant interaction between encoding time and test session ($\beta = 0.24$,

Table 12. Predictors of picture naming accuracy for the main 24-hour analysis.

Fixed effects	β	*SE*	*Z*	*p*
(Intercept)	-1.39	0.36	-3.88	<.001
group	**0.56**	**0.25**	**2.20**	**.028**
pictureType	**0.21**	**0.07**	**3.14**	**.002**
learnTime	0.18	0.17	1.08	.281
time(0-12)	**0.48**	**0.11**	**4.35**	**<.001**
time(12-24)	**0.49**	**0.10**	**4.71**	**<.001**
group*pictureType	-0.01	0.04	-0.17	.863
group*learnTime	-0.08	0.14	-0.53	.596
group*time(0-12)	-0.17	0.11	-1.53	.126
group*time(12-24)	0.00	0.10	0.01	.993
pictureType*learnTime	-0.02	0.04	-0.46	.648
pictureType*time(0-12)	-0.05	0.11	-0.46	.647
pictureType*time(12-24)	0.00	0.10	0.00	.997
learnTime*time(0-12)	**0.62**	**0.11**	**5.78**	**<.001**
learnTime*time(12-24)	-0.05	0.10	-0.52	.606

Note. Model formed from 4220 observations, collected from 30 participants across 24 items. The model include by-participant random slopes for learnTime, and by-item slopes for group, learnTime and pictureType. Three- and four-way interactions were pruned from the model (χ^2 = 3.68, p = .93).

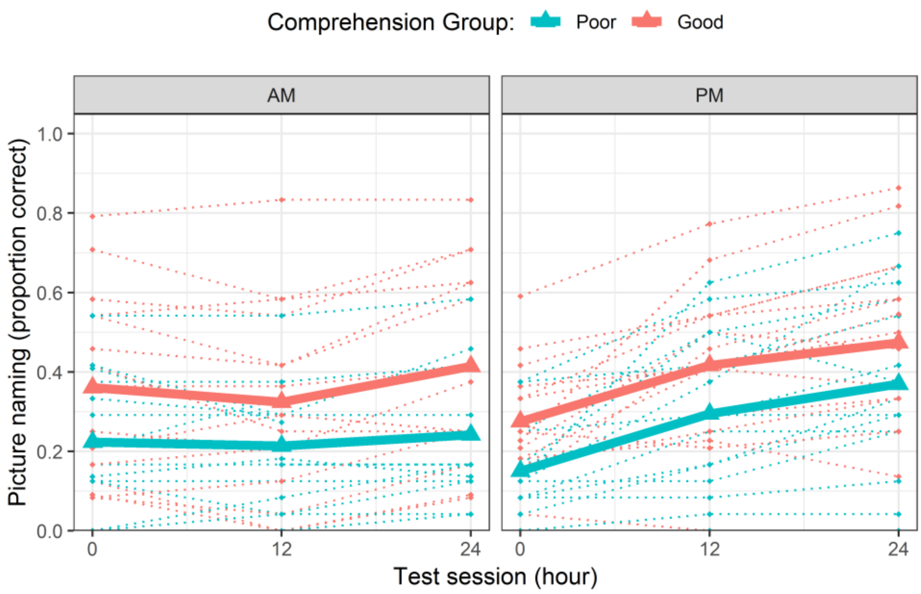


Figure 14. Mean picture naming accuracy across the 0-, 12-, and 24-hour tests for the AM and PM encoding sessions separately, averaged across the two picture types. Dotted lines mark performance for each participant, with thick lines representing mean scores per comprehension group.

SE = 0.06, *Z* = 4.23, $p < .001$): performance improved from PM encoding (0-hour: *M* = .21, *SD* = .41) to the delayed test (*M* = .28, *SD* = .45), whereas there was a decline in performance from AM encoding (0-hour: *M* = .30, *SD* = .46) to the delayed test (*M* = .24, *SD* = .43).

Response time

We analysed response time data from the accurate responses only, and removed trials that had either a prolonged onset or were preceded by vocalizations indicating earlier retrieval (*n* = 56). The initial model featured skewed residuals, and thus a Box-Cox transform was applied to the data to improve normality (raw scores are reported for ease of interpretation). The model summary is presented in Table 13. Training images (*M* = 1531 ms, *SD* = 1137 ms) were named more quickly than the novel photographs (*M* = 2055 ms, *SD* = 1513 ms) – although the training images were always presented second during testing and so also reflected a second retrieval attempt. Response time decreased across all three test sessions. This decrease in response time interacted with encoding time, such that there was a greater reduction between 0-12 hours for the PM encoding time than the AM encoding time, and vice versa for 12-24 hours (Figure 15). Across both encoding conditions therefore, periods of sleep always facilitated retrieval time (*M* = -477 ms) more than periods awake (*M* = 40 ms). However, there were no group differences for good versus poor comprehenders overall or in interaction with any other variable.

In analysing the delayed picture naming data, the effect of picture type on naming speed was maintained ($\beta = -1.09$, *SE* = 0.12, *t* = -8.79, $p < .001$), with training images being named more quickly (*M* = 1663 ms, *SD* = 1299 ms) than novel photographs (*M* = 2429 ms, *SD* = 2134 ms). There was weak statistical evidence for a decline in response times from the 0-hour to the delayed tests ($\beta = -.37$, *SE* = 0.19, *t* = -1.92, *p* = .070), but this was in the context of an interaction with comprehension group ($\beta = 0.31$, *SE* = 0.13, *t* = 2.43, *p* = .015). In contrast to our hypotheses, poor comprehenders showed *bigger* reductions in response times (0-hour: *M* = 2149 ms, *SD* = 1454 ms; delayed: *M* = 1894 ms, *SD* = 1923 ms) than good comprehenders (0-hour: *M* = 1980 ms, *SD* = 1567 ms; delayed: *M* = 2022 ms, *SD* = 1996 ms) over this period. There was also a significant interaction between group and encoding time ($\beta = -0.40$, *SE* = 0.13, *t* = -3.01, *p* = .003), with poor comprehenders faster to respond in

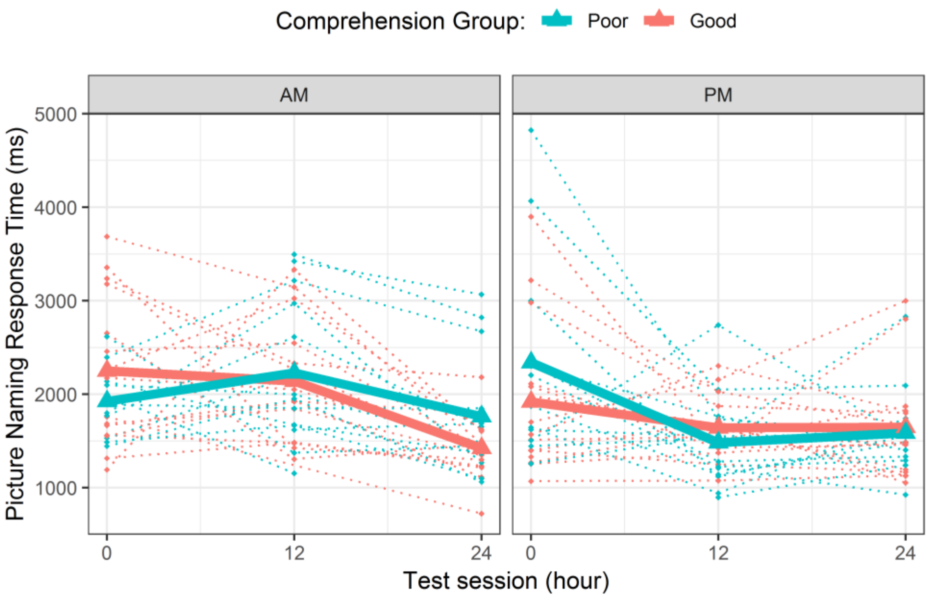


Figure 15. Mean picture naming response times across the 0-, 12-, and 24-hour tests for the AM and PM encoding sessions separately, averaged across the two picture types. Dotted lines mark performance for each participant, with thick lines representing mean scores per comprehension group.

Table 13. Predictors of picture naming response times in the main 24-hour analysis.

Fixed effects	β	*SE*	*t*	*p*
(Intercept)	1406.75	0.30	4722.84	<.001
group	-0.30	0.20	-1.49	.148
pictureType	**-0.80**	**0.06**	**-12.65**	**<.001**
learnTime	**-0.21**	**0.10**	**-2.11**	**.047**
time(0-12)	**-0.80**	**0.17**	**-4.82**	**<.001**
time(12-24)	**-0.74**	**0.15**	**-5.09**	**<.001**
group*pictureType	-0.02	0.06	-0.34	.735
group*learnTime	-0.02	0.07	0.27	.790
group*time(0-12)	0.25	0.16	1.50	.135
group*time(12-24)	-0.18	0.14	-1.26	.208
pictureType*learnTime	0.06	0.06	1.04	.300
pictureType*time(0-12)	-0.21	0.16	-1.30	.193
pictureType*time(12-24)	0.23	0.15	1.66	.097
learnTime*time(0-12)	**-0.66**	**0.16**	**-4.10**	**<.001**
learnTime*time(12-24)	**0.46**	**0.14**	**3.19**	**.001**

Note. Model formed from 1271 observations, collected from 30 participants across 24 items. The model included by-item slopes for learnTime only. Three- and four-way interactions were pruned from the model (χ^2 = 4.61, p = .87). The analysis was performed on transformed data.

the AM (*M* = 1855 ms, *SD* = 1176 ms) versus PM (*M* = 2183 ms, *SD* = 1176 ms) encoding condition, and the opposite trend for good comprehenders (AM: *M* = 2122 ms, *SD* = 2000 ms; PM: *M* = 1870 ms, *SD* = 1530 ms). However, it should also be noted that poor comprehenders contributed fewer trials to these analyses (due to their lower accuracy), and so their estimates may be less reliable.

7.4.3 Stem completion

As with the other tasks, there were significant improvements in memory for the new word forms across test sessions (see Table 14 for model summary), and this improvement interacted with encoding time for the 0-12 hour tests: items that were learned in the evening improved more between the first two sessions than items that were learned in the morning (Figure 16). As with picture naming accuracy, there was no interaction between encoding time and test session for the 12-24 hour tests: task performance improved across the 12-24 hour tests for both the AM and PM encoding conditions. The data did not support the hypothesis that poor comprehenders would show broadening impairments with consolidation on this task: poor comprehenders showed weaker recall overall (*M* = .29, *SD* = .46) than good comprehenders (*M* = .42, *SD* = .49), but there were no interactions with test session or encoding time.

In analysing the delayed test for the stem completion task, there remained an overall comprehension group difference in recall (β = 0.39, *SE* = 0.16, *Z* = 2.38, *p* = .017), with good comprehenders (*M* = .38, *SD* = .49) outperforming poor comprehenders (*M* = .26, *SD* = .44) across sessions, but there was no significant change in performance over time. There was an interaction between encoding time and test session (β = 0.19, *SE* = 0.07, *Z* = 2.78, *p* = .005): learning was poorer in the PM-encoding condition (*M* = .26, *SD* = .44) but improved by the delayed test (*M* = .35, *SD* = .48), whereas the higher performance in the the AM-encoding condition (0-hour: *M* = .35, *SD* = .48) showed a slight decline across this period (delayed: *M* = .33, *SD* = .47).

7.4.4 Object-pair location task

This nonverbal task was designed to test the language specificity of poor comprehenders' difficulties. In contrast to our hypotheses, poor comprehenders also performed more poorly on this task (*M* = .38, *SD* = .49) across test points than good

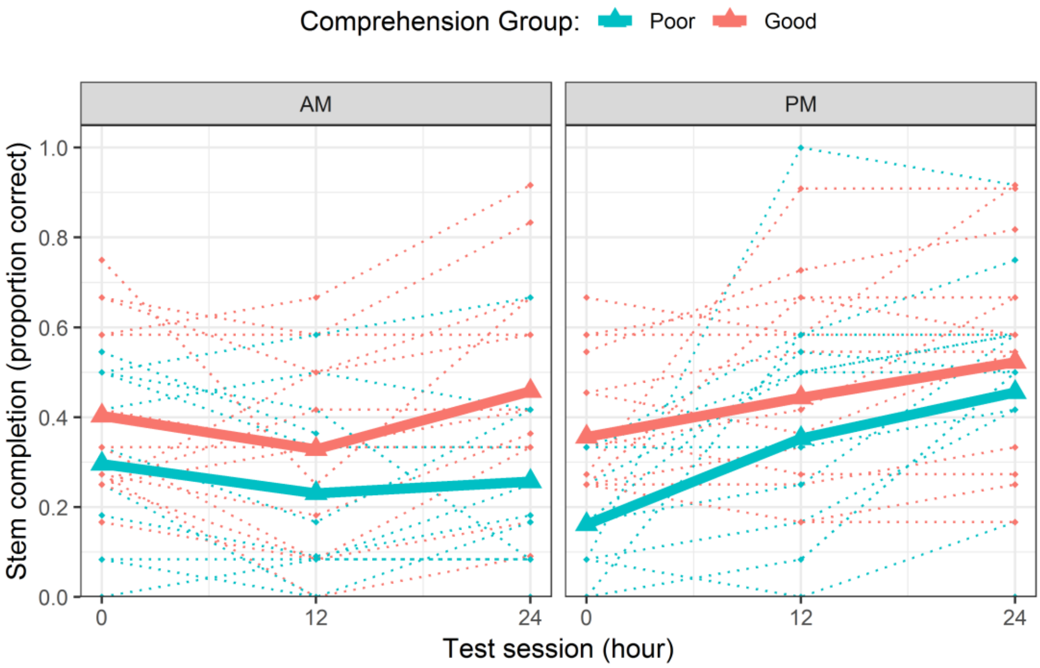


Figure 16. Mean stem completion accuracy across the 0-, 12-, and 24-hour tests for the AM and PM encoding sessions separately. Dotted lines mark performance for each participant, with thick lines representing mean scores per comprehension group.

Table 14. Predictors of stem completion accuracy in the main 24-hour analysis.

Fixed effects	β	*SE*	*t*	*p*
(Intercept)	-0.98	0.31	-3.14	.002
group	**0.47**	**0.22**	**2.17**	**.030**
learnTime	0.14	0.14	0.98	.325
time(0-12)	**0.29**	**0.14**	**2.07**	**.038**
time(12-24)	**0.56**	**0.14**	**4.06**	**<.001**
group*learnTime	-0.02	0.12	-0.18	.860
group*time(0-12)	-0.25	0.14	-1.80	.072
group*time(12-24)	0.08	0.14	0.60	.546
learnTime*time(0-12)	**0.73**	**0.14**	**5.11**	**<.001**
learnTime*time(12-24)	0.01	0.14	0.10	.924
group*learnTime*time(0-12)	-0.23	0.14	-1.62	.105
group*learnTime*time(12-24)	-0.17	0.14	-1.23	.220

Note. Model formed from 2109 observations, collected from 30 participants across 24 items. The model included by-participant slopes for learnTime and by-item slopes for group and learnTime.

comprehenders (*M* = .49, *SD* = .50; β = 0.30, *SE* = 0.11, *Z* = 2.67, *p* = .008; Figure 17). There was a general deterioration in performance between the 0-hour and 12-hour tests (β = -1.61, *SE* = 0.14, *Z* = -11.16, *p* < .001), and this again interacted with encoding time (β = 0.31, *SE* = 0.14, *Z* = 2.16, *p* = .031). There was a smaller decline in performance for the PM-encoded condition that featured sleep between the 0-hour (*M* = .59, *SD* = .49) and 12-hour (*M* = .35, *SD* = .48) tests than there was for the AM-encoded condition (0-hour: *M* = .66, *SD* = .47; 12-hour: *M* = .32, *SD* = .47). However, there was no change in performance between 12-24 hours, alone (*p* = .999) or in interaction with encoding time (*p* = .312), suggesting no further benefits for post-sleep wake, or for sleep to recover information lost from morning.

The comprehension group difference was maintained in the follow-up analyses (β = 0.24, *SE* = 0.11, *Z* = 2.28, *p* = .023), with poor comprehenders showing weaker memory (*M* = .34, *SD* = .48) across sessions than good comprehenders (*M* = .42, *SD* = .49). All participants showed a steep decline in performance (β = -1.44, *SE* = 0.15, *Z* = -9.45, *p* < .001), from a mean proportion of .63 (*SD* = .48) correct after learning to .13 (*SD* = .34) at the delayed follow-up. However, there also emerged a three-way interaction between comprehension group, encoding time and test session (β = -0.27, *SE* = 0.08, *Z* = -3.32, *p* < .001): poor comprehenders were poorer at learning in the

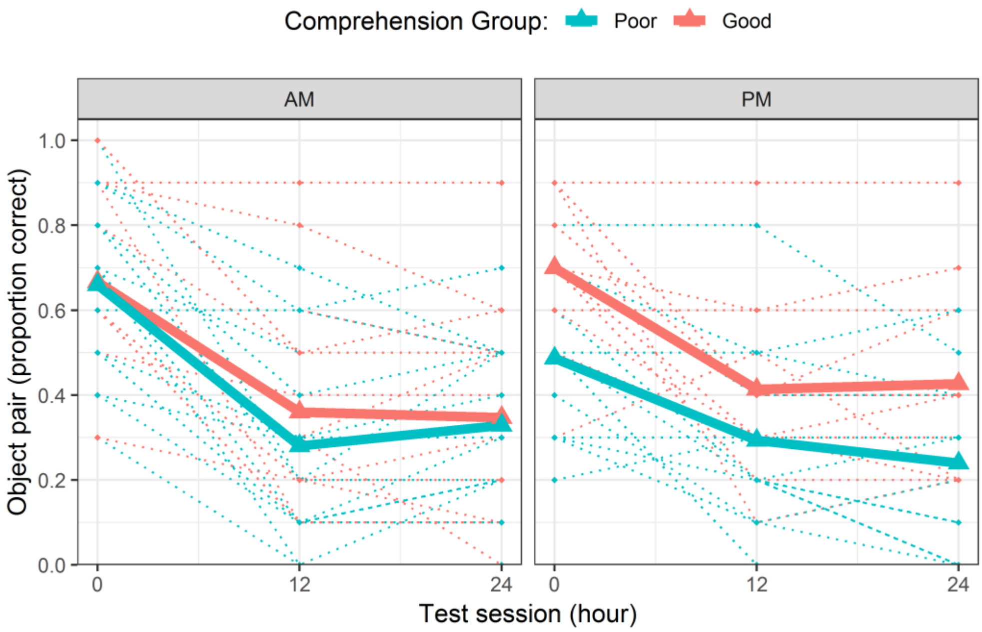


Figure 17. *Mean object-pair accuracy across the 0-, 12-, and 24-hour tests for the AM and PM encoding sessions separately. Dotted lines mark performance for each participant, with thick lines representing mean scores per comprehension group.*

evening (M = .49, SD = .50) but showed a weaker decline by the delayed follow-up (M = .14; SD = .35) than when they learned the items in the morning (0-hour: M = .66, SD = .48; delayed: M = .09, SD = .28). Good comprehenders did not show such large immediate differences between the AM-encoding (M = .67, SD = .47) and PM-encoding (M = .7, SD = .46), with both declining similarly by the delayed test (AM: M = .18, SD = .39; PM: M = .13, SD = .33).

7.4.5 Exploring individual differences in semantic knowledge

The group contrasts were one way of examining the hypothesis that weak semantic knowledge may constrain later consolidation of new word-forms, in line with previous studies that had indicated a retention deficit for poor comprehenders. Using a broad comprehension measure enabled us to identify children with weak understanding in the context of good phonological abilities. However, the poor comprehenders in the present study did not have as weak comprehension skills as previous samples, and there was substantial overlap in the range of vocabulary abilities within each group (standardised t-scores of good comprehenders: 48-70; poor comprehenders: 36-76). To further explore our original hypothesis, we carried out an additional analysis to assess whether the expressive vocabulary scores – as a measure of individual differences in lexical-semantic knowledge - predicted subsequent consolidation of new words. Using the vocabulary assessment enabled us to better capture the aspect of poor comprehenders' difficulties that we proposed to be most influential in their consolidation difficulties (the depth and richness of lexical-semantic knowledge), and provided a more meaningful raw score than the comprehension measure since different ability levels read different passages on the YARC. The stem completion data were used for this analysis to test our key original hypothesis that poor semantic knowledge would have broadening influences on new word-form knowledge over consolidation. Furthermore, recall of word-forms is particularly sensitive to sleep-associated improvements (in the present data and previous studies, e.g., Weighall et al. (2016)). This additional analysis therefore provides a novel opportunity to directly compare how vocabulary knowledge influences the retention and consolidation of new words when the *same* children learn at different times of the day.

For this analysis, we included all participant datasets, including the four children who did not meet the revised comprehension group selection criteria.

However, one child was missing a vocabulary score, resulting in 33 participants. We entered vocabulary score as a fixed effect alongside encoding time (AM vs. PM), test session (0-, 12-, 24-hour), and all interactions (Appendix D12). As would be expected, vocabulary ability was a highly significant predictor of overall performance ($\beta = 0.66$, $SE = 0.15$, $Z = 4.48$, $p < .001$). Most interestingly, there was a three-way interaction between vocabulary ability, encoding time, and 12-24 test session ($\beta = -0.34$, $SE = 0.13$, $Z = -2.57$, $p = .010$). As depicted in Figure 18 children with good vocabulary knowledge showed more improvements in recall over sleep (AM-encoded) than wake (PM-encoded) during this 12-24 hour period. Although in a similar direction for the relative sleep and wake comparisons, there was no evidence for an interaction with vocabulary ability across the 0-12-hour sessions ($p = .82$).

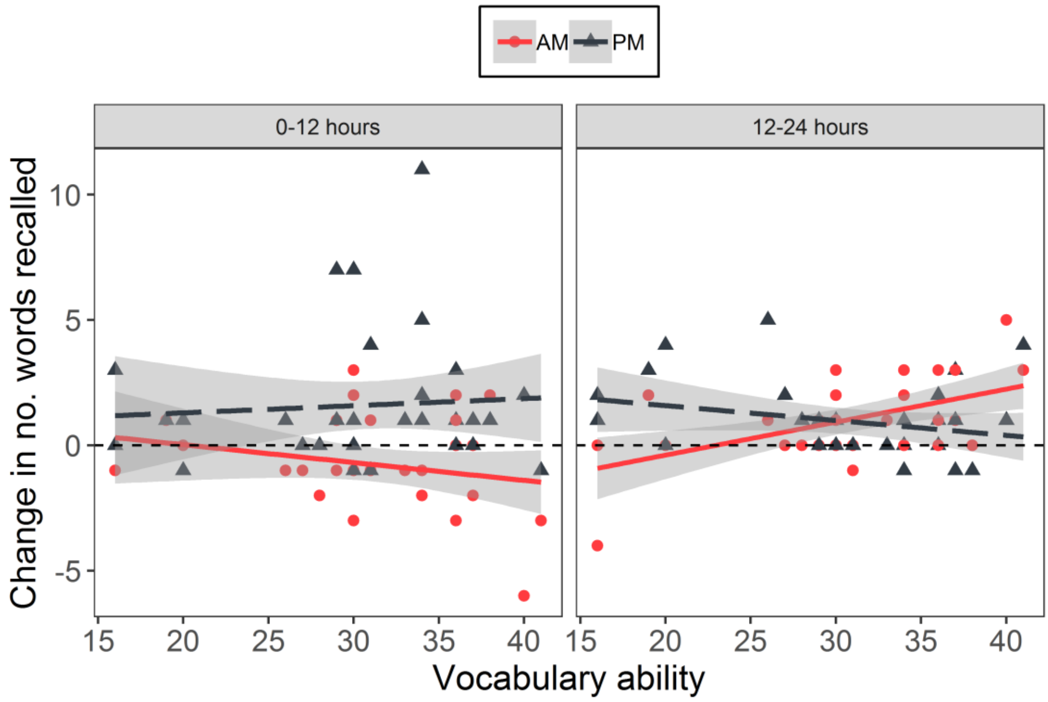


Figure 18. Change in stem completion accuracy across 12-hour periods for each of the AM-PM encoding conditions, plotted against participants' vocabulary ability scores.

7.5 Discussion

In this study, we sought to understand the learning and consolidation of new vocabulary in children with weak semantic knowledge, as is characteristic of children with poor reading comprehension. Poor comprehenders were relatively impaired at learning new vocabulary compared to good comprehenders and – in contrast to previous studies showing a primary deficit in semantic tasks – we showed that this weakness was general to all types of memory tasks. We tracked new word memory across wake and sleep, but saw no indication that poor comprehenders had weaker consolidation for their new vocabulary within the 24 hour period of the experiment, nor by the 1-to-2-month follow-up. Thus, although clearly demonstrating weaker vocabulary learning overall, consolidation mechanisms themselves were not a specific point of weakness for this group of children. On the contrary, there were clear sleep-associated benefits for performance across both comprehension groups, and these were long-lasting when sleep could occur soon after learning. When a day of wake intervened before opportunities to consolidate, an exploratory analysis (pooling across both good and poor comprehenders) suggested that expressive vocabulary ability may be a better predictor of vocabulary consolidation than broader comprehension profiles. As such, it may be that weak semantic knowledge affects what can be later prioritised during consolidation, rather than the consolidation process itself.

7.5.1 A benefit for sleep in learning and consolidating new vocabulary

In the context of the CLS model, sleep is proposed to strengthen memory for newly learned vocabulary (Davis & Gaskell, 2009). In both the stem completion and picture naming tasks, there was a clear benefit for sleep in the first 12 hours of learning which boosted recall in the second relative to the immediate test session. Both of these tasks required recall of the newly learned word-forms, cued by either the starting sound or the associated picture. Consistent with a number of previous studies (e.g., Chapter 3; Tamminen & Gaskell, 2013), there were no sleep-associated benefits for recall of semantic details in the definitions task (but see Henderson et al., 2013c, for contrasting findings). This pattern of results may suggest that sleep is most beneficial for strengthening word-form representations, but may also have been influenced by task order: encountering the pictures in the picture naming tasks may have refreshed memory when retrieving details for the definitions task, thereby overriding any consolidation benefit in the present study.

Memory also improved across the 12-24 hour period for the stem completion and picture naming tasks. This period featured sleep for the AM-encoded items, but comparable improvements were seen for the PM-encoded items, suggesting that wake is less detrimental to memory *after* (versus before) a period of sleep. In line with the CLS account, one possible explanation for this may be that sleep strengthened the neocortical representations of the new words, such that wake-based decay of hippocampal representations are less detrimental to retrieval accuracy after compared to before sleep (Hardt, Nader, & Nadel, 2013). In the more stable form, new representations may also be able to better benefit from retrieval practice to continue processes of consolidation (Antony et al., 2017). Interestingly, the changes in explicit memory (i.e., accuracy) were dissociated from changes in implicit access to them, with improvements in retrieval time seeming specific to sleep-based processes regardless of whether items are learned in the morning or the evening. These differences perhaps suggest that wake-based consolidation processes following sleep rely on different mechanisms than the sleep-associated improvements themselves.

Taken together, it seems that the first twelve hours after encoding are particularly important for this age group: sleep within this time window led to continued improvements across the 24-hour period. This finding corroborates those of Gais et al. (2006), who showed that sleep following learning was more beneficial to memory than sleep after an intervening period of day or night wakefulness (see also McGregor et al., 2013). Importantly, we extend earlier findings to show that these sleep-associated improvements support the longer-term retention of new vocabulary, with benefits for PM-encoded information still apparent 4-10 weeks later. In contrast, a day's wakefulness before opportunities to consolidate risks longer-term forgetting of new information.

7.5.2 Vocabulary learning in poor comprehenders

The poor comprehenders in our sample showed generally weaker vocabulary learning than the good comprehenders, consistent with their poorer existing expressive vocabulary knowledge and with previous studies demonstrating weaker vocabulary acquisition for this group (e.g., Cain et al., 2004). However, we had predicted that poor comprehenders would show equivalent learning of word *forms* to good comprehenders at least at the immediate test point, given that phonological skills are a relative strength for these children (e.g., Nation & Snowling, 1998). Two previous studies had showed only delayed impairment for phonological aspects of new word knowledge in children with comprehension and vocabulary weaknesses (Nation et al., 2007; Ricketts et al., 2008), consistent with evidence of delayed influences of semantic knowledge for word-form recall (Henderson et al., 2013c), but the present findings do not support this pattern. The broader weaknesses seen likely result from the more challenging nature of the present experiment: we taught children significantly more words than in the study of Nation et al. (2007) and used tasks that assessed explicit recall of the new words at each test point. Indeed, an exploratory analysis of the multiple-choice recognition data from training showed only slight group differences that did not reach statistical significance (p = .065; Appendix D13). This suggests that previous studies have perhaps failed to capture the encoding deficit that has been observed here. However, there are also sample differences between the present experiment and previous studies: as a result of recruitment difficulty, our control group were more above average comprehension ability than the poor comprehender group were below. It is therefore plausible that that the broader group differences may be attributable to the good comprehenders' superior learning in this instance, rather than poor comprehenders' weaknesses *per se*.

7.5.3 Vocabulary consolidation in children with poor semantic knowledge

We set out to test the hypothesis that poor comprehenders would show weaker consolidation of new vocabulary, in the context of their poorer semantic learning. The data did not support this hypothesis: there was no evidence that sleep-associated consolidation of new words was particularly problematic for poor comprehenders. On the contrary, memory for the new words was remarkably stable for both groups even when tested 1-to-2 months later. There was a slight indication of weaker overnight consolidation of word-forms in the stem completion task when poor comprehenders learned in the morning, but this difference was not statistically significant. However, our exploratory analysis of individual differences was more strongly indicative of this pattern: from 12-24 hours, vocabulary was a more positive predictor of recall improvements for the AM condition (i.e., overnight) than the PM condition. It may be then that vocabulary differences better capture differences in consolidation than the comprehension profiles alone, which likely have heterogeneous aetiologies.

Interestingly, there was no evidence of an interaction with vocabulary at 0-12 hours. That is, there are clear benefits for sleep within the first 12 hours regardless (as described above), whereas consolidation is more reliant on prior knowledge when there is a wake delay before opportunities to consolidate. One possible explanation for this finding is that the new lexical representations of children with poorer vocabulary deteriorate more during the day and are less able to recover overnight. However, the lack of interaction between vocabulary ability and change in performance over the first 12 hours – at least not in a way that could be detected by explicit retrieval measures – presents a challenge for this account. Alternatively, we propose that those with weak vocabulary are less likely to prioritise new words in later consolidation when there is an intervening period of wakefulness, whilst children with good vocabulary are better able to recover and consolidate these more fragile memories. This might be because superior vocabulary knowledge affords more robust connections to prior knowledge that are less prone to decay and/or interference. Alternatively, those with good vocabulary knowledge may be better able to capitalise upon repeat testing benefits for subsequent consolidation (Roediger & Karpicke, 2006). However, it is important to remember that this finding was a result of exploratory analyses: future studies should seek to replicate and test these alternative hypotheses. If supported, this could have important implications for timing vocabulary instruction to best support longer-term retention of new words, particularly for those with weaker language skills.

7.5.4 Broader declarative memory consolidation

Turning to the nonverbal declarative memory task, memory for the locations declined steeply from 0-12 hours, but there was evidence of less forgetting when sleep featured during this period. This sleep-associated benefit was smaller than in previous studies (Henderson et al., 2012; Wilhelm et al., 2008) and – counter to our expectations – we did not see a later-emerging benefit for sleep during the 12-24 hour period when items were learned in the morning. This is in contrast to the vocabulary tasks, in which memory improved across both encoding conditions across the 12-24 hour period. However, one key difference between our paradigm and that of previous studies (e.g., Henderson et al., 2012) is the removal of feedback during the test sessions. This decision was made to enable comparable tracking of memory across the three time points without further opportunities for learning, yet this removal of feedback may account for the weaker benefits for sleep seen in the present experiment. Despite these differences, performance on this nonverbal task was consistent with the conclusions above in that the first 12 hours was most important for the benefits of sleep to emerge.

Counter to our predictions that poor comprehenders would have specific language learning difficulties (based on findings from other aspects of memory, e.g., Nation, Adams, Bowyer-Crane, & Snowling, 1999), the poor comprehender group also performed more poorly in the nonverbal task than the good comprehender group.

There was some evidence of a difference in nonverbal ability between the two groups which – although not statistically significant – may have related to these broader differences in learning ability. Alternatively, it could be argued that the task itself was not independent of language skills, and that verbal strategies may have offered support in encoding items and their locations.

There were also group differences in longer-term consolidation for this task: poor comprehenders showed less decline by the delayed follow-up when they learned the pair locations in the evening compared to the morning, whereas this was not the case for good comprehenders. However, poor comprehenders also started at a lower level of performance in the PM encoding condition, presenting a challenge to interpreting what drives these differences in memory decline: it may be that poor comprehenders simply had less knowledge to lose after learning in the evening, or that sleep was particularly beneficial after weaker encoding. This weaker evening encoding for poor comprehenders was also seen in the picture naming response times and was numerically reflected in other measures, perhaps suggesting that immediate sleep could yet prove to be more broadly beneficial for this group of children.

7.5.5 Limitations and conclusions

It should be noted that the small sample size remains an issue for the present study, as it has done with many previous comparisons between good and poor comprehenders. Our research questions were motivated by previous studies of poor comprehenders, presenting an ideal focus group for assessing contributions of semantic knowledge to vocabulary consolidation (James et al., 2017). However, despite using a within-subjects design for all experimental manipulations and increasing statistical power via the number of items learned, the challenge of recruiting an atypical group – and for such an intensive study – still undermines our ability to draw strong conclusions from the data. Given the heterogeneity of poor comprehenders' difficulties (e.g., Colenbrander et al., 2016; Nation et al., 2002), it is perhaps not surprising that an additional exploratory analysis with vocabulary as a continuous predictor offered the most insight into our predicted relationship. Vocabulary differences were apparent in previous studies of interest, and we propose that this continuous approach may be most fruitful in furthering our understanding of vocabulary learning in children with semantic weaknesses.

In summary, the present study showed that children with weaker reading comprehension learn vocabulary at a slower rate than those with good comprehension skills, and that this relative impairment is apparent even when new vocabulary is taught directly (i.e., not reliant on text comprehension). Although previous studies had shown weaker vocabulary retention for this group, it does not appear to be the case that poor comprehenders have problems specific to offline consolidation mechanisms. As a whole, there was clear evidence that sleep soon after learning can have long-lasting benefits for memory, and that this is the case regardless of language ability. When learning was followed by day of wake however, new words were less likely to be retained for the longer term, and this was particularly the case for children with poorer existing vocabulary knowledge. This suggests that previous vocabulary-related differences in retention might not relate to the offline consolidation mechanisms *per se*, but to the likelihood that information is prioritised for later consolidation. Given that literacy instruction typically features in the morning in the UK education system, this finding – if supported by future studies - would have important implications for how vocabulary instruction can be better timed to support struggling learners.

Chapter 8. General Discussion

The research presented in this thesis sought to further our understanding of individual differences in learning and consolidating new vocabulary. It is known from previous literature that sleep-associated processes play an important role in consolidating new words encountered during the day, strengthening memory for new words and enabling their integration with existing vocabulary knowledge. However, the benefits of offline consolidation are variable across experimental designs and across individuals, highlighting the importance of understanding factors that influence longer-term vocabulary retention beyond those that facilitate initial learning. In Chapter 1, existing evidence was reviewed to suggest that existing vocabulary knowledge may support the longer-term consolidation of new words, akin to the ways in which cognitive schema support the acquisition of broader knowledge. The nine experiments presented adopted varied approaches to test this hypothesis, aiming to dissociate between general associative relationships – i.e., children who are good at consolidating new words have a larger vocabulary – and an account that proposes active support from existing knowledge in offline consolidation of new words (James et al., 2017). These studies have informed the ways in which prior knowledge does – and as importantly, does not – support new learning, how these processes unfold over time, and how learning mechanisms change in their relative importance across development. I will begin this discussion chapter by summarising each of the key findings (see also Table 15), before addressing their broader theoretical contributions and implications for vocabulary development.

8.1 Summary of experimental findings

8.1.1 Chapter 3

The first experiments presented in this thesis manipulated novel words' connections to semantic knowledge to test the hypothesis that words trained into more richly populated areas of semantic memory would benefit in overnight consolidation. Adults and children were taught novel concepts associated with low- and high- density semantic neighbourhoods, and completed memory tasks before and after opportunities for consolidation (same day, next day, week later). Recall of word-forms improved with opportunities for consolidation, and these effects were bigger in children than in adults. In contrast to the original hypotheses, novel items associated with high-density

Table 15. Summary of offline consolidation and prior knowledge effects across all experiments.

	Group	Mean form recall improvement (Day1-Day2)	Mean form recall improvement (Day2-Day8)	Local prior knowledge overall	Local prior knowledge in consolidation	Global prior knowledge overall	Global prior knowledge in consolidation
Chapter 3 – Semantic neighbours							
Experiment 1	Adults	12.8%	3.2%	-	-	NA	NA
Experiment 2	Children	31.7%	29.1%	↓	→	NA	NA
Experiment 3	Adults	15.5%	1.9%	↓ / ↑	→	NA	NA
Chapter 4 – Form neighbours (taught)							
Experiment 1	Children	5.6%	13.4%	↑	↓	↑	→
Experiment 2	Adults	12.1%	2.2%	(↑)	(↓)	↑	→
Experiment 3	Children	7.5%	14.6%	↑	↓	↑	→
Chapter 5 – Form neighbours (stories)							
Experiment 1	Children	4.2%	14.31%	↓	→	↑	→ / ↑
Experiment 2	Adults	0.5%	3.4%	↑	→	↑	→
Chapter 7 – Poor comprehenders							
AM-PM Experiment	Children	12.3%	NA	NA	NA	↑	→ / ↑

Notes. Recall improvements denote the improvements observed in the stem completion task, as a percentage of total items trained. For prior knowledge, arrows mark the direction of any significant effect (for any task): ↑ indicates a benefit for prior knowledge, ↓ indicates interference from prior knowledge, and → means that effects do not change with consolidation. Where more than one arrow is shown, this marks a contrast in the direction of influence across different tasks. Parentheses indicate relationships that did not reach statistical significance in the individual experiment analyses, but showed a consistent pattern in cross-experiment analyses.

neighbourhoods were recalled more poorly than items associated with low-density neighbourhoods, and this did not change with opportunities for consolidation. It was concluded that a locally dense network of semantic connections may hinder (rather than support) the learning of new associated concepts, and that these influences can occur early and persist across consolidation.

8.1.2 Chapter 4

The experiments in Chapter 4 complemented those in Chapter 3 by instead manipulating the number of word-form neighbours, to test the hypothesis that prior knowledge is accessed via form-based connections during learning and consolidation. Children and adults were taught pseudowords that had either no, one, or many word-form neighbours, and were again tested on their memory on the same day, the next day, and one week later. In contrast to Chapter 3, an abundance of local *form*-based connections exerted a positive influence on new learning. Children were better at recalling pseudowords with neighbours than pseudowords without neighbours at the immediate test point, but this benefit disappeared over a period of consolidation: pseudowords without neighbours received the greatest benefit from offline consolidation, and reached comparable levels of recall at later test points. As above, this consolidation effect was not as strong for adults, supporting the proposal that individuals receive greater benefit from offline consolidation processes earlier in development.

The experiments in Chapter 4 showed a robust early benefit for local prior knowledge but – somewhat surprisingly – this benefit showed limited interactions with global vocabulary knowledge. Vocabulary ability was a highly significant predictor of new word memory, but it was not influenced by the presence of many word-form neighbours or by a period of consolidation. There was some evidence that pseudowords with only one neighbour were more sensitive to individual differences (Chapter 4 Experiment 2), but there was still no evidence that vocabulary knowledge predicted consolidation of the new word-forms in any condition.

8.1.3 Chapter 5

In Chapter 5, we examined whether word-form neighbours similarly influenced new word knowledge during incidental word learning from stories, in contrast to the explicit teaching paradigm employed in Chapter 4. It was predicted that presenting the stimuli in spoken stories would minimise strategic access to prior knowledge during learning, resulting in weaker connections to prior knowledge that could be strengthened via offline consolidation. For children, there was some evidence that this presentation format might alter the influence of prior knowledge in learning pseudowords: they showed no influence of local word neighbours in recalling the pseudowords, and performed worse at recognising word-forms with many neighbours. For adults, effects of prior knowledge were similar to those seen in Chapter 4: adults were better at recalling word-forms with many neighbours than no neighbours. In neither experiment did these influences of word-form neighbours change with consolidation.

Vocabulary ability was once again a highly significant predictor of vocabulary learning in this task, but there remained no interaction between global vocabulary knowledge and participants' benefit (or hindrance) from word-form neighbours. However, in using the story paradigm, a relationship emerged between children's expressive vocabulary ability and overnight improvements in the picture-form matching task. This suggests that story contexts may enable the formation of richer semantic connections between new words and existing vocabulary knowledge. Connections to semantic knowledge in this sense are proposed to be more distant and varied than the close competing neighbours trained in Chapter 3.

8.1.4 Chapter 7

The final experiment took an alternative approach to understanding individual differences in vocabulary learning by closely examining sleep-associated consolidation in children proposed to vary in semantic knowledge. If broad connections to semantic knowledge are important for consolidating newly learned vocabulary, then those who typically have poor semantic knowledge are hypothesised to show poorer vocabulary consolidation. This hypothesis was tested by teaching good and poor comprehenders new vocabulary at the beginning or end of the day, and assessing memory 0-, 12- and 24-hours later – thereby isolating periods of wake and sleep-associated consolidation. Children improved in their ability to recall the new words after sleep compared to wake, and these benefits were long-lasting if sleep could occur within the first 12 hours of learning. However, counter to our hypotheses, the poor comprehenders were generally weaker across all measures of vocabulary learning than good comprehenders, and these relative impairments remained consistent over time.

An exploratory analysis did provide evidence of a relationship between global vocabulary knowledge and overnight improvements in recall of the new word-forms. Most interestingly, this association was only seen for vocabulary learned earlier in the day, whereas all children (regardless of vocabulary ability) were able to strengthen their new word knowledge if sleep followed soon after learning. This suggests that individual differences in sleep-associated consolidation may not relate to the overnight consolidation mechanisms themselves, but rather the likelihood that weakened traces can be recovered and strengthened during offline consolidation.

In sum, the present experiments were designed to capture the ways in which prior linguistic knowledge might drive individual differences in learning and consolidating new vocabulary. In manipulating local connections to prior knowledge, it became clear that such direct similarities to existing knowledge can influence new learning prior to opportunities for offline consolidation, but also that these similarities can hinder or help in different circumstances. In measuring global vocabulary knowledge, it was apparent that vocabulary ability is always a highly significant predictor of new vocabulary learning, but that its supporting role during consolidation may depend on the learning context more than was previously predicted. In this final discussion, I will first consider the ways in which these findings contribute to our understanding of offline consolidation in vocabulary learning, in the context of the CLS model. I will then return to the theoretical predictions presented in Chapter 1 to address what the role of prior knowledge may be within these models of learning, and consider the evidence that these factors contribute to individual differences in vocabulary development.

8.2 Offline consolidation of new vocabulary

The Complementary Learning Systems account of word learning proposes that this process engages two neural systems: the hippocampal system required for the rapid acquisition of a new word, and the slower-learning neocortical system that enables integration with existing vocabulary knowledge (Davis & Gaskell, 2009). Some of the clearest evidence for these dual systems in language learning arises from findings that explicit knowledge of trained pseudowords is rapidly acquired, but there are delays in the emergence of lexical competition between new and existing words (e.g., Dumay & Gaskell, 2007; Dumay et al., 2004). However, a number of studies have now provided evidence that these slower learning processes can also support improvements in explicit memory for the new words (Dumay et al., 2004; Henderson et al., 2012), which was the focus of the present experiments.

8.2.1 Evidence for improvements in word knowledge “offline”

In eight experiments (Chapters 3, 4, 5), we taught participants pseudowords and assessed memory for them on the same day, the next day, and one week later. As shown in Table 15, recall of the new word-forms consistently improved beyond the end of training on Day 1 (as measured by a stem completion task). These improvements were largely attributed to benefits of offline consolidation, but it is important to note that these studies also incorporated additional presentations of the pseudowords in the additional tasks administered (e.g., in prompting for meaning recall, in semantic categorisation, and in recognition tasks). Although additional presentations were minimal, these may have acted as feedback for the pseudowords not quite remembered in the recall task (e.g., Krishnan, Sellars, Wood, Bishop, & Watkins, 2018), or facilitated memory in similar ways to spaced learning (Sobel, Cepeda, & Kapler, 2011). It seems likely that additional presentations did contribute to the improvements seen across the week, but an interpretation based primarily on offline consolidation mechanisms is favoured in light of the ninth experiment (Chapter 7): improvements in memory performance between repeated tests were seen only across or after a period of sleep had occurred, whereas no such improvements were seen across an initial period of wake. These findings corroborate those of previous studies showing greater improvements in word knowledge over sleep compared to wake (e.g., Dumay & Gaskell, 2007; Gais et al., 2006; Henderson et al., 2012). Importantly, we showed that opportunities to consolidate soon after learning can have long-lasting benefits on memory for school-aged children, which were still apparent 4-10 weeks later.

Interestingly, it is clear from the present experiments that opportunities for offline consolidation facilitate *improvements* in word-form recall. These gains in memory support an “active consolidation” mechanism (Diekelmann & Born, 2010; Ellenbogen, Payne, & Stickgold, 2006), as opposed to a passive role for sleep in preventing memory decay. Such conclusions are consistent with studies that have measured neural activity during sleep (e.g., Smith et al., 2017; Tamminen et al., 2010), converging on evidence that low-frequency oscillatory activity occurring in deep sleep stages may facilitate replay of hippocampal traces to neocortical systems (Staresina et al., 2015). This is not to say that sleep does not also benefit the maintenance of items, an issue that has been at the centre of much recent debate (Dumay, 2016; Fenn & Hambrick, 2013; Schreiner & Rasch, 2018). However, one key difference between the opposing conclusions of the maintenance versus gains debate is the type of memory task used (recall of word-forms versus paired associates). This brings us onto the next topic for discussion: what aspects of word knowledge benefit from sleep?

8.2.2 What benefits from offline consolidation?

Davis and Gaskell (2009) previously noted the greater evidence for sleep benefits in memory for word-forms, compared to the semantic aspects of word knowledge. Despite always training the pseudowords alongside a meaning (e.g., definition, novel object), the benefits of offline consolidation seen in the present studies were largely specific to word-form knowledge: there were consistent improvements in recall of the new word-forms, and also nearly always in their recognition if tested independently from the semantic mapping. In contrast, there was no evidence of an offline benefit for recalling or providing associated definitions (Chapters 3, 7), or in recognition tasks that incorporated the semantic element (with one exception discussed below, Section 8.3.2). Thus, it does appear that it is word-form knowledge primarily that benefits from a period of offline consolidation, at least in tasks that assess explicit knowledge.

Weaker benefits for offline consolidation in semantic components of word learning may relate to their associative nature. I consider two – not mutually exclusive – ways to conceptualise these relationships. First, these findings may result from the way in which semantic knowledge was assessed within these experiments: whilst assessment of the new word-forms typically relied on accessing the word-form *items* only, the measures of semantic knowledge tended to assess the *association* between form and semantic knowledge (i.e., required knowledge of the form-semantic mappings to access newly learned semantic information). To better understand the influences of offline consolidation on these different elements of new knowledge in more comparable ways, it will be important to assess learning of the new semantic information independently of new form knowledge – e.g., by cueing definition knowledge using part-definitions, and by testing familiarity with the new objects presented.

Second, the semantic components of word learning are also more associative in their relationship with existing knowledge. For example, in Chapter 3, we trained novel concepts that made explicit reference to an existing concept (e.g., a *chicken* that sleeps upside down). In Chapters 5 and 7, the novel items were related to classes of items already known (e.g., type of drink, type of cat). These influences of prior semantic knowledge were arguably minimised for the experiments in Chapter 4, that made use of novel objects (Horst & Hout, 2016), yet a number of these still represented familiar toys. Schema-related elements of new knowledge may receive direct benefits from existing knowledge and not require further consolidation offline, as will be discussed below (Section 8.3.1; see also van Kesteren et al. (2013)). Whilst we made some attempt to manipulate the relations between novel concepts and existing knowledge in Chapter 3, it is clear that further studies are warranted to better understand how multiple divergent and overlapping associations may support new learning.

8.3 Situating prior knowledge in neural models of learning and consolidation

The evidence presented in this thesis thus supports that sleep is beneficial for consolidating at least some aspects of new vocabulary knowledge. However, the factors that drive variability in these processes are not well understood. In Chapter 1, it was proposed that one source of variability in vocabulary consolidation lies in the learner's prior linguistic knowledge. Since Davis and Gaskell described the utility of the CLS framework for understanding word learning in 2009, Complementary Learning Systems Theory has been extended to acknowledge the benefits of cognitive schema during learning (Kumaran et al., 2016; McClelland, 2013). In these amended models, the neocortical system is described as "prior knowledge dependent", to capture the ways in which new information is more rapidly acquired if it is consistent with known information (e.g., learning a pigeon versus a penguin as an example of a bird). In Chapter 1, it was similarly argued that vocabulary knowledge may act as a linguistic schema, based on evidence that existing vocabulary knowledge predicts overnight consolidation of new vocabulary (Henderson et al., 2015; James et al., 2017).

However, whilst a range of experimental and correlational evidence is suggestive of a role of prior knowledge in vocabulary learning, the precise nature of these prior knowledge relationships have not been well-specified: what aspects of a lexical representation might benefit from prior knowledge and why? Under what circumstances are individuals most likely to benefit from prior knowledge? What should the timescale of this facilitation should be, and how will prior knowledge impact the benefits of sleep for new learning? One possibility for this latter question is that the hippocampus binds relevant connections to neocortical knowledge as part of the newly formed memory trace. Thus, as hippocampal replay occurs offline, these prior knowledge connections cause co-activation of the new and existing representations and the two can become integrated more efficiently than when memories share fewer connections to existing knowledge (e.g., Lewis & Durrant, 2011). By this account, any prior knowledge benefits observed should *strengthen* over sleep. An alternative interpretation posits that prior knowledge may speed neocortical learning from the outset: without risk of catastrophic interference, there may be minimal demands on offline consolidation mechanisms to integrate the new information with existing knowledge. By this "cortical learning" account therefore, information that does *not* benefit from prior knowledge is most reliant on hippocampal mechanisms during learning, and thus benefits most from replay during sleep.

The work presented in this thesis addressed these questions using two approaches: manipulating "local" prior knowledge to assess learning and consolidation for items varying in their potential connections to existing knowledge, and measuring individual differences in "global" prior knowledge to better understand how individuals might benefit from their existing knowledge. These two approaches were intended to be complementary: we predicted that associations with global prior knowledge during learning and consolidation would be stronger for items that we manipulated to have more local connections to prior knowledge, under the assumption that individuals with better global knowledge would know more of the relevant local connections. However, there was very little evidence that this was the case, and – whilst both approaches inform our understanding of prior knowledge in vocabulary learning – they appeared to capture different aspects of this relationship. As such, this discussion will primarily address how each approach has informed models of learning independently, drawing comparisons where appropriate.

8.3.1 How has manipulating "local" prior knowledge informed the CLS model?

One way to better understand the role of prior knowledge in consolidation is to manipulate prior knowledge on an item-level, comparing the memory trajectories for new information that has more versus less potential to relate to prior knowledge. The data presented in this thesis are clear in demonstrating that these "local" connections to prior knowledge can influence memory early in the learning process: regardless of the type of manipulation (semantic, word-form) and learning context (explicit teaching, incidental learning through stories), any effects of prior knowledge were apparent when tested on the same day as learning. Interestingly, the experiments in Chapter 4 showed this influence to diminish with consolidation, as items with weaker connections to prior knowledge were preferentially strengthened during offline consolidation. These data favour the cortical learning account described above, whereby neocortical learning can proceed immediately in the context of prior knowledge, "fast-tracking" consolidation processes such that offline replay is not required.

This neocortical learning account is also supported by a handful of recent studies. For example, in a study by van Kesteren et al. (2013), associative memories benefited from schematic knowledge immediately in learning and persisted over consolidation, whereas memory for the new items did not show these benefits until after sleep. Mirković and Gaskell (2016) also showed that participants who napped for 90 minutes after learning new vocabulary improved in their memory for the arbitrary components more so than participants who stayed awake during an equivalent period, whereas there were no such sleep-associated differences for the more systematic elements. It was argued that the overlapping nature of systematic elements during learning enabled the formation of a distributed neocortical representation without further need for offline replay - much in the way that the overlap between new representations and existing ones are proposed to support neocortical learning in the present experiments.

However, evidence for local prior knowledge benefits is not evidence in itself for a consolidated neocortical representation: novel vocabulary is still processed in the context of neocortical knowledge prior to a new episodic trace being formed in the hippocampus. Indeed, an alternative interpretation for the weakening influence of word-form neighbours in Chapter 4 is that the items initially benefiting from prior knowledge showed *reduced* offline consolidation effects because of their difficulty to integrate with overlapping items in the neocortical system. It is not possible to determine the extent to which any items became consolidated into existing linguistic knowledge in the present studies: the lack of semantic density effect for the implicit task in Chapter 3 suggested new items had not been integrated, but no other experiments measured integration in this thesis. It seems likely that the underlying lexical representations of knowledge-related words likely change in ways not captured by our measures. For example, Havas et al. (2017) trained participants on new words that were more versus less similar to their native language and showed that – whilst participants benefited from sleep in recognition only for word-forms less similar to their existing knowledge – sleep enabled the engagement of language-similar word-forms with existing knowledge as marked by a semantic priming task.

A final issue for our studies of local prior knowledge is to consider the divergent effects seen between Chapters 3 and 4: why do semantic neighbours interfere with new learning, yet form neighbours benefit new learning? We could speculate that these differences may relate to the consequences that semantic- and form-based errors have for communication. If an individual does not possess the correct vocabulary for a concept, using a word that refers to a close semantic neighbour will still communicate relevant meaning (e.g., referring to a *chimpanzee* as a *monkey*). This makes it relatively less important to learn new words with close semantic relationships, whereas learning concepts distantly related to our existing knowledge enables communication in a broader range of contexts. In contrast, the mappings between word-forms and their concepts can appear relatively arbitrary – and were in the present experiments – such that a word-form error may communicate the wrong meaning entirely (e.g., referring to *money* as a *monkey*). On this basis, it is important to prioritise learning distinctions between novel and existing word-forms to ensure correct terminology is used, whereas errors for word-forms without neighbours have less catastrophic consequences for communication. However, the contrasting influences of semantic and word-form variables on vocabulary learning are not consistent across studies (Storkel, 2009), and may relate more to the specific stimuli used in the present experiments.

The evidence presented in this thesis therefore supports that local prior knowledge can influence new word knowledge from the outset, and that this explicit knowledge of a new word is not additionally influenced by prior knowledge over consolidation. Whilst these findings favour a neocortical learning account for prior knowledge within in the CLS framework, the precise mechanisms underlying the engagement of prior knowledge in these paradigms will be better informed by experiments assessing the integration of new words within neocortical vocabulary. Our understanding of prior knowledge influences in language learning will also benefit from manipulating more distant and wide-ranging relationships, as the very close form manipulations used in this thesis (i.e., single sound/letter change) were not sensitive to individual differences in prior knowledge. One particularly useful approach may be to use manipulations that better reflect the morphological structure of the language system. For example, how does having a network of words from the Latin root *nov* (e.g., *novel, innovate, novice*) support the learning of a related word (e.g., *novitiate*)? These similarities better address the ways in which prior linguistic knowledge could support new learning, combining both form and meaning in meaningful ways. Manipulating local prior knowledge in this way may also relate more closely to measures of global vocabulary knowledge, given previous evidence that children with superior vocabulary ability appear to be more sensitive to derivational relationships (Freyd & Baron, 1982).

8.3.2 How have studies of "global" prior knowledge informed these models?

The second approach to assessing contributions of prior knowledge to word learning was to measure the prior knowledge of the learners themselves. Across all experiments that used standardised assessments of vocabulary knowledge (Chapters 4, 5, 7), vocabulary ability was a strong and highly significant predictor of word learning performance on every experimental task (see Table 15). Although in many respects not surprising – one would expect that those who are good at learning new words to also have good vocabulary knowledge already – we had not anticipated such strong relationships in adults, especially in a sample drawn from an undergraduate population. The pervasiveness of this relationship highlights that vocabulary knowledge remains variable even within student populations, and that this remains a strikingly good marker of new vocabulary learning across ages.

Previous studies had shown that such individual differences in existing vocabulary knowledge were further predictive in overnight vocabulary consolidation beyond initial differences in learning (James et al., 2017). Within the context of the CLS model, it was proposed that the greater availability of prior knowledge would be able to bolster the neocortical consolidation of newly acquired words. However, there was very limited evidence of such a relationship in the present experiments: the relationship with global vocabulary knowledge always held across test sessions, and rarely changed with consolidation. As in Section 8.3.1, this supports an early role for prior linguistic knowledge in acquiring new vocabulary, although there was no evidence that individuals with more expressive vocabulary knowledge were any better at learning the pseudowords that we manipulated to be more closely related to existing knowledge. Clearly then, it is necessary to reconsider the relationship found in previous studies, and the alternative ways in which language schemas may yet prove useful in consolidation.

The analysis in James et al. (2017) was conducted on a measure of integration: children with better expressive vocabulary knowledge showed bigger increases in lexical competition over a night's sleep. However, studies of lexical competition are highly restrictive in their stimulus design (e.g., training *cathedruke* to overlap strongly with existing word *cathedral*; Gaskell & Dumay, 2003), and so we chose to focus on explicit recall measures in the present experiments to enable more flexibility in manipulating potential connections to existing knowledge. In support of this decision, previous studies provided evidence that similar relationships between existing vocabulary knowledge and overnight improvements held for recall of new forms (Henderson et al., 2015). The lack of evidence for such a relationship in the present experiments leads to us to consider previous findings in two alternative ways. First, perhaps existing vocabulary knowledge is a better predictor of lexical integration, meaning that the tasks used in these thesis failed to capture the ways in which existing knowledge might support consolidation. Second, it may be that the stimuli used to assess integration are more sensitive to vocabulary-related differences. As described above, studies of lexical competition train pseudowords that overlap closely with a single real word (i.e., *cathedruke* - cathedral; Gaskell & Dumay, 2003), which may allow for variability in individuals accessing the neighbour, and/or may provide a direct route to semantic knowledge via the neighbouring word. These stimulus-related differences were supported in Chapter 4 (Experiment 2), in which pseudowords overlapping with only a single existing word were shown to be most sensitive to vocabulary-related differences in performance. Thus, although sleep clearly plays an important role in strengthening new vocabulary (Chapter 7), the magnitude of relationships with prior knowledge may have been overstated by previous experiments.

What evidence is there then for a relationship between existing vocabulary knowledge and overnight consolidation in this thesis? In Chapter 5, children with better existing vocabulary knowledge improved more overnight in their ability to select the correct referent for pseudowords encountered in stories than children with weaker vocabulary knowledge. There are a number of key differences between this experiment and those of previous chapters that may inform *when* prior knowledge plays a supporting role in consolidation. First, presenting items in stories provides opportunities for developing rich and varied connections with semantic knowledge, as was also the case for Henderson et al. (2015). In line with this proposal, Henderson & James (2018) showed that children with good vocabulary showed superior consolidation of pseudowords learned across multiple story contexts compared to repeated contexts. The earlier experiments presented in this thesis were lacking in these opportunities to build such broad semantic connections from context, which may be crucial in benefiting from prior knowledge during consolidation.

Second, the relationship of the novel items with existing knowledge may also facilitate these connections, as the items in the stories were related to similar real objects (e.g., a coat made on Saturn). In contrast, the experiments in Chapter 4 used novel objects to limit contributions of semantic knowledge (Horst & Hout, 2016). The same novel objects have been shown not to influence vocabulary learning in other studies (Hawkins & Rastle, 2016), whilst familiar objects support vocabulary learning (Havas et al., 2017). Similar item relationships could also underlie the AM-PM experiment findings presented in Chapter 7, the other experiment to find an association between existing vocabulary knowledge and overnight consolidation of word-forms. In this experiment, children learned new animals and plants via explicit training, and were encouraged to draw comparisons with known exemplars in the definitions task.

The proposal that item-level associations with semantic knowledge may facilitate consolidation is somewhat at odds with earlier experiments in this thesis that highlighted the ways in which semantic similarities can present a challenge to learning: in Chapter 3, definitions incorporating high-density semantic concepts inhibited new learning compared to those incorporating low-density concepts. Perhaps then, it is the opportunity to engage this item knowledge flexibly and internally generate connections that is most important. This proposal can also help to account for vocabulary-related differences in studies training single neighbouring words (e.g., *cathedruke, dolpheg),* which were proposed above to enable access to semantic knowledge. Indeed, participants in Davis et al. (2009) reported increased meaning attribution to the novel words the day after learning, despite not being taught any semantic information alongside the word-forms. Future experiments should directly assess contributions of semantic knowledge to learning knowledge-relevant concepts with and without story contexts to assess how and whether both elements contribute to new learning.

The findings from the AM-PM experiment (Chapter 7) also posed an intriguing new question: does prior knowledge actually support the offline consolidation process itself, or processes that happen prior to sleep? For new words trained in the morning, existing vocabulary ability predicted later overnight consolidation of new word-forms. However, vocabulary ability did not show a relationship with overnight consolidation if that period of offline consolidation could happen soon after learning (i.e., when the words were learned in the evening). There are likely multiple contributing factors to what information gets prioritised in memory consolidation, such as recency, reward, motivation, or saliency (see Diekelmann, Wilhelm, & Born, 2009, for a review). The present findings indicate that connections to prior knowledge may support this later consolidation for new vocabulary, whereas learning new words before bed enables all children to benefit from recency. This finding clearly warrants replication, and should be thoroughly examined to further our understanding of prior knowledge processes in systems consolidation of new vocabulary.

The studies presented in this thesis therefore demonstrate that individual differences in global vocabulary knowledge are strongly related to word-learning ability, but that they are not always further predictive of individual differences in overnight consolidation. Instead, it appears that opportunities to capitalise upon existing knowledge in rich and varied ways are important for supporting this process. Furthermore, the nature of underlying support may be in prioritising language for later consolidation rather than offline consolidation processes themselves.

8.4 Vocabulary learning from childhood to adulthood

In Section 1.6.1, we stressed the need for more developmental comparisons to better understand the changing role of sleep-associated consolidation across development. Specifically, it was predicted that adults would always gain from superior language knowledge when new words shared neighbours, but that overnight consolidation benefits would be stronger for children where new words and/or concepts shared few similarities with existing knowledge. This hypothesis was supported in Chapter 4: children showed bigger benefits of offline consolidation than adults, and this was especially the case for items that had no neighbours in the English language (James et al., 2018).

It was also stressed that multi-faceted approaches are needed to understand developmental differences, given that children and adults typically show vastly different initial levels of learning, and weaker memories typically receive greater benefits from offline consolidation (Diekelmann et al., 2009). The majority of the experiments presented in this thesis left initial performance levels to vary between groups, and most usually involved more experimental items for adults than children. There were two exceptions to this: in Chapter 5, children and adults were both presented with 15 items embedded in the stories, and in Chapter 3, the third experiment matched adults to children in their initial proportion correct after training. Regardless of the experimental set-up, children still showed greater benefits from offline consolidation than adults across all 8 experiments: they usually improved to a greater extent overnight, and always showed more substantial improvements across the course of the week (*Table 15*). These improvements were largely seen in the recall of word-forms, which we predicted would show the largest developmental differences on the basis that item-level components are less susceptible to prior knowledge influences than associations (Section 1.6.1; van Kesteren et al., 2013).

Interestingly, the advantage for children over adults in consolidation seems to be most consistent for continued consolidation across the course of the week (Figure 19). If children's consolidation benefits are attributable to sleep-associated processes (i.e., their greater amounts of slow-wave sleep; Wilhelm et al., 2012), then it seems plausible that these relative benefits would accumulate across the course of the week. Developmental changes in sleep are also tightly linked to ongoing neural maturation (Buchmann et al., 2011), perhaps suggesting that the developing brain is more plastic to incorporating new words into existing vocabulary knowledge without further exposures. However, one could also consider that children might be more likely to engage with the novel vocabulary in between testing sessions, given that participating in the study is viewed as a class activity that their peers also engage in.

Whilst we did see the predicted benefit of offline consolidation for children relative to adults, the present experiments did not find evidence that adults benefit more from their richer prior knowledge when learning new vocabulary than children:

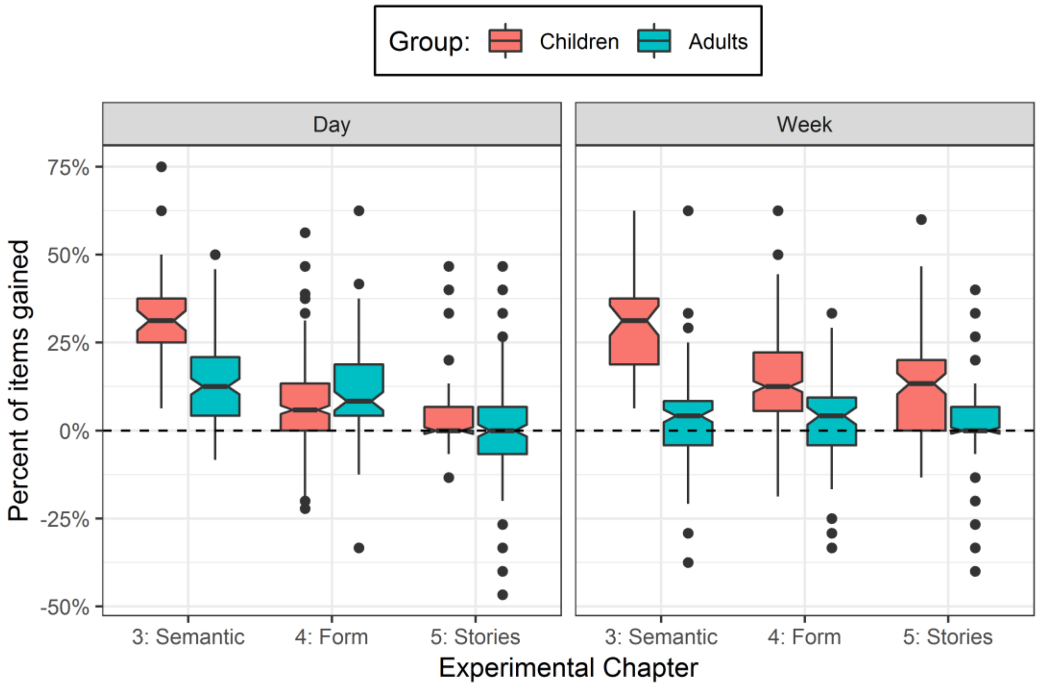


Figure 19. *Box plot summarising improvements in stem completion performance across time points, for each experimental design that assessed memory one day and one week later (Chapters 3-5). The Day data (left) plots percent correct on Day2 – Day 1. The Week data (right) plots percent correct on Day 8 - Day 2. Where there was more than one experiment per child/adult group, data plotted are collapsed across both. Notches mark 95% confidence intervals around the median.*

both groups showed local effects of prior knowledge (Chapters 3-5). This consistency across age groups is in line with a recent study by Brod and Shing (in press), which also showed no evidence that young adults utilise prior knowledge more than children during learning. Whilst, adults were more likely to retain their prior knowledge benefit across the week (Chapter 4), they did not exacerbate with opportunities for neocortical engagement, suggesting that the developmental differences were driven by offline consolidation processes independently of prior knowledge influences. There was some limited evidence that adults and children might engage prior knowledge differently across tasks. In Chapter 3, adults were influenced by semantic knowledge in tasks assessing word-form memory, whereas children only showed effects of semantic neighbours when required to explicitly engage semantic knowledge. Adults also benefited from local knowledge when learning pseudowords from stories, whereas children showed interference from local knowledge under these circumstances, potentially as a consequence of reduced opportunities to engage strategies during learning. These studies hint at the automaticity with which adults might utilise prior knowledge during learning, but do not provide evidence that they benefit more from doing so.

Important next steps require polysomnography to assess whether children's enhanced consolidation can be attributed to their greater proportions of slow-wave sleep compared to adults (as has been demonstrated for procedural memory, Wilhelm et al., 2013). It may also be that children benefited more from repeat testing in the present studies, and/or the spaced exposures that emerged from varied test tasks (i.e., testing recall of word-forms at each test point, but also including subsequent tasks that presented the word-forms). This seems unlikely to fully account for the developmental differences seen in the present studies, considering that repeat testing effects seem to be a relatively ubiquitous phenomenon with very little evidence for a developmental differences (see Toppino & Gerbier, 2014, for a review), and that benefits of sleep outweigh the benefits of repeat testing in both groups (Dumay & Gaskell, 2007; Henderson et al., 2012). However, directly comparing child and adult consolidation with and without repeat testing within the same study of language learning will facilitate a better understanding of how each group capitalises upon these opportunities prior to consolidation. Understanding the relative contributions of different learning mechanisms across development is crucial if we are to understand how interventions may need to be differently targeted across development.

8.5 If the "rich" *always* get richer, how do we help the "poor"?

Finally, it is important to consider the practical implications of the work presented in this thesis. A key motivation for understanding mechanisms behind vocabulary learning is to better target interventions for struggling children – particularly given the necessity of good vocabulary knowledge for broader academic success (e.g., Spencer et al., 2016). All experiments that measured existing vocabulary ability (Chapters 4, 5, 7) showed it to be a strong and highly significant predictor of word learning performance in all tasks. Similarly, children with poor comprehension skills were also generally poor at learning across all aspects of word knowledge. These relationships are attributable to the learner in the present studies, rather than their opportunities and exposure to texts: individuals with good literacy skills learn more vocabulary even when exposure is equated. However, it is not difficult to see how such vast differences in vocabulary ability can emerge across development in combination with environmental factors.

Vocabulary ability did not generally interact with any other variable manipulated in the present experiments, making it challenging to pinpoint particular points of weakness that could be useful for informing practice. Chapter 5 provided some evidence that children with weaker vocabulary were poorer at consolidating semantic mappings after learning from texts, whereas this interaction was not seen for novel objects trained explicitly (Chapter 4). Tighter experimental comparisons are needed to draw strong conclusions in this regard, comparing not only the learning of identical concepts across equal numbers of exposures, but also manipulating whether equivalent semantic information is presented explicitly or embedded within texts. Still, this pattern of findings corroborates previous research that children benefit from direct vocabulary instruction alongside learning from stories (Wilkinson & Houston-Price, 2013).

More interestingly, the experiment in Chapter 7 showed that existing vocabulary was a better predictor of overnight consolidation for words learned earlier in the day than those learned in the evening, suggesting that those with weaker vocabulary knowledge might benefit from learning new vocabulary closer to bedtime. This was an exploratory finding and requires replication, but is consistent with recent findings from Walker et al., (in prep) that show children with poorer vocabulary knowledge to benefit from offline consolidation for novel vocabulary learned in the evening but not the morning. More broadly, these benefits for sleep sooner after learning are supported by earlier studies highlighting better retention for vocabulary learned immediately before bed (Gais et al., 2006; McGregor et al., 2013). It may be then that evening homework could be better geared towards vocabulary learning for struggling individuals in order to maximise potential for remediation. Employing vocabulary intervention closer to bedtimes is becoming increasingly plausible as individuals can engage with independent learning from digital apps. Understanding how to best capitalise on these flexible learning opportunities may prove a promising avenue for developing robust and long-lasting interventions.

8.6 Conclusions

The work presented in this thesis addressed whether the “rich get richer” in vocabulary consolidation: I sought to better understand the ways in which prior linguistic knowledge may support the consolidation of newly learned vocabulary. The experimental evidence was clear in demonstrating that this relationship does not generalise to all learning situations. A dense semantic neighbourhood was shown to slow vocabulary learning, whereas similarity of new words to known forms facilitated word learning. However, these local influences did not appear to underlie individual differences in vocabulary acquisition. Instead, the evidence suggests building semantic connections from context likely matters in supporting later lexical consolidation, and that these connections are better established by those with richer existing vocabulary knowledge. Together these studies demonstrate the complex and multi-faceted aspects of new word learning, and the varied ways in which existing knowledge might interact with this process: the rich persist in getting richer overall, yet there are many ways to be rich.

Appendices

Appendix A

A1.Semantic Neighbour Experiments - Pseudoword stimuli

Word form	Distractor (Experiments 2 & 3)
attay (attie)	attoe
bligma	-
chipod	-
dratus	dratas
myord	myird
oggice	-
glupor	glupear
peflin	peflon
rejele	-
sponto	spontie
trimpy	trimpo
waypo	waypi
ammert	-
bryet (bryat)	bryit
clivel	-
ellnog	ellnig
marpan	marpun
oppult	oppilt
philok	-
plymie	plymoo
shimal (shamal)	shamil
tesdar	-
vorgal (vorgol)	vorgil
whoma	whomie

Note. Words in paretheses note slightly different pronunciations used in Experiments 2 and 3. Only a subset of items were used in Experiment 2; items that did not feature are marked with a '–' in the distractor column.

A2.Semantic Neighbour Experiments - Novel concepts

	Counterbalance Condition 1	**Counterbalance Condition 2**
Low density	A beetle that walks sideways*	A beetle that sleeps upside down
	A tortoise that is hairy*	A tortoise that is sparkly
	A camel that lives in caves*	A camel that lives in cities
	A worm that swings from trees*	A worm that lays blue eggs
	A bull that eats cheese	A bull that eats clothing
	A gorilla that has green skin	A gorilla that has big ears
	A mirror used by witches*	A mirror used by fairies
	A bridge made of pearls*	A bridge made of cardboard
	A tractor used for travelling into space*	A tractor used for carrying drinks
	Rice eaten for breakfast	Rice eaten at birthdays*
	A raisin that has a red outside	A raisin that is orange inside
	A shield that is cylindrical	A shield that is furry
High density	A chicken that sleeps upside down*	A chicken that walks sideways
	A penguin that is sparkly*	A penguin that is hairy
	A deer that lives in cities*	A deer that lives in caves
	A goat that lays blue eggs*	A goat that swings from trees
	A duck that eats clothing	A duck that eats cheese
	A ostrich that has big ears	A ostrich that has green skin
	A sofa used by fairies*	A sofa used by witches
	A shirt made of cardboard*	A shirt made of pearls
	A bike used for carrying drinks*	A bike used for travelling into space
	Lettuce eaten at birthdays	Lettuce eaten for breakfast*
	A prune that is orange inside	A prune that has a red outside
	An apartment that is furry	An apartment that is cylindrical

Note. * marks subset of items used in Experiment 2

A3.Semantic Neighbour Experiment 1 (Adults) – Cued form recall

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula:
## acc ~ time + density + (1 + density | ID) + (1 + density | wordno
)
##    Data: CR
## Control: glmerControl(optimizer = "bobyqa")
##
##      AIC      BIC   logLik deviance df.resid
##   5253.5   5318.7  -2616.8   5233.5     4982
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -5.7102 -0.5969 -0.2991  0.6281  7.3667
##
## Random effects:
##  Groups Name        Variance Std.Dev. Corr
##  ID     (Intercept) 1.62830  1.2760
##         density1    0.06418  0.2533   -0.17
##  wordno (Intercept) 0.71514  0.8457
##         density1    0.04240  0.2059   -0.58
## Number of obs: 4992, groups:  ID, 71; wordno, 24
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.59407    0.23263  -2.554  0.01066 *
## time1        0.30121    0.02574  11.704  < 2e-16 ***
## time2        0.11255    0.04247   2.650  0.00804 **
## density1    -0.07313    0.06315  -1.158  0.24684
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) time1  time2
## time1    -0.012
## time2     0.001  0.024
## density1 -0.342 -0.008 -0.002
```

Interaction pruned from the model: $\chi^2 = 0.28$, $p = .868$

A4.Semantic Neighbour Experiment 1 (Adults) – Cued meaning recall

```
## Cumulative Link Mixed Model fitted with the Laplace approximation
##
## formula: acc ~ time + density + (1 + density | ID) + (1 | itemno)
## data:    def
##
##  link  threshold nobs logLik   AIC     niter     max.grad cond.H
##  logit flexible  4992 -3060.82 6139.64 689(4088) 1.30e-03 9.5e+02
##
## Random effects:
##  Groups Name        Variance Std.Dev. Corr
##  ID     (Intercept) 3.3782   1.8380
##         density2    0.4706   0.6860   -0.566
##  itemno (Intercept) 0.3682   0.6068
## Number of groups:  ID 71,  itemno 24
##
## Coefficients:
##          Estimate Std. Error z value Pr(>|z|)
## time1    -0.10154    0.02480  -4.094 4.24e-05 ***
## time2    -0.14337    0.04296  -3.337 0.000846 ***
## density1 -0.08064    0.13707  -0.588 0.556306
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Threshold coefficients:
##     Estimate Std. Error z value
## 0|1  -0.9909     0.2386  -4.152
## 1|2  -0.7962     0.2385  -3.339
```

Interaction pruned from the model: LR stat = 1.16, *p* = .561

A5.Semantic Neighbour Experiment 1 (Adults) – Semantic categorisation (accuracy)

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: acc ~ time * density + (1 + density | ID) + (1 | itemno)
##    Data: semCat
## Control: glmerControl(optimizer = "bobyqa")
##
##      AIC      BIC   logLik deviance df.resid
##   4972.3   5037.5  -2476.2   4952.3     4979
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -5.4967  0.2052  0.3850  0.5582  1.5013
##
## Random effects:
##  Groups Name        Variance Std.Dev. Corr
##  ID     (Intercept) 0.6192   0.7869
##         density1    0.0432   0.2079   -0.05
##  itemno (Intercept) 0.2749   0.5243
## Number of obs: 4989, groups:  ID, 71; itemno, 24
##
## Fixed effects:
##                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)     1.450817   0.148005   9.802   <2e-16 ***
## time1          -0.006118   0.025244  -0.242    0.809
## time2          -0.035929   0.044237  -0.812    0.417
## density1        0.044868   0.116721   0.384    0.701
## time1:density1 -0.036232   0.025210  -1.437    0.151
## time2:density1  0.012736   0.044034   0.289    0.772
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) time1  time2  dnsty1 tm1:d1
## time1        0.005
## time2        0.002  0.005
## density1    -0.002 -0.006  0.002
## tim1:dnsty1 -0.007  0.027  0.003  0.007
## tim2:dnsty1  0.002  0.003 -0.003  0.002  0.002
```

A6.Semantic Neighbour Experiment 1 (Adults) – Semantic categorisation (RT)

```
## Linear mixed model fit by maximum likelihood . t-tests use
##   Satterthwaite's method [lmerModLmerTest]
## Formula: logRT ~ time + density + (1 + time + density | ID) + (1 |
itemno)
##    Data: trimRT
##
##      AIC      BIC   logLik deviance df.resid
##    756.8    856.6   -362.4    724.8     3760
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.6608 -0.6317 -0.1152  0.5110  5.2209
##
## Random effects:
##  Groups   Name        Variance  Std.Dev. Corr
##  ID       (Intercept) 0.0263619 0.16236
##           time1       0.0020984 0.04581  -0.30
##           time2       0.0033775 0.05812   0.03  0.18
##           density1    0.0006896 0.02626   0.12 -0.10  0.40
##  itemno   (Intercept) 0.0021257 0.04611
##  Residual             0.0627583 0.25052
## Number of obs: 3776, groups:  ID, 70; itemno, 24
##
## Fixed effects:
##               Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)  6.7891463  0.0219962 86.0771707 308.651  < 2e-16 ***
## time1       -0.0515216  0.0062244 69.8888725  -8.277 5.70e-12 ***
## time2       -0.0368099  0.0087964 67.0756337  -4.185 8.49e-05 ***
## density1    -0.0001153  0.0107252 28.0825103  -0.011    0.992
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) time1  time2
## time1    -0.229
## time2     0.035  0.143
## density1  0.030 -0.024  0.094
```

Interaction pruned from the model: χ^2 = 1.85, *p* = .396

A7.Semantic Neighbour Experiment 2 (Children) – Cued form recall

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: acc ~ time + density + (1 | ID) + (1 | wordno)
##    Data: CR
## Control: glmerControl(optimizer = "bobyqa")
##
##      AIC      BIC   logLik deviance df.resid
##   2290.5   2325.3  -1139.3   2278.5     2410
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -5.4123 -0.5063  0.0949  0.4945 12.1755
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  ID     (Intercept) 0.9907   0.9954
##  wordno (Intercept) 0.9143   0.9562
## Number of obs: 2416, groups:  ID, 51; wordno, 16
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.01928    0.28210   0.068    0.946
## time1        0.95322    0.04518  21.097   <2e-16 ***
## time2        0.91478    0.06851  13.353   <2e-16 ***
## density1     0.06838    0.24505   0.279    0.780
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) time1  time2
## time1    -0.012
## time2     0.023  0.214
## density1  0.001  0.009  0.000
```

Interaction pruned from the model: χ^2 = 0.88, p = .644

A8.Semantic Neighbour Experiment 2 (Children) – Form recognition

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: acc ~ time + density + (1 + density | ID) + (1 | itemno)
##    Data: recog
## Control: glmerControl(optimizer = "bobyqa")
##
##      AIC      BIC   logLik deviance df.resid
##   1397.0   1443.4   -690.5   1381.0     2408
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -9.5940  0.1282  0.2084  0.3255  1.7946
##
## Random effects:
##  Groups Name        Variance Std.Dev. Corr
##  ID     (Intercept) 1.0410   1.0203
##         density1    0.2398   0.4897   -0.12
##  itemno (Intercept) 0.5104   0.7144
## Number of obs: 2416, groups:  ID, 51; itemno, 16
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.88991    0.25398  11.378  < 2e-16 ***
## time1        0.39168    0.04853   8.070 7.01e-16 ***
## time2        0.21102    0.10138   2.081   0.0374 *
## density1    -0.29020    0.21673  -1.339   0.1806
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) time1  time2
## time1     0.143
## time2     0.041  0.086
## density1 -0.066 -0.009 -0.001
```

Interaction pruned from the model: $\chi^2 = 1.76$, $p = .414$

A9.Semantic Neighbour Experiment 2 (Children) – Cued meaning recall

```
## Cumulative Link Mixed Model fitted with the Laplace approximation
##
## formula: acc ~ time + density + (1 | ID) + (1 | itemno)
## data:    def
##
##  link  threshold nobs logLik   AIC     niter     max.grad cond.H
##  logit flexible  2416 -1009.32 2032.63 388(2089) 2.50e-04 9.3e+02
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  ID     (Intercept) 1.9485   1.3959
##  itemno (Intercept) 0.4593   0.6777
## Number of groups:  ID 51,  itemno 16
##
## Coefficients:
##           Estimate Std. Error z value Pr(>|z|)
## time1     0.009438   0.043296   0.218  0.82743
## time2    -0.068104   0.075694  -0.900  0.36826
## density1 -0.475824   0.181695  -2.619  0.00882 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Threshold coefficients:
##     Estimate Std. Error z value
## 0|1   2.1059     0.2738   7.690
## 1|2   2.1909     0.2743   7.987
```

Interaction pruned from the model: LR stat = 0.75, p = .687

A10. Semantic Neighbour Experiment 2 (Children) – Semantic categorisation (accuracy)

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: acc ~ time + density + (1 | ID) + (1 | itemno)
##    Data: semCat
## Control: glmerControl(optimizer = "bobyqa")
##
##      AIC      BIC   logLik deviance df.resid
##   3228.5   3263.2  -1608.3   3216.5     2410
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.1593 -1.0866  0.6541  0.8294  1.2761
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  ID     (Intercept) 0.08988  0.2998
##  itemno (Intercept) 0.10906  0.3302
## Number of obs: 2416, groups:  ID, 51; itemno, 16
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.40084    0.10198   3.930 8.48e-05 ***
## time1        0.02358    0.02980   0.791    0.429
## time2       -0.02841    0.05223  -0.544    0.586
## density1     0.01905    0.09282   0.205    0.837
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) time1  time2
## time1     0.008
## time2    -0.010 -0.017
## density1  0.000  0.000  0.000
```

Interaction pruned from the model: $\chi^2 = 0.80$, $p = .669$

A11. Semantic Neighbour Experiment 2 (Children) – Semantic categorisation (RT)

```
## Linear mixed model fit by maximum likelihood . t-tests use
##   Satterthwaite's method [lmerModLmerTest]
## Formula: logRT ~ time + density + (1 + density | ID) + (1 | itemn
o)
##    Data: trimRT
##
##      AIC      BIC   logLik deviance df.resid
##   1596.0   1641.7   -789.0   1578.0     1176
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.1480 -0.6663 -0.1587  0.5507  3.2052
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  ID       (Intercept) 0.033437 0.18286
##           density1    0.005029 0.07091  0.03
##  itemno   (Intercept) 0.001714 0.04140
##  Residual             0.203949 0.45161
## Number of obs: 1185, groups:  ID, 40; itemno, 16
##
## Fixed effects:
##               Estimate Std. Error         df t value Pr(>|t|)
## (Intercept)  7.410e+00  3.347e-02  3.991e+01 221.364  < 2e-16 ***
## time1       -7.197e-02  9.393e-03  1.124e+03  -7.662 3.93e-14 ***
## time2       -2.996e-02  1.627e-02  1.121e+03  -1.841   0.0658 .
## density1     3.018e-02  2.025e-02  2.302e+01   1.490   0.1497
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) time1  time2
## time1     0.001
## time2    -0.009 -0.019
## density1  0.011 -0.010 -0.002
```

Interaction pruned from the model: $\chi^2 = 0.53$, $p = .767$

A12. Semantic Neighbour Experiment 3 (Adults) – Cued form recall

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: acc ~ time + density + (1 + density | ID) + (1 | word)
##    Data: CR
## Control: glmerControl(optimizer = "bobyqa")
##
##      AIC      BIC   logLik deviance df.resid
##   4497.3   4548.8  -2240.7   4481.3     4600
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.4752 -0.5432 -0.2998  0.5458  6.1626
##
## Random effects:
##  Groups Name        Variance Std.Dev. Corr
##  ID     (Intercept) 1.54591  1.2433
##         density1    0.04986  0.2233   -0.16
##  word   (Intercept) 0.78600  0.8866
## Number of obs: 4608, groups:  ID, 67; word, 24
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.10911    0.24026  -4.616 3.91e-06 ***
## time1        0.38719    0.02908  13.316  < 2e-16 ***
## time2        0.07067    0.04569   1.547   0.1219
## density1    -0.10866    0.04859  -2.236   0.0253 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) time1  time2
## time1    -0.029
## time2     0.010  0.059
## density1 -0.052 -0.017 -0.003
```

Interaction pruned from the model: $\chi^2 = 1.92$, $p = .383$

A13. Semantic Neighbour Experiment 3 (Adults) – Form recognition

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: acc ~ time + density + (1 + density | word) + (1 | ID)
##    Data: recog
## Control: glmerControl(optimizer = "bobyqa")
##
##      AIC      BIC   logLik deviance df.resid
##   1972.8   2023.5   -978.4   1956.8     4192
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -8.0765  0.1071  0.1839  0.2837  0.9439
##
## Random effects:
##  Groups Name        Variance Std.Dev. Corr
##  ID     (Intercept) 1.1791   1.0858
##  word   (Intercept) 0.9232   0.9609
##         density1    0.1103   0.3321   0.68
## Number of obs: 4200, groups:  ID, 65; word, 24
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.38264    0.26233  12.895  < 2e-16 ***
## time1        0.11018    0.04255   2.589  0.00962 **
## time2       -0.04296    0.07728  -0.556  0.57828
## density1     0.21232    0.10771   1.971  0.04870 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) time1 time2
## time1    0.024
## time2    0.013  0.027
## density1 0.397  0.004 0.000
```

Interaction pruned from the model: χ^2 = 0.10, p = .951

A14. Semantic Neighbour Experiment 3 (Adults) – Cued meaning recall

```
## Cumulative Link Mixed Model fitted with the Laplace approximation
##
## formula: acc ~ time + density + (1 + density | ID) + (1 | itemno)
## data:    def
##
##  link  threshold nobs logLik   AIC     niter     max.grad cond.H
##  logit flexible  4536 -2140.07 4298.15 565(3281) 2.90e-03 8.4e+02
##
## Random effects:
##  Groups Name        Variance Std.Dev. Corr
##  ID     (Intercept) 1.6869   1.2988
##         density2    1.2899   1.1357   -0.629
##  itemno (Intercept) 0.6463   0.8039
## Number of groups:  ID 66,  itemno 24
##
## Coefficients:
##           Estimate Std. Error z value Pr(>|z|)
## time1    -0.009897   0.029638  -0.334    0.738
## time2    -0.015426   0.053488  -0.288    0.773
## density1 -0.071082   0.186777  -0.381    0.704
##
## Threshold coefficients:
##     Estimate Std. Error z value
## 0|1   1.8623     0.2157   8.632
## 1|2   1.9846     0.2161   9.182
```

Interaction pruned from the model: LR stat = 0.01, p = .995

A15. Semantic Neighbour Experiment 3 (Adults) – Semantic categorisation (accuracy)

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: acc ~ time + density + (1 | ID) + (1 | itemno)
##    Data: semCat
## Control: glmerControl(optimizer = "bobyqa")
##
##      AIC      BIC   logLik deviance df.resid
##   5688.3   5726.4  -2838.2   5676.3     4194
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.9825 -1.0498  0.6715  0.8874  1.2909
##
## Random effects:
##  Groups Name        Variance Std.Dev.
##  ID     (Intercept) 0.05497  0.2345
##  itemno (Intercept) 0.10207  0.3195
## Number of obs: 4200, groups:  ID, 65; itemno, 24
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.27322    0.07829   3.490 0.000483 ***
## time1       -0.01405    0.02299  -0.611 0.541020
## time2       -0.03565    0.03886  -0.917 0.358951
## density1     0.04800    0.07253   0.662 0.508077
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) time1  time2
## time1    -0.013
## time2     0.027  0.051
## density1  0.002  0.000  0.000
```

Interaction pruned from the model: χ^2 = 1.77, p = .414

A16. Semantic Neighbour Experiment 3 (Adults) – Semantic categorisation (RT)

```
## Linear mixed model fit by maximum likelihood . t-tests use
##   Satterthwaite's method [lmerModLmerTest]
## Formula: logRT ~ time + density + (1 + time | ID) + (1 | itemno)
##    Data: trimRT
##
##      AIC      BIC   logLik deviance df.resid
##   1398.0   1464.5   -687.0   1374.0     1877
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.1467 -0.6323 -0.0734  0.5949  3.4407
##
## Random effects:
##  Groups   Name        Variance  Std.Dev. Corr
##  ID       (Intercept) 0.0569382 0.23862
##           time1       0.0054786 0.07402  0.62
##           time2       0.0087954 0.09378  0.12 0.26
##  itemno   (Intercept) 0.0005499 0.02345
##  Residual             0.1055014 0.32481
## Number of obs: 1889, groups:  ID, 49; itemno, 24
##
## Fixed effects:
##              Estimate Std. Error        df t value Pr(>|t|)
## (Intercept)  6.854348   0.035620 47.421019 192.432  < 2e-16 ***
## time1       -0.080613   0.012350 45.607432  -6.527 4.85e-08 ***
## time2       -0.048206   0.017611 39.582669  -2.737  0.00923 **
## density1     0.008922   0.008967 23.694375   0.995  0.32980
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) time1  time2
## time1     0.516
## time2     0.145  0.258
## density1 -0.002  0.003 -0.004
```

Interaction pruned from the model: $\chi^2 = 0.22$, $p = .897$

Appendix B

B1. Form Neighbour Experiments - Pseudoword stimuli

No	One	Many
femod*	pungus*	ballow*
marpan*	rafar*	dester*
parung*	regby*	gumble*
peflin*	suburt*	mowel*
tesdar*	tabric*	nusty*
vorgal*	wabon*	solly*
hovvy	lentig	fallet
sabam	pilbar	hender

Note. Asterisk denotes subset selected for Experiment 3

B2. Form Neighbour Experiment 1 (Children) - Receptive vocabulary measures

We also included a receptive vocabulary measure in the form of a shortened adapted version of the British Picture Vocabulary Scale Third Edition (Dunn, Dunn, Styles, & Sewell, 2009). A subset of items were administered that ranged in difficulty from the recommended start point for the youngest children, and increased in difficulty until the item that was approximately two standard deviations above the average score for the oldest children. Every fourth item was selected, resulting in 27 test items (plus two training items with feedback). Children were read each word aloud and asked to select which of four pictures presented via a projector at the front of the classroom represented the word. Children circled the number corresponding to their answer in answer booklets.

Analyses

Neither vocabulary measure was a significantly stronger correlate with overall cued recall performance than the other (z = 0.34, p = .734). Therefore, we proceeded to use the expressive vocabulary measure for comparability with key studies of interest (Storkel & Hoover, 2011; Henderson et al., 2015).

Performance on the expressive vocabulary task was significantly better predictor for average recognition performance (r = .53) than receptive vocabulary (r = .36; z = 2.82, p = .005), and was therefore used as the vocabulary predictor for recognition.

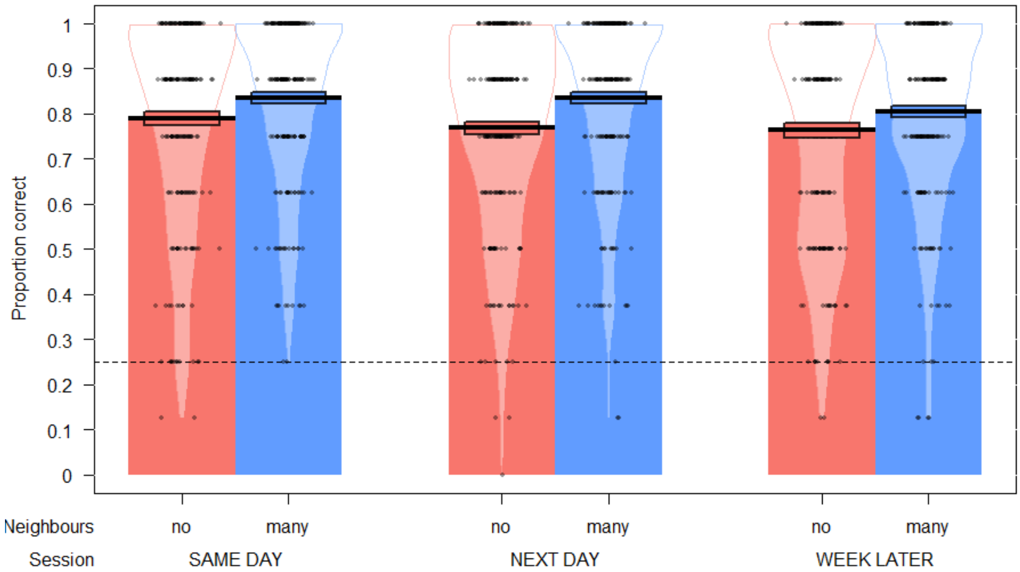
B3. Form Neighbour Experiment 1 (Children) - Recognition analyses

Predictors of recognition performance in Experiment 1

Fixed effects	*b*	*SE*	*z*	*p*	
(Intercept)	2.00	0.19	10.72	<.001	
delay1	-0.04	0.03	-1.54	.124	
delay2	-0.08	0.05	-1.73	.084	
neighb	0.27	0.17	1.60	.110	
vocab	**0.79**	**0.09**	**8.54**	**<.001**	
delay1:neighb	0.01	0.03	0.48	.633	
delay2:neighb	-0.05	0.05	-1.13	.260	
neighb:vocab	0.04	0.04	0.85	.394	
Random effects	**Variance**	***SD***	**Correlations**		
participant: (intercept)	1.43	1.20			
participant: (slope) neighb	0.04	0.20	-0.05		
item: (intercept)	0.43	0.66			
item: (slope) vocab	0.01	0.12	0.23		
item: (slope) delay1	0.01	0.07	0.47	-0.60	
item: (slope) delay2	0.02	0.12	0.15	-0.64	0.09

Note. Model formed from 11,055 observations, collected from 232 participants across 16 items. Orthogonal contrasts were used for the three-level factor of session: delay1 (Session 1 vs. Sessions 2&3), delay2 (Session 2 vs. Session 3).

B4. Form Neighbour Experiment 1 (Children) – Recognition graph



RDI plot of recognition performance in Experiment 1, plotted by neighbour condition and test session. Thick black horizontal bars represent the mean for each condition, and surrounding boxes marked +/-1 standard error of the mean. The dashed line indicates chance performance.

B5. Form Neighbour Experiment 2 (Adults) - Receptive vocabulary measures

Receptive vocabulary was measured using a shortened adapted version of the Peabody Picture Vocabulary Test 4th Edition (Dunn & Dunn, 2007). A subset of items were selected that began at the recommended start point for adults and increased in difficulty until the item that was approximately two standard deviations above the average score. Every third item was selected, resulting in 23 test items (plus two training items with feedback). Participants were presented with one item at a time, and asked to select which of four pictures presented in the web browser represented the written word.

Analyses

As with Experiment 1, neither vocabulary measure correlated more strongly with average recall performance than the other (z = 0.48, p = .632), and so we continued to use the expressive vocabulary measure as the predictor in this analysis.

Neither vocabulary measure better predicted performance in the recognition task (z = 0.24, p = .811). Modelling therefore proceeded with expressive vocabulary as in all previous analyses.

B6. Form Neighbour Experiment 2 (Adults) - Recognition analyses

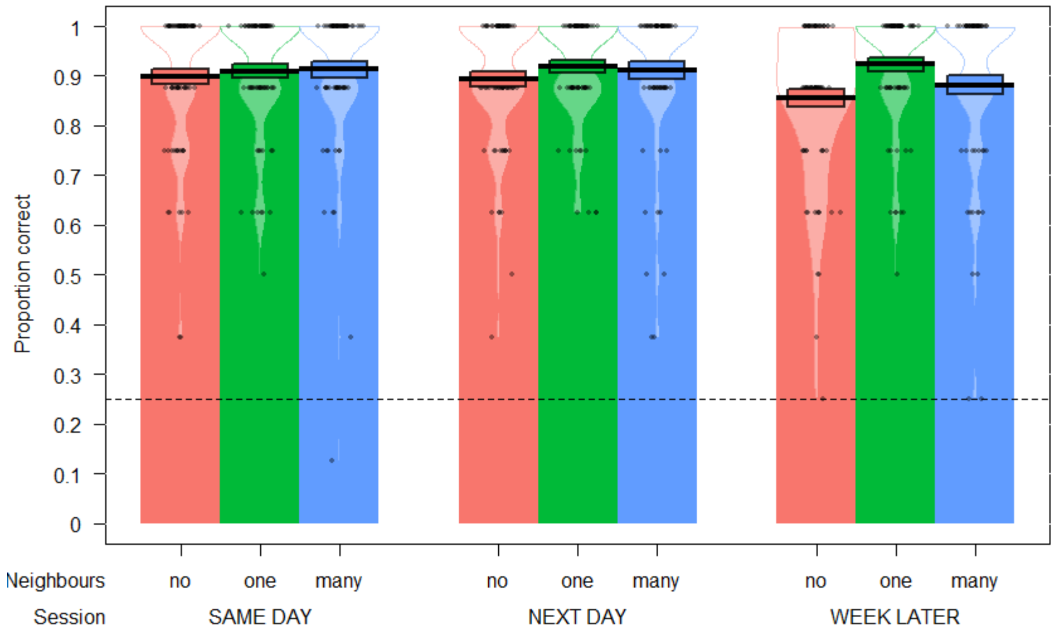
Predictors of recognition performance in Experiment 2

Fixed effects	*b*	*SE*	*z*	*p*
(Intercept)	2.93	0.22	13.42	<.001
delay1	-0.05	0.04	-1.34	.181
delay2	-0.10	0.06	-1.66	.097
neighb1	0.22	0.12	1.77	.077
neighb2	-0.02	0.22	-0.09	.925
vocab	**0.36**	**0.15**	**2.43**	**.015**
delay1:neighb1	0.03	0.02	1.07	.285
delay2:neighb1	0.05	0.04	1.24	.216
delay1:neighb2	-0.06	0.04	-1.39	.165
delay2:neighb2	-0.13	0.08	-1.64	.100
delay1:vocab	0.00	0.03	0.10	.922
delay2:vocab	0.05	0.06	0.96	.339
neighb1:vocab	0.03	0.05	0.58	.562
neighb2:vocab	-0.08	0.09	-0.88	.379
delay1:neighb1:vocab	-0.00	0.02	-0.01	.995
delay2:neighb1:vocab	0.06	0.04	1.55	.122
delay1:neighb2:vocab	0.04	0.04	0.89	.374
delay2:neighb2:vocab	-0.02	0.07	-0.24	.810

Random effects	Variance	*SD*	Correlations	
participant: (intercept)	1.25	1.12		
participant: (slope) neighb1	0.03	0.17	0.57	
participant: (slope) neighb2	0.08	0.28	0.70	0.99
item: (intercept)	0.62	0.79		
item: (slope) vocab	0.03	0.18	0.14	

Note. Model formed from 5592 observations, collected from 79 participants across 24 items. Orthogonal contrasts were used for three-level factors: delay1 (Session 1 vs. Sessions 2&3), delay2 (Session 2 vs. Session 3), neighb1 (no vs. one&many), neighb2 (one vs. many).

B7. Form Neighbour Experiment 2 (Adults) – Recognition graph



RDI plot of recognition performance in Experiment 2, plotted by neighbour condition and test session. Thick black horizontal bars represent the mean for each condition, and surrounding boxes marked +/-1 standard error of the mean. The dashed line indicates chance performance.

B8. Form Neighbour Experiment 3 (Children) – Recognition analyses

Predictors of form recognition performance in Experiment 3

Fixed effects	*b*	*SE*	*z*	*p*		
(Intercept)	2.67	0.19	13.89	<.001		
delay1	**0.28**	**0.06**	**4.94**	**<.001**		
delay2	**0.39**	**0.11**	**3.48**	**<.001**		
neighb1	-0.07	0.11	-0.59	.554		
neighb2	0.08	0.19	0.44	.659		
vocab	**0.53**	**0.14**	**3.89**	**<.001**		
delay1:neighb1	**-0.06**	**0.03**	**-2.18**	**.029**		
delay2:neighb1	-0.01	0.06	-0.09	.928		
delay1:neighb2	**-0.11**	**0.05**	**-2.13**	**.034**		
delay2:neighb2	0.00	0.09	-0.05	.961		
delay1:vocab	-0.07	0.06	-1.10	.271		
delay2:vocab	0.14	0.10	1.36	.175		
neighb1:vocab	0.07	0.05	1.34	.180		
neighb2:vocab	0.07	0.08	0.87	.384		
Random effects	Variance	*SD*	Correlations			
participant: (intercept)	0.71	0.84				
participant: (slope) delay1	0.06	0.25	0.60			
participant: (slope) delay2	0.06	0.24	0.53	0.83		
participant: (slope) neighb1	0.04	0.20	-0.46	-0.74	-0.25	
participant: (slope) neighb2	0.07	0.26	-0.21	-0.35	0.22	0.89
item: (intercept)	0.33	0.58				
item: (slope) vocab	0.06	0.24	0.66			

Note. Model formed from 3834 observations, collected from 72 participants across 18 items. Orthogonal contrasts were used for three-level factors: delay1 (Session 1 vs. Sessions 2&3), delay2 (Session 2 vs. Session 3), neighb1 (no vs. one&many), neighb2 (one vs. many).

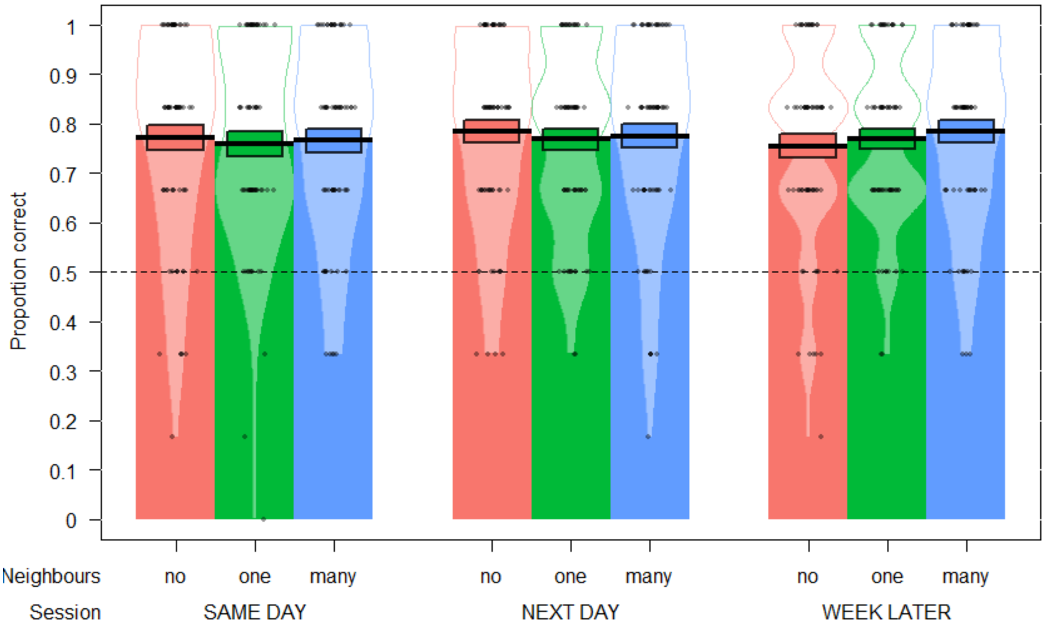
B9. Form Neighbour Experiment 3 (Children) – Picture-form recognition analyses

Fixed effects	*b*	*SE*	*z*	*p*
(Intercept)	1.53	0.20	7.57	<.001
delay1	0.02	0.03	0.63	.532
delay2	-0.03	0.05	-0.52	.602
vocab	**0.34**	**0.11**	**3.05**	**.002**
neighb1	0.03	0.13	0.27	.786
neighb2	0.05	0.23	0.24	.809
delay1:vocab	0.04	0.03	1.45	.148
delay2:vocab	-0.01	0.05	-0.14	.893
delay1:neighb1	0.01	0.02	0.45	.652
delay2:neighb1	0.04	0.04	1.13	.260
delay1:neighb2	0.00	0.04	0.13	.901
delay2:neighb2	0.02	0.06	0.25	.802
vocab:neighb1	-0.02	0.05	-0.49	.625
vocab:neighb2	0.03	0.09	0.30	.764

Random effects	Variance	*SD*	Correlations	
participant: (intercept)	0.55	0.74		
participant: (slope) neighb1	0.01	0.12	-0.28	
participant: (slope) neighb2	0.12	0.35	0.11	0.23
item: (intercept)	0.55	0.74		
item: (slope) vocab	0.05	0.22	0.89	

Note. Model formed from 3816 observations, collected from 72 participants across 18 items. Orthogonal contrasts were used for three-level factors: delay1 (Session 1 vs. Sessions 2&3), delay2 (Session 2 vs. Session 3), neighb1 (no vs. one&many), neighb2 (one vs. many).

B10. Form Neighbour Experiment 3 (Children) – Picture-form recognition graph



RDI plot of picture-form recognition performance in Experiment 3, plotted by neighbour condition and test session. Thick black horizontal bars represent the mean for each condition, and surrounding boxes marked +/-1 standard error of the mean. The dashed line indicates chance performance.

Appendix C

C1. Story Experiment 1 (Children) – Cued form recall

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula:
## acc ~ session * neighb * vocabS + (1 + neighb | ID) + (1 + vocabS
|
##     item)
##    Data: stemComp
## Control:
## glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 1e+06)
)
##
##      AIC      BIC   logLik deviance df.resid
##   2220.5   2392.5  -1083.3   2166.5     4293
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.2035 -0.2865 -0.1487 -0.0706 18.0603
##
## Random effects:
##  Groups Name        Variance Std.Dev. Corr
##  ID     (Intercept) 1.02257  1.0112
##         neighb1     0.05949  0.2439   -0.16
##         neighb2     0.24874  0.4987   -0.02  0.15
##  item   (Intercept) 0.93341  0.9661
##         vocabS      0.05646  0.2376   0.66
## Number of obs: 4320, groups:  ID, 97; item, 15
##
## Fixed effects:
##                         Estimate Std. Error z value Pr(>|z|)
## (Intercept)            -3.444804   0.299546 -11.500  < 2e-16 ***
## session1                0.676171   0.073660   9.180  < 2e-16 ***
## session2                0.838915   0.078952  10.626  < 2e-16 ***
## neighb1                 0.286008   0.203775   1.404    0.160
## neighb2                 0.135256   0.336752   0.402    0.688
## vocabS                  0.693534   0.163743   4.236 2.28e-05 ***
## session1:neighb1        0.017549   0.055096   0.319    0.750
## session2:neighb1       -0.058948   0.060233  -0.979    0.328
## session1:neighb2       -0.073280   0.084525  -0.867    0.386
## session2:neighb2       -0.026053   0.087567  -0.298    0.766
## session1:vocabS        -0.007169   0.071471  -0.100    0.920
## session2:vocabS        -0.003727   0.080611  -0.046    0.963
## neighb1:vocabS         -0.083202   0.089114  -0.934    0.350
## neighb2:vocabS          0.092238   0.143173   0.644    0.519
## session1:neighb1:vocabS  0.081339   0.050447   1.612    0.107
## session2:neighb1:vocabS  0.022723   0.060355   0.376    0.707
## session1:neighb2:vocabS -0.023929   0.086720  -0.276    0.783
## session2:neighb2:vocabS  0.032986   0.089442   0.369    0.712
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

C2. Story Experiment 1 (Children) – Form recognition

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: acc ~ session * neighb + vocabS * neighb + session * voc
abS +
##     (1 | ID) + (1 + vocabS | item)
##    Data: FR
## Control: glmerControl(optimizer = "bobyqa")
##
##      AIC      BIC   logLik deviance df.resid
##   4704.1   4818.7  -2334.0   4668.1     4287
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.0304 -0.8724  0.4514  0.6082  1.6936
##
## Random effects:
##  Groups Name        Variance Std.Dev. Corr
##  ID     (Intercept) 0.27619  0.5255
##  item   (Intercept) 0.07829  0.2798
##         vocabS      0.02230  0.1493   0.99
## Number of obs: 4305, groups:  ID, 97; item, 15
##
## Fixed effects:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)       1.159330   0.097938  11.837  < 2e-16 ***
## session1          0.225817   0.025193   8.963  < 2e-16 ***
## session2          0.170189   0.047521   3.581 0.000342 ***
## neighb1          -0.121607   0.057913  -2.100 0.035746 *
## neighb2          -0.026543   0.099028  -0.268 0.788674
## vocabS            0.296115   0.076234   3.884 0.000103 ***
## session1:neighb1  0.030851   0.018063   1.708 0.087644 .
## session2:neighb1  0.043883   0.033785   1.299 0.193980
## session1:neighb2  0.010490   0.029714   0.353 0.724077
## session2:neighb2  0.080166   0.056088   1.429 0.152921
## neighb1:vocabS   -0.069790   0.038431  -1.816 0.069372 .
## neighb2:vocabS   -0.007632   0.064497  -0.118 0.905807
## session1:vocabS   0.035607   0.025093   1.419 0.155900
## session2:vocabS   0.062152   0.047647   1.304 0.192087
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interaction pruned from the model: $\chi^2 = 0.53$, $p = .970$

C3. Story Experiment 1 (Children) – Picture-form recognition

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: acc ~ session * neighb + vocabS * neighb + session * voc
abS +
##     (1 + neighb | ID) + (1 + vocabS | item)
##    Data: PR
## Control: glmerControl(optimizer = "bobyqa")
##
##      AIC      BIC   logLik deviance df.resid
##   4808.2   4953.5  -2381.1   4762.2     4068
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.3658 -1.0023  0.4862  0.6575  1.4262
##
## Random effects:
##  Groups Name        Variance Std.Dev. Corr
##  ID     (Intercept) 0.31665  0.5627
##         neighb1     0.02573  0.1604    0.31
##         neighb2     0.07505  0.2740   -0.06 -0.23
##  item   (Intercept) 0.04450  0.2110
##         vocabS      0.04379  0.2093   0.56
## Number of obs: 4091, groups:  ID, 92; item, 15
##
## Fixed effects:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)        0.970422   0.089013  10.902  < 2e-16 ***
## session1           0.102831   0.025221   4.077 4.56e-05 ***
## session2          -0.007887   0.045301  -0.174 0.861780
## neighb1           -0.004995   0.049876  -0.100 0.920220
## neighb2            0.001411   0.086706   0.016 0.987019
## vocabS             0.352621   0.088952   3.964 7.36e-05 ***
## session1:neighb1  0.025598   0.017727   1.444 0.148734
## session2:neighb1  0.046134   0.031428   1.468 0.142116
## session1:neighb2  0.050173   0.030729   1.633 0.102522
## session2:neighb2  0.031218   0.055106   0.567 0.571040
## neighb1:vocabS    0.043897   0.049558   0.886 0.375739
## neighb2:vocabS    0.138487   0.086203   1.607 0.108158
## session1:vocabS   0.091069   0.025807   3.529 0.000417 ***
## session2:vocabS   0.021885   0.046592   0.470 0.638557
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interaction pruned from the model: χ^2 = 1.90, *p* = .754

C4. Story Experiment 2 (Adults) – Cued form recall

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: acc ~ session * neighb + session * vocabS + neighb * voc
abS +
##     (1 + neighb | ID) + (1 + vocabS | item)
##    Data: stemComp
## Control: glmerControl(optimizer = "bobyqa")
##
##      AIC      BIC   logLik deviance df.resid
##   4936.9   5090.4  -2445.5   4890.9     5827
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.7276 -0.4493 -0.2640 -0.1116  6.3904
##
## Random effects:
##  Groups Name        Variance Std.Dev. Corr
##  ID     (Intercept) 1.83146  1.3533
##         neighb1     0.07288  0.2700    0.35
##         neighb2     0.32715  0.5720   -0.04  0.12
##  item   (Intercept) 0.41302  0.6427
##         vocabS      0.02432  0.1559   0.00
## Number of obs: 5850, groups:  ID, 130; item, 15
##
## Fixed effects:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -1.9128957  0.2109239  -9.069  < 2e-16 ***
## session1          0.0625037  0.0270554   2.310 0.020876 *
## session2          0.1361171  0.0457950   2.972 0.002956 **
## neighb1           0.1657845  0.1253384   1.323 0.185937
## neighb2           0.4700863  0.2179749   2.157 0.031036 *
## vocabS            0.4827190  0.1349641   3.577 0.000348 ***
## session1:neighb1  0.0104060  0.0192964   0.539 0.589699
## session2:neighb1  0.0003329  0.0328160   0.010 0.991906
## session1:neighb2  0.0141498  0.0318926   0.444 0.657282
## session2:neighb2 -0.0364620  0.0541262  -0.674 0.500535
## session1:vocabS  -0.0171180  0.0277945  -0.616 0.537976
## session2:vocabS   0.0305816  0.0469195   0.652 0.514537
## neighb1:vocabS   -0.0232294  0.0490411  -0.474 0.635733
## neighb2:vocabS   -0.1185620  0.0884394  -1.341 0.180050
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Three-way interaction pruned from the model: $\chi^2 = 2.21$, $p = .697$

C5. Story Experiment 2 (Adults) – Recognition

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: acc ~ session * neighb + session * vocabS + neighb * voc
abS +
##     (1 + neighb | ID) + (1 | item)
##    Data: recog
## Control: glmerControl(optimizer = "bobyqa")
##
##      AIC      BIC   logLik deviance df.resid
##   6268.1   6408.2  -3113.0   6226.1     5829
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -5.6817 -0.7453  0.3528  0.5987  2.1527
##
## Random effects:
##  Groups Name        Variance Std.Dev. Corr
##  ID     (Intercept) 1.15520  1.0748
##         neighb1     0.04878  0.2209   0.32
##         neighb2     0.22980  0.4794   0.25 0.29
##  item   (Intercept) 0.18597  0.4312
## Number of obs: 5850, groups:  ID, 130; item, 15
##
## Fixed effects:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)       1.165666   0.150785   7.731 1.07e-14 ***
## session1         -0.107829   0.023127  -4.663 3.12e-06 ***
## session2         -0.056229   0.038802  -1.449 0.147303
## neighb1           0.073302   0.084965   0.863 0.388286
## neighb2           0.093157   0.150427   0.619 0.535728
## vocabS            0.352840   0.099766   3.537 0.000405 ***
## session1:neighb1  0.020256   0.016086   1.259 0.207941
## session2:neighb1  0.005643   0.026843   0.210 0.833492
## session1:neighb2 -0.002038   0.028574  -0.071 0.943151
## session2:neighb2 -0.009758   0.048304  -0.202 0.839903
## session1:vocabS   0.007591   0.021730   0.349 0.726832
## session2:vocabS   0.002301   0.036809   0.063 0.950151
## neighb1:vocabS    0.009861   0.029304   0.337 0.736493
## neighb2:vocabS    0.001401   0.057731   0.024 0.980635
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Three-way interaction pruned from the model: $\chi^2 = 1.18$, $p = .881$

Appendix D

D1. AM-PM Experiment – Word stimuli

list	category	item
1	birds	hoopoe
1	birds	galah
1	birds	drongo
1	trees	baobab
1	trees	deglupta
1	trees	banyan
1	monkeys	gelada
1	monkeys	mandrill
1	monkeys	lesula
1	cats	caracal
1	cats	oncilla
1	cats	serval
2	fish	oranda
2	fish	blenny
2	fish	dorado
2	flowers	trillium
2	flowers	zinnia
2	flowers	protea
2	rodents	marmot
2	rodents	woylie
2	rodents	agouti
2	dogs	dingo
2	dogs	komondor
2	dogs	lycaon

D2.AM-PM Experiment – Definitions analysis (24-hour)

```
## Cumulative Link Mixed Model fitted with the Laplace approximation
##
## formula: acc ~ group * learnTime + learnTime * time + group * tim
e + (1 +
##     learnTime | ID) + (1 + learnTime | item)
## data:    def
##
##  link  threshold nobs logLik   AIC     niter      max.grad cond.H
##  logit flexible  2122 -1933.90 3901.79 2208(8836) 1.28e-03 6.6e+0
1
##
## Random effects:
##  Groups Name        Variance Std.Dev. Corr
##  ID     (Intercept) 1.0469   1.0232
##         learnTimePM 0.4727   0.6876   -0.054
##  item   (Intercept) 0.8958   0.9464
##         learnTimePM 0.3932   0.6270   -0.384
## Number of groups:  ID 30,  item 24
##
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## group1               0.48213    0.19934   2.419   0.0156 *
## learnTime1           0.13522    0.10081   1.341   0.1798
## time2-1              0.15395    0.10889   1.414   0.1574
## time3-2              0.11283    0.11027   1.023   0.3062
## group1:learnTime1   -0.10342    0.07774  -1.330   0.1834
## learnTime1:time2-1   0.15601    0.10877   1.434   0.1515
## learnTime1:time3-2   0.03505    0.11022   0.318   0.7505
## group1:time2-1      -0.07019    0.10885  -0.645   0.5191
## group1:time3-2       0.03756    0.11030   0.341   0.7335
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Threshold coefficients:
##     Estimate Std. Error z value
## 0|1  -1.3518     0.2706  -4.995
## 1|2   0.5003     0.2691   1.859
```

Three-way interaction pruned: LR.stat = 0.77, p = .679

D3.AM-PM Experiment – Definitions analysis (follow-up)

```
## Cumulative Link Mixed Model fitted with the Laplace approximation
##
## formula: acc ~ group * learnTime * time + (1 + time | ID) + (1 +
time |
##     item)
## data:    def
##
##  link  threshold nobs logLik   AIC     niter      max.grad cond.H
##  logit flexible  1422 -1306.73 2643.47 1747(6992) 5.18e-04 6.2e+0
1
##
## Random effects:
##  Groups Name        Variance Std.Dev. Corr
##  ID     (Intercept) 0.98946  0.9947
##         time4       0.81799  0.9044   0.063
##  item   (Intercept) 0.64052  0.8003
##         time4       0.08921  0.2987   1.000
## Number of groups:  ID 30,  item 24
##
## Coefficients:
##                           Estimate Std. Error z value Pr(>|z|)
## group1                    0.448442   0.212081   2.114  0.03447 *
## learnTime1                0.035041   0.055363   0.633  0.52678
## time1                    -0.297951   0.104718  -2.845  0.00444 **
## group1:learnTime1         0.032951   0.055450   0.594  0.55235
## group1:time1             -0.044789   0.100191  -0.447  0.65485
## learnTime1:time1          0.007604   0.055318   0.137  0.89067
## group1:learnTime1:time1   0.091544   0.055398   1.652  0.09844 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Threshold coefficients:
##     Estimate Std. Error z value
## 0|1  -0.7992     0.2898  -2.757
## 1|2   0.8617     0.2900   2.972
```

D4.AM-PM Experiment – Picture naming analysis (24-hour) – accuracy

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula:
## acc ~ group + task + learnTime + time + group:task + group:learnT
ime +
##     group:time + task:learnTime + task:time + learnTime:time +
##     (1 + group + learnTime + task | item) + (1 + learnTime |      I
D)
##    Data: picName
## Control:
## glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 20000)
)
##
##      AIC      BIC   logLik deviance df.resid
##   3756.9   3934.7  -1850.5   3700.9     4192
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -5.6999 -0.4770 -0.1998  0.4294 13.9751
##
## Random effects:
##  Groups Name        Variance Std.Dev. Corr
##  ID     (Intercept) 1.5802   1.2571
##         learnTime1  0.5428   0.7368   -0.22
##  item   (Intercept) 1.7420   1.3198
##         group1      0.2145   0.4631   -0.47
##         learnTime1  0.1836   0.4285   -0.46  0.57
##         task1       0.0539   0.2322   -0.40  0.04 -0.26
## Number of obs: 4220, groups:  ID, 30; item, 24
##
## Fixed effects:
##                      Estimate Std. Error z value Pr(>|z|)
## (Intercept)        -1.3895580  0.3585823  -3.875 0.000107 ***
## group1              0.5583634  0.2541462   2.197 0.028019 *
## task1               0.2058816  0.0654714   3.145 0.001663 **
## learnTime1          0.1827214  0.1695093   1.078 0.281059
## time2-1             0.4774752  0.1097318   4.351 1.35e-05 ***
## time3-2             0.4873769  0.1035817   4.705 2.54e-06 ***
## group1:task1       -0.0075375  0.0438003  -0.172 0.863369
## group1:learnTime1  -0.0761363  0.1436193  -0.530 0.596024
## group1:time2-1     -0.1665971  0.1088410  -1.531 0.125857
## group1:time3-2      0.0009450  0.1024681   0.009 0.992642
## task1:learnTime1   -0.0197313  0.0432602  -0.456 0.648313
## task1:time2-1      -0.0488697  0.1067898  -0.458 0.647222
## task1:time3-2      -0.0003871  0.1013181  -0.004 0.996951
## learnTime1:time2-1  0.6232117  0.1077284   5.785 7.25e-09 ***
## learnTime1:time3-2 -0.0529997  0.1026194  -0.516 0.605527
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Three- and four-way interactions pruned: χ^2= 3.68, p = .931

D5.AM-PM Experiment – Picture naming analysis (follow-up) - accuracy

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula:
## acc ~ group + task + learnTime + time + group:task + group:learnT
ime +
##     group:time + task:learnTime + task:time + learnTime:time +
##     (1 + learnTime + time + group | item) + (1 + learnTime +
##     time | ID)
##    Data: picName
## Control: glmerControl(optimizer = "bobyqa")
##
##      AIC      BIC   logLik deviance df.resid
##   2467.3   2627.8  -1206.6   2413.3     2793
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.4631 -0.4431 -0.2298  0.2480  6.4474
##
## Random effects:
##  Groups Name        Variance Std.Dev. Corr
##  ID     (Intercept) 1.4549   1.2062
##         learnTime1  0.3485   0.5903   -0.26
##         time1       0.1159   0.3405    0.29  0.50
##  item   (Intercept) 1.5611   1.2494
##         learnTime1  0.2530   0.5030   -0.06
##         time1       0.1603   0.4004    0.09 -0.03
##         group1      0.1376   0.3710   -0.15  0.13  0.69
## Number of obs: 2820, groups:  ID, 30; item, 24
##
## Fixed effects:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)        -1.775791   0.345561  -5.139 2.76e-07 ***
## group1              0.560474   0.243401   2.303  0.02130 *
## task1               0.173446   0.054013   3.211  0.00132 **
## learnTime1         -0.020582   0.165315  -0.125  0.90092
## time1              -0.004729   0.122714  -0.039  0.96926
## group1:task1       -0.011913   0.053990  -0.221  0.82536
## group1:learnTime1   0.013851   0.124814   0.111  0.91164
## group1:time1       -0.070622   0.085534  -0.826  0.40900
## task1:learnTime1   -0.014161   0.053346  -0.265  0.79066
## task1:time1        -0.026744   0.053399  -0.501  0.61648
## learnTime1:time1    0.239077   0.056539   4.229 2.35e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Three- and four-way interactions pruned: χ^2= 1.88, *p* = .865

D6.AM-PM Experiment – Picture naming analysis (24-hour) – response time

```
## Linear mixed model fit by maximum likelihood . t-tests use
##   Satterthwaite's method [lmerModLmerTest]
## Formula:
## RTtrans ~ group + task + learnTime + time + group:task + group:learnTime
+
##     group:time + task:learnTime + task:time + learnTime:time +
##     (1 + learnTime | item) + (1 | ID)
##    Data: picName
##
##      AIC      BIC   logLik deviance df.resid
##   5720.3   5823.3  -2840.2   5680.3     1251
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.2493 -0.6360 -0.0848  0.6030  3.5411
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  ID       (Intercept) 1.00755  1.0038
##  item     (Intercept) 1.12507  1.0607
##           learnTime1  0.09805  0.3131   0.10
##  Residual             4.58648  2.1416
## Number of obs: 1271, groups:  ID, 30; item, 24
##
## Fixed effects:
##                      Estimate Std. Error         df  t value Pr(>|t|)
## (Intercept)        1406.74565    0.29786   45.14566 4722.839  < 2e-16 ***
## group1               -0.29722    0.19955   27.87936   -1.489  0.14759
## task1                -0.80410    0.06357 1206.38709  -12.649  < 2e-16 ***
## learnTime1           -0.20817    0.09890   22.41161   -2.105  0.04672 *
## time2-1              -0.80113    0.16625 1216.17612   -4.819 1.63e-06 ***
## time3-2              -0.74303    0.14603 1206.49603   -5.088 4.19e-07 ***
## group1:task1         -0.02107    0.06215 1201.99245   -0.339  0.73468
## group1:learnTime1     0.01807    0.06787 1241.51042    0.266  0.79011
## group1:time2-1        0.24606    0.16440 1213.90966    1.497  0.13472
## group1:time3-2       -0.18226    0.14483 1208.61176   -1.258  0.20848
## task1:learnTime1      0.06377    0.06146 1216.59836    1.038  0.29963
## task1:time2-1        -0.20662    0.15862 1201.73562   -1.303  0.19296
## task1:time3-2         0.23426    0.14121 1197.92916    1.659  0.09739 .
## learnTime1:time2-1   -0.65787    0.16039 1211.31920   -4.102 4.38e-05 ***
## learnTime1:time3-2    0.45676    0.14300 1205.29106    3.194  0.00144 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Three- and four-way interactions pruned: χ^2= 4.61, p = .867

D7.AM-PM Experiment – Picture naming analysis (follow-up) – response time

```
## Linear mixed model fit by maximum likelihood . t-tests use
##   Satterthwaite's method [lmerModLmerTest]
## Formula:
## RTtrans ~ group + task + learnTime + time + group:task + group:learnTime
+
##     group:time + task:learnTime + task:time + learnTime:time +
##     (1 + time | item) + (1 | ID)
##    Data: picName
##
##      AIC      BIC   logLik deviance df.resid
##   3590.8   3663.2  -1779.4   3558.8      666
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.80400 -0.67453 -0.02675  0.58202  3.03941
##
## Random effects:
##  Groups   Name        Variance Std.Dev. Corr
##  ID       (Intercept) 1.1772   1.0850
##  item     (Intercept) 1.9547   1.3981
##           time1       0.4014   0.6335   0.38
##  Residual             9.4328   3.0713
## Number of obs: 682, groups:  ID, 29; item, 24
##
## Fixed effects:
##                      Estimate Std. Error         df  t value Pr(>|t|)
## (Intercept)         1.489e+03  3.887e-01  3.733e+01 3831.971  < 2e-16 ***
## group1             -1.745e-01  2.485e-01  2.759e+01   -0.702  0.48858
## task1              -1.089e+00  1.239e-01  6.256e+02   -8.790  < 2e-16 ***
## learnTime1          1.258e-01  1.343e-01  6.619e+02    0.937  0.34931
## time1              -3.718e-01  1.932e-01  1.841e+01   -1.924  0.06991 .
## group1:task1       -8.711e-03  1.234e-01  6.191e+02   -0.071  0.94376
## group1:learnTime1 -3.969e-01  1.319e-01  6.564e+02   -3.009  0.00272 **
## group1:time1        3.105e-01  1.279e-01  6.526e+02    2.429  0.01543 *
## task1:learnTime1    8.484e-02  1.205e-01  6.247e+02    0.704  0.48145
## task1:time1        -9.269e-02  1.206e-01  6.245e+02   -0.769  0.44241
## learnTime1:time1   -1.167e-01  1.231e-01  6.456e+02   -0.947  0.34379
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Three- and four-way interactions pruned: χ^2= 2.39, *p* = .793

D8.AM-PM Experiment – Stem completion analysis (24-hour)

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula:
## acc ~ group * learnTime * time + (1 + group + learnTime | item) +
##     (1 + learnTime | ID)
##    Data: stemComp
## Control: glmerControl(optimizer = "bobyqa")
##
##      AIC      BIC   logLik deviance df.resid
##   2174.4   2293.2  -1066.2   2132.4     2088
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.4867 -0.5379 -0.2579  0.5902 11.7764
##
## Random effects:
##  Groups Name        Variance Std.Dev. Corr
##  ID     (Intercept) 1.1459   1.0705
##         learnTime1  0.3488   0.5906   0.13
##  item   (Intercept) 1.3155   1.1469
##         group1      0.1345   0.3667   -0.18
##         learnTime1  0.1038   0.3221   -0.08  0.33
## Number of obs: 2109, groups:  ID, 30; item, 24
##
## Fixed effects:
##                           Estimate Std. Error z value Pr(>|z|)
## (Intercept)               -0.98010    0.31214  -3.140  0.00169 **
## group1                     0.47466    0.21918   2.166  0.03034 *
## learnTime1                 0.13947    0.14177   0.984  0.32523
## time2-1                    0.29359    0.14172   2.072  0.03830 *
## time3-2                    0.55538    0.13685   4.058 4.94e-05 ***
## group1:learnTime1         -0.02190    0.12394  -0.177  0.85973
## group1:time2-1            -0.25351    0.14108  -1.797  0.07235 .
## group1:time3-2             0.08257    0.13665   0.604  0.54571
## learnTime1:time2-1         0.72504    0.14182   5.112 3.18e-07 ***
## learnTime1:time3-2         0.01307    0.13681   0.096  0.92390
## group1:learnTime1:time2-1 -0.22960    0.14153  -1.622  0.10475
## group1:learnTime1:time3-2 -0.16731    0.13632  -1.227  0.21968
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

D9.AM-PM Experiment – Stem completion analysis (follow-up)

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: acc ~ group * learnTime + group * time + learnTime * time + (1 +
##     time + learnTime | item) + (1 + learnTime | ID)
##    Data: stemComp
## Control: glmerControl(optimizer = "bobyqa")
##
##      AIC      BIC   logLik deviance df.resid
##   1517.8   1601.8   -742.9   1485.8     1394
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.5955 -0.5535 -0.3185  0.6376  6.4766
##
## Random effects:
##  Groups Name        Variance Std.Dev. Corr
##  ID     (Intercept) 0.6519   0.8074
##         learnTime1  0.3084   0.5554   0.15
##  item   (Intercept) 1.1091   1.0531
##         time1       0.1519   0.3897    0.67
##         learnTime1  0.1401   0.3742   -0.01 -0.24
## Number of obs: 1410, groups:  ID, 30; item, 24
##
## Fixed effects:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)        -1.10466    0.27201  -4.061 4.89e-05 ***
## group1              0.38899    0.16355   2.378  0.01739 *
## learnTime1         -0.13808    0.14805  -0.933  0.35099
## time1               0.06899    0.10684   0.646  0.51844
## group1:learnTime1   0.10581    0.12362   0.856  0.39204
## group1:time1       -0.10429    0.06807  -1.532  0.12550
## learnTime1:time1    0.18995    0.06838   2.778  0.00547 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

Three-way interaction pruned: χ^2= 1.13, p = .288

D10. AM-PM Experiment – Object pair analysis (24-hour)

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: acc ~ group * learnTime * time + (1 | item) + (1 + learnTime |
##     ID)
##    Data: op
## Control: glmerControl(optimizer = "bobyqa")
##
##      AIC      BIC   logLik deviance df.resid
##   1970.0   2057.8   -969.0   1938.0     1774
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.5754 -0.6126 -0.3110  0.6669  4.3365
##
## Random effects:
##  Groups Name        Variance Std.Dev. Corr
##  ID     (Intercept) 0.28743  0.5361
##         learnTime1  0.09316  0.3052   0.14
##  item   (Intercept) 1.38829  1.1783
## Number of obs: 1790, groups:  ID, 30; item, 20
##
## Fixed effects:
##                            Estimate Std. Error z value Pr(>|z|)
## (Intercept)               -0.316586   0.287288  -1.102  0.27047
## group1                     0.304434   0.113953   2.672  0.00755 **
## learnTime1                -0.091090   0.080734  -1.128  0.25920
## time2-1                   -1.611611   0.144376 -11.163  < 2e-16 ***
## time3-2                    0.000169   0.143456   0.001  0.99906
## group1:learnTime1          0.108609   0.081027   1.340  0.18011
## group1:time2-1             0.012521   0.140383   0.089  0.92893
## group1:time3-2            -0.004457   0.143466  -0.031  0.97522
## learnTime1:time2-1         0.305019   0.141449   2.156  0.03105 *
## learnTime1:time3-2        -0.145154   0.143500  -1.012  0.31176
## group1:learnTime1:time2-1 -0.229515   0.140553  -1.633  0.10248
## group1:learnTime1:time3-2  0.220934   0.143561   1.539  0.12382
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

D11. AM-PM Experiment – Object pair analysis (follow-up)

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: acc ~ group * learnTime * time + (1 + time | item) + (1 | ID)
##    Data: op
## Control: glmerControl(optimizer = "bobyqa")
##
##      AIC      BIC   logLik deviance df.resid
##   1191.5   1252.6   -583.8   1167.5     1188
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.5838 -0.4828 -0.2464  0.5888  4.8071
##
## Random effects:
##  Groups Name        Variance Std.Dev. Corr
##  ID     (Intercept) 0.1435   0.3788
##  item   (Intercept) 0.4932   0.7023
##         time1       0.2894   0.5379   -0.02
## Number of obs: 1200, groups:  ID, 30; item, 20
##
## Fixed effects:
##                        Estimate Std. Error z value Pr(>|z|)
## (Intercept)            -0.77801    0.19454  -3.999 6.35e-05 ***
## group1                  0.23919    0.10502   2.278 0.022753 *
## learnTime1             -0.08009    0.07894  -1.015 0.310304
## time1                  -1.43718    0.15216  -9.445  < 2e-16 ***
## group1:learnTime1      -0.04099    0.07973  -0.514 0.607189
## group1:time1           -0.03215    0.07911  -0.406 0.684500
## learnTime1:time1        0.10686    0.07921   1.349 0.177346
## group1:learnTime1:time1 -0.26533    0.07984  -3.323 0.000889 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

D12. AM-PM Experiment – Exploratory vocabulary analysis

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula: acc ~ vocab * learnTime * time + (1 + learnTime | item)
+ (1 |
##     ID)
##    Data: stemComp
##
##      AIC      BIC   logLik deviance df.resid
##   2491.1   2583.2  -1229.6   2459.1     2312
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.5969 -0.6024 -0.3012  0.6765  8.1734
##
## Random effects:
##  Groups Name        Variance Std.Dev. Corr
##  ID     (Intercept) 0.59928  0.7741
##  item   (Intercept) 1.01566  1.0078
##         learnTime1  0.07911  0.2813   -0.17
## Number of obs: 2328, groups:  ID, 33; item, 24
##
## Fixed effects:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)             -0.852525   0.252458  -3.377 0.000733 **
*
## vocab                    0.664468   0.148321   4.480 7.47e-06 **
*
## learnTime1               0.171437   0.078977   2.171 0.029951 *
## time2-1                  0.253908   0.130079   1.952 0.050944 .
## time3-2                  0.466321   0.124601   3.742 0.000182 **
*
## vocab:learnTime1         0.102883   0.057817   1.779 0.075164 .
## vocab:time2-1           -0.165967   0.143298  -1.158 0.246783
## vocab:time3-2            0.077570   0.131632   0.589 0.555663
## learnTime1:time2-1       0.628461   0.130154   4.829 1.37e-06 **
*
## learnTime1:time3-2      -0.007927   0.124503  -0.064 0.949234
## vocab:learnTime1:time2-1  0.032224   0.142843   0.226 0.821521
## vocab:learnTime1:time3-2 -0.339437   0.131904  -2.573 0.010071 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

D13. AM-PM Experiment – Training data

	AM-encoding	PM-encoding
	M (*SD*)	*M* (*SD*)
Poor Comprehenders	.89 (.31)	.89 (.32)
Good Comprehenders	.94 (.25)	.94 (.24)

```
## Generalized linear mixed model fit by maximum likelihood (Laplace
##   Approximation) [glmerMod]
##  Family: binomial  ( logit )
## Formula:
## acc ~ group + learnTime + modality + difficulty + (1 + learnTime |
##     ID) + (1 | item)
##    Data: training
##
##      AIC      BIC   logLik deviance df.resid
##   1593.2   1646.9   -787.6   1575.2     2871
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -6.8342  0.1647  0.2323  0.3150  0.7782
##
## Random effects:
##  Groups Name        Variance Std.Dev. Corr
##  ID     (Intercept) 0.6608   0.8129
##         learnTime1  0.1456   0.3816   -0.37
##  item   (Intercept) 0.2306   0.4802
## Number of obs: 2880, groups:  ID, 30; item, 24
##
## Fixed effects:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.78920    0.20179  13.822   <2e-16 ***
## group1       0.30101    0.16338   1.842   0.0654 .
## learnTime1  -0.05823    0.11470  -0.508   0.6117
## modality1    0.00479    0.06718   0.071   0.9432
## difficulty1  0.01435    0.06718   0.214   0.8309
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) group1 lrnTm1 mdlty1
## group1       0.032
## learnTime1  -0.208  0.038
## modality1    0.001  0.000  0.000
## difficulty1  0.004  0.000  0.000  0.000
```

References

Antony, J. W., Ferreira, C. S., Norman, K. A., & Wimber, M. (2017). Retrieval as a fast route to memory consolidation. *Trends in Cognitive Sciences, 21*, 573-576.

Anwyl-Irvine, A., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. (2018). Gorillas in our Midst: An online behavioral experiment builder. *bioRxiv*, 438242.

Ashworth, A., Hill, C. M., Karmiloff-Smith, A., & Dimitriou, D. (2014). Sleep enhances memory consolidation in children. *Journal of Sleep Research, 23*, 304-310.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390-412.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database (release 2). *Distributed by the Linguistic Data Consortium, University of Pennsylvania*.

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

Baldo, J. V., Bunge, S. A., Wilson, S. M., & Dronkers, N. (2010). Is relational reasoning dependent on language? A voxel-based lesion symptom mapping study. *Brain and Language, 113*, 59-64.

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English lexicon project. *Behavior Research Methods, 39*, 445-459.

Barnes, M. A., Stuebing, K. K., Fletcher, J. M., Barth, A. E., & Francis, D. J. (2016). Cognitive difficulties in struggling comprehenders and their relation to reading comprehension: A comparison of group selection and regression-based models. *Journal of Research on Educational Effectiveness, 9*, 153-172.

Barnhoorn, J. S., Haasnoot, E., Bocanegra, B. R., & van Steenbergen, H. (2015). QRTEngine: An easy solution for running online reaction time experiments using Qualtrics. *Behavior Research Methods, 47*, 918-929.

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

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

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. *R package version, 1*.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015b). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*, 1-48.

Batterink, L. J., Oudiette, D., Reber, P. J., & Paller, K. A. (2014). Sleep facilitates learning a new linguistic rule. *Neuropsychologia, 65*, 169-179.

Batterink, L. J., & Paller, K. A. (2017). Sleep-based memory processing facilitates grammatical generalization: evidence from targeted memory reactivation. *Brain and Language, 167*, 83-93.

Beck, I. L., Perfetti, C. A., & McKeown, M. G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology, 74*, 506.

Biemiller, A. (2003). Vocabulary: Needed if more children are to read well. *Reading Psychology, 24*, 323-335.

Biemiller, A. (2006). Vocabulary development and instruction: A prerequisite for school learning. *Handbook of Early Literacy Research, 2*, 41-51.

Biemiller, A., & Slonim, N. (2001). Estimating root word vocabulary growth in normative and advantaged populations: Evidence for a common sequence of vocabulary acquisition. *Journal of Educational Psychology, 93*, 498-520.

Blakemore, S.-J., & Bunge, S. A. (2012). At the nexus of neuroscience and education. *Developmental Cognitive Neuroscience, 2*, S1-S5.

Borovsky, A., Ellis, E. M., Evans, J. L., & Elman, J. L. (2015). Lexical leverage: category knowledge boosts real-time novel word recognition in 2-year-olds. *Developmental Science*.

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

Braginsky, M. (2018). ggpirate. Retrieved from https://github.com/mikabr/ggpirate

Braze, D., Tabor, W., Shankweiler, D. P., & Mencl, W. E. (2007). Speaking up for vocabulary: Reading skill differences in young adults. *Journal of Learning Disabilities, 40*, 226-243.

Brod, G., & Shing, Y. L. (in press). A Boon and a Bane: Comparing the Effects of Prior Knowledge on Memory Across the Lifespan. *Developmental Psychology*.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods, 46*, 904-911.

Buchanan, L., Westbury, C., & Burgess, C. (2001). Characterizing semantic space: Neighborhood effects in word recognition. *Psychonomic Bulletin & Review, 8*, 531-544.

Buchmann, A., Ringli, M., Kurth, S., Schaerer, M., Geiger, A., Jenni, O. G., & Huber, R. (2011). EEG sleep slow-wave activity as a mirror of cortical maturation. *Cerebral Cortex, 21*, 607-615.

Burge, B., Styles, B., Brzyska, B., Cooper, L., Shamsan, Y., Saltini, F., & Twist, L. (2010). New Group Reading Test (NGRT): London: GL Assessment.

Burgoyne, K., Kelly, J., Whiteley, H., & Spooner, A. (2009). The comprehension skills of children learning English as an additional language. *British Journal of Educational Psychology, 79*, 735-747.

Cain, K., & Oakhill, J. (2011). Matthew effects in young readers reading comprehension and reading experience aid vocabulary development. *Journal of Learning Disabilities, 44*, 431-443.

Cain, K., Oakhill, J., & Lemmon, K. (2004). Individual differences in the inference of word meanings from context: The influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology, 96*, 671.

Cain, K., & Oakhill, J. V. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and writing, 11*, 489-503.

Catts, H. W., Adlof, S. M., & Weismer, S. E. (2006). Language deficits in poor comprehenders: A case for the simple view of reading. *Journal of Speech, Language, and Hearing Research, 49*, 278-293.

Catts, H. W., Compton, D., Tomblin, J. B., & Bridges, M. S. (2012). Prevalence and nature of late-emerging poor readers. *Journal of Educational Psychology*, *104*, 166.

Christensen, R. H. B. (2015). ordinal-Regression Models for Ordinal Data. R package version 2015.6-28.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12*, 335-359.

Clarke, P. J., Snowling, M. J., Truelove, E., & Hulme, C. (2010). Ameliorating children's reading-comprehension difficulties: A randomized controlled trial. *Psychological Science, 21*, 1106-1116.

Clay, F., Bowers, J. S., Davis, C. J., & Hanley, D. A. (2007). Teaching adults new words: the role of practice and consolidation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 970.

Colenbrander, D., Kohnen, S., Smith-Lock, K., & Nickels, L. (2016). Individual differences in the vocabulary skills of children with poor reading comprehension. *Learning and Individual Differences*, *50*, 210-220.

Colenbrander, D., Nickels, L., & Kohnen, S. (2017). Similar but different: differences in comprehension diagnosis on the Neale Analysis of Reading Ability and the York Assessment of Reading for Comprehension. *Journal of Research in Reading*, *40*, 403-419.

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology, 33*, 497-505.

Coutanche, M. N., & Thompson-Schill, S. L. (2014). Fast mapping rapidly integrates information into existing memory networks. *Journal of Experimental Psychology: General*, *143*, 2296.

Cutting, L. E., Clements-Stephens, A., Pugh, K. R., Burns, S., Cao, A., Pekar, J. J., . . . Rimrodt, S. L. (2013). Not All Reading Disabilities Are Dyslexia: Distinct Neurobiology of Specific Comprehension Deficits. *Brain Connectivity, 3*, 199-211.

Davies, R. A., Arnell, R., Birchenough, J. M., Grimmond, D., & Houlson, S. (2017). Reading through the life span: Individual differences in psycholinguistic effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*, 1298.

Davis, C., & Andrews, S. (1996). *The role of computational modeling in studies of visual word recognition.* Paper presented at the Symposium on Computer Models of Cognition: Possibilities and Pitfalls. 31st.

Davis, M. H., Di Betta, A. M., Macdonald, M. J., & Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. *Journal of Cognitive Neuroscience, 21*, 803-820.

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

DeMaster, D. M., & Ghetti, S. (2013). Developmental differences in hippocampal and cortical contributions to episodic retrieval. *Cortex, 49,* 1482-1493.

Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience, 11*, 114-126.

Diekelmann, S., Wilhelm, I., & Born, J. (2009). The whats and whens of sleep-dependent memory consolidation. *Sleep Medicine Reviews, 13*, 309-321.

Dimitriou, D., Karmiloff-Smith, A., Ashworth, A., & Hill, C. M. (2013). Impaired sleep-related learning in children with Williams syndrome. *Pediatrics Research International Journal, 2013*.

Drosopoulos, S., Schulze, C., Fischer, S., & Born, J. (2007). Sleep's function in the spontaneous recovery and consolidation of memories. *Journal of Experimental Psychology: General*, *136*, 169.

Dumay, N. (2016). Sleep not just protects memories against forgetting, it also makes them more accessible. *Cortex*, *74*, 289-296.

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

Dumay, N., Gaskell, M. G., Feng, X., Forbus, K., Gentner, D., & Regier, T. (2004). *A day in the life of a spoken word.* Paper presented at the Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society.

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody picture vocabulary test*: Pearson Assessments.

Dunn, L. M., Dunn, D. M., Styles, B., & Sewell, J. (2009). *The British Picture Vocabulary Scale III*. London: GL Assessment.

Ellenbogen, J. M., Payne, J. D., & Stickgold, R. (2006). The role of sleep in declarative memory consolidation: passive, permissive, active or none? *Current Opinion in Neurobiology, 16*, 716-722.

Elliot, C., Smith, P., & McCulloch, K. (1997). British Abilities Scale second edition (BAS II) technical manual. *Windsor, Berkshire: NFER-Nelson*.

Feinberg, I., & Campbell, I. G. (2010). Sleep EEG changes during adolescence: an index of a fundamental brain reorganization. *Brain and Cognition, 72*, 56-65.

Fenn, K. M., & Hambrick, D. Z. (2013). What drives sleep-dependent memory consolidation: Greater gain or less loss? *Psychonomic Bulletin & Review, 20*, 501-506.

Fernandes, T., Kolinsky, R., & Ventura, P. (2009). The metamorphosis of the statistical segmentation output: Lexicalization during artificial language learning. *Cognition, 112*, 349-366.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers, 35*, 116-124.

Freyd, P., & Baron, J. (1982). Individual differences in acquisition of derivational morphology. *Journal of Verbal Learning and Verbal Behavior, 21*, 282-295.

Friedrich, M., Wilhelm, I., Born, J., & Friederici, A. D. (2015). Generalization of word meanings during infant sleep. *Nature Communications, 6*.

Gais, S., Lucas, B., & Born, J. (2006). Sleep after learning aids memory recall. *Learning & Memory, 13*, 259-262.

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

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

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

German, D. J., & Newman, R. S. (2004). The impact of lexical factors on children's word-finding errors. *Journal of Speech, Language, and Hearing Research, 47*, 624-636.

Ghetti, S., & Bunge, S. A. (2012). Neural changes underlying the development of episodic memory during middle childhood. *Developmental Cognitive Neuroscience, 2*, 381-395.

Ghetti, S., DeMaster, D. M., Yonelinas, A. P., & Bunge, S. A. (2010). Developmental differences in medial temporal lobe function during memory encoding. *The Journal of Neuroscience, 30*, 9548-9556.

Giedd, J. N., Vaituzis, A. C., Hamburger, S. D., Lange, N., Rajapakse, J. C., Kaysen, D., . . . Rapoport, J. L. (1996). Quantitative MRI of the temporal lobe, amygdala, and hippocampus in normal human development: ages 4–18 years. *Journal of Comparative Neurology, 366*, 223-230.

Gilmore, J. H., Shi, F., Woolson, S. L., Knickmeyer, R. C., Short, S. J., Lin, W., . . . Shen, D. (2012). Longitudinal development of cortical and subcortical gray matter from birth to 2 years. *Cerebral Cortex, 22*, 2478-2485.

Gogtay, N., Nugent, T. F., Herman, D. H., Ordonez, A., Greenstein, D., Hayashi, K. M., . . . Rapoport, J. L. (2006). Dynamic mapping of normal human hippocampal development. *Hippocampus, 16*, 664-672.

Gómez, R. L., Bootzin, R. R., & Nadel, L. (2006). Naps promote abstraction in language-learning infants. *Psychological Science, 17*, 670-674.

Gómez, R. L., & Edgin, J. O. (2015). Sleep as a Window Into Early Neural Development: Shifts in Sleep-Dependent Learning Effects Across Early Childhood. *Child Development Perspectives, 9*, 183-189.

Groch, S., McMakin, D., Guggenbühl, P., Rasch, B., Huber, R., & Wilhelm, I. (2016). Memory cueing during sleep modifies the interpretation of ambiguous scenes in adolescents and adults. *Developmental Cognitive Neuroscience, 17*, 10-18.

Groen, M. A., Veenendaal, N. J., & Verhoeven, L. (2018). The role of prosody in reading comprehension: evidence from poor comprehenders. *Journal of Research in Reading*.

Grondin, R., Lupker, S. J., & McRae, K. (2009). Shared features dominate semantic richness effects for concrete concepts. *Journal of Memory and Language, 60*, 1-19.

Hardt, O., Nader, K., & Nadel, L. (2013). Decay happens: the role of active forgetting in memory. *Trends in Cognitive Sciences, 17*, 111-120.

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

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

Henderson, L., Powell, A., Gaskell, M. G., & Norbury, C. (2014). Learning and consolidation of new spoken words in autism spectrum disorder. *Developmental Science, 17*, 858-871.

Henderson, L., Snowling, M., & Clarke, P. (2013a). Accessing, integrating, and inhibiting word meaning in poor comprehenders. *Scientific Studies of Reading, 17*, 177-198.

Henderson, L. M., Devine, K., Weighall, A., & Gaskell, G. (2015). When the daffodat flew to the intergalactic zoo: Off-line consolidation is critical for word learning from stories. *Developmental Psychology, 51*, 406.

Henderson, L. M., & James, E. (2018). Consolidating new words from repetitive versus multiple stories: Prior knowledge matters. *Journal of Experimental Child Psychology, 166*, 465-484.

Henderson, L. M., Weighall, A., Brown, H., & Gaskell, G. (2013b). Online lexical competition during spoken word recognition and word learning in children and adults. *Child development, 84*, 1668-1685.

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

Henderson, L. M., Weighall, A. R., Brown, H., & Gaskell, M. G. (2012). Consolidation of vocabulary is associated with sleep in children. *Developmental Science, 15*, 674-687.

Hennies, N., Lambon Ralph, M. A., Kempkes, M., Cousins, J. N., & Lewis, P. A. (2016). Sleep Spindle Density Predicts the Effect of Prior Knowledge on Memory Consolidation. *The Journal of Neuroscience, 36*, 3799-3810.

Hills, T. T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. (2009). Longitudinal Analysis of Early Semantic Networks Preferential Attachment or Preferential Acquisition? *Psychological Science, 20*, 729-739.

Himmer, L., Müller, E., Gais, S., & Schönauer, M. (2017). Sleep-mediated memory consolidation depends on the level of integration at encoding. *Neurobiology of Learning and Memory, 137*, 101-106.

Hoover, J. R., Storkel, H. L., & Hogan, T. P. (2010). A cross-sectional comparison of the effects of phonotactic probability and neighborhood density on word learning by preschool children. *Journal of Memory and Language, 63*, 100-116.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and writing, 2*, 127-160.

Horst, J. S., & Hout, M. C. (2016). The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behavior Research Methods, 48,* 1393-1409.

Horváth, K., Liu, S., & Plunkett, K. (2015a). A Daytime Nap Facilitates Generalization of Word Meanings in Young Toddlers. *Sleep, 39*, 203-207.

Horváth, K., Myers, K., Foster, R., & Plunkett, K. (2015b). Napping facilitates word learning in early lexical development. *Journal of Sleep Research, 24,* 503-509.

Hulme, C., & Snowling, M. J. (2011). Children's reading comprehension difficulties: Nature, causes, and treatments. *Current Directions in Psychological Science, 20*, 139-142.

Hulme, R. C., Barsky, D., & Rodd, J. M. (2018). Incidental Learning and Long-Term Retention of New Word Meanings From Stories: The Effect of Number of Exposures. *Language Learning*.

Hupbach, A., Gomez, R. L., Bootzin, R. R., & Nadel, L. (2009). Nap-dependent learning in infants. *Developmental Science, 12,* 1007-1012.

James, E., Gaskell, M. G., & Henderson, L. M. (2018). Offline consolidation supersedes prior knowledge benefits in children's (but not adults') word learning. *Developmental Science,* e12776.

James, E., Gaskell, M. G., Weighall, A., & Henderson, L. (2017). Consolidation of vocabulary during sleep: The rich get richer? *Neuroscience & Biobehavioral Reviews, 77,* 1-13.

Jenni, O. G., & Carskadon, M. A. (2004). Spectral analysis of the sleep electroencephalogram during adolescence. *Sleep, 27*, 774-783.

Justice, L., Mashburn, A., & Petscher, Y. (2013). Very early language skills of fifth-grade poor comprehenders. *Journal of Research in Reading, 36*, 172-185.

Kaushanskaya, M., & Marian, V. (2009). The bilingual advantage in novel word learning. *Psychonomic Bulletin & Review, 16*, 705-710.

Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading, 12*, 281-300.

Keenan, J. M., Hua, A. N., Meenan, C. E., Pennington, B. F., Willcutt, E., & Olson, R. K. (2014). Issues in identifying poor comprehenders. *L'Année psychologique, 114*, 753-777.

Knudsen, E. I. (2004). Sensitive periods in the development of the brain and behavior. *Journal of Cognitive Neuroscience, 16*, 1412-1425.

Krishnan, S., Sellars, E., Wood, H., Bishop, D. V., & Watkins, K. E. (2018). The influence of evaluative right/wrong feedback on phonological and semantic processes in word learning. *Royal Society Open Science, 5*, 171496.

Kumaran, D., Hassabis, D., & McClelland, J. L. (2016). What Learning Systems do Intelligent Agents Need? Complementary Learning Systems Theory Updated. *Trends in Cognitive Sciences, 20*, 512-534.

Kumaran, D., & McClelland, J. L. (2012). Generalization through the recurrent interaction of episodic memories: a model of the hippocampal system. *Psychological Review, 119*, 573.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods, 44*, 978-990.

Kurth, S., Ringli, M., Geiger, A., LeBourgeois, M., Jenni, O. G., & Huber, R. (2010). Mapping of cortical activity in the first two decades of life: a high-density sleep electroencephalogram study. *The Journal of Neuroscience, 30*, 13211-13219.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software, 82*.

Landi, N., & Perfetti, C. A. (2007). An electrophysiological investigation of semantic and phonological processing in skilled and less-skilled comprehenders. *Brain and Language, 102*, 30-45.

Landi, N., & Ryherd, K. (2017). Understanding specific reading comprehension deficit: A review. *Language and Linguistics Compass, 11*, e12234.

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

Lee, J. K., Nordahl, C. W., Amaral, D. G., Lee, A., Solomon, M., & Ghetti, S. (2015). Assessing hippocampal development and language in early childhood: Evidence from a new application of the Automatic Segmentation Adapter Tool. *Human Brain Mapping, 36*, 4483-4496.

Lewis, P. A., & Durrant, S. J. (2011). Overlapping memory replay during sleep builds cognitive schemata. *Trends in Cognitive Sciences, 15*, 343-351.

Malow, B. A., Marzec, M. L., McGrew, S. G., Wang, L., Henderson, L. M., & Stone, W. L. (2006). Characterizing sleep in children with autism spectrum disorders: a multidimensional approach. *Sleep, 29*, 1563.

Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND: Cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PloS One*, *7*, e43230.

Marshall, L., & Born, J. (2007). The contribution of sleep to hippocampus-dependent memory consolidation. *Trends in Cognitive Sciences, 11*, 442-450.

Marshall, L., Helgadóttir, H., Mölle, M., & Born, J. (2006). Boosting slow oscillations during sleep potentiates memory. *Nature, 444*, 610-613.

Maski, K., Holbrook, H., Manoach, D., Hanson, E., Kapur, K., & Stickgold, R. (2015). Sleep dependent memory consolidation in children with autism spectrum disorder. *Sleep, 38*, 1955.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44*, 314-324.

Mattys, S. L., & Clark, J. H. (2002). Lexical activity in speech processing: Evidence from pause detection. *Journal of Memory and Language, 47*, 343-359.

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

McClelland, J. L. (2013). Incorporating rapid neocortical learning of new schema-consistent information into complementary learning systems theory. *Journal of Experimental Psychology: General, 142*, 1190.

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

McGaugh, J. L. (2000). Memory--a century of consolidation. *Science, 287*, 248-251.

McGregor, K. K., Licandro, U., Arenas, R., Eden, N., Stiles, D., Bean, A., & Walker, E. (2013). Why words are hard for adults with developmental language impairments. *Journal of Speech, Language, and Hearing Research, 56*, 1845-1856.

McKague, M., Pratt, C., & 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*, 231-262.

McMurray, B., Kapnoula, E., & Gaskell, M. (2016). Learning and integration of new word-forms: Consolidation, pruning and the emergence of automaticity. *Speech Perception and Spoken Word Recognition. London, UK: Taylor and Francis*.

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods, 37*, 547-559.

Meade, G., Midgley, K. J., Dijkstra, T., & Holcomb, P. J. (2018). Cross-language neighborhood effects in learners indicative of an integrated lexicon. *Journal of Cognitive Neuroscience, 30*, 70-85.

Metsala, J. L. (1997). An examination of word frequency and neighborhood density in the development of spoken-word recognition. *Memory & cognition, 25*, 47-56.

Mirković, J., & Gaskell, M. G. (2016). Does Sleep Improve Your Grammar? Preferential Consolidation of Arbitrary Components of New Linguistic Knowledge. *PloS One, 11*, e0152489.

Mirman, D., & Magnuson, J. S. (2008). Attractor dynamics and semantic neighborhood density: processing is slowed by near neighbors and speeded by distant neighbors. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 65.

Mitchell, C., & McMurray, B. (2009). On leveraged learning in lexical acquisition and its relationship to acceleration. *Cognitive Science, 33*, 1503-1523.

Mölle, M., Eschenko, O., Gais, S., Sara, S. J., & Born, J. (2009). The influence of learning on sleep slow oscillations and associated spindles and ripples in humans and rats. *European Journal of Neuroscience, 29*, 1071-1081.

Mölle, M., Marshall, L., Gais, S., & Born, J. (2004). Learning increases human electroencephalographic coherence during subsequent slow sleep oscillations. *Proceedings of the National Academy of Sciences of the United States of America, 101*, 13963-13968.

Moscovitch, M., Rosenbaum, R. S., Gilboa, A., Addis, D. R., Westmacott, R., Grady, C., . . . Winocur, G. (2005). Functional neuroanatomy of remote episodic, semantic and spatial memory: a unified account based on multiple trace theory. *Journal of Anatomy, 207*, 35-66.

Nadel, L., Winocur, G., Ryan, L., & Moscovitch, M. (2007). Systems consolidation and hippocampus: two views. *Debates in Neuroscience, 1*, 55-66.

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

Nation, K., Adams, J. W., Bowyer-Crane, C. A., & Snowling, M. J. (1999). Working memory deficits in poor comprehenders reflect underlying language impairments. *Journal of Experimental Child Psychology, 73*, 139-158.

Nation, K., Clarke, P., Marshall, C. M., & Durand, M. (2004). Hidden language impairments in children: Parallels between poor reading comprehension and specific language impairment? *Journal of Speech, Language, and Hearing Research, 47*, 199-211.

Nation, K., Clarke, P., & Snowling, M. J. (2002). General cognitive ability in children with reading comprehension difficulties. *British Journal of Educational Psychology, 72*, 549-560.

Nation, K., Cocksey, J., Taylor, J. S., & Bishop, D. V. (2010). A longitudinal investigation of early reading and language skills in children with poor reading comprehension. *Journal of Child Psychology and Psychiatry, 51*, 1031-1039.

Nation, K., & Snowling, M. (1997). Assessing reading difficulties: The validity and utility of current measures of reading skill. *British Journal of Educational Psychology, 67*, 359-370.

Nation, K., & Snowling, M. J. (1998). Semantic processing and the development of word-recognition skills: Evidence from children with reading comprehension difficulties. *Journal of Memory and Language, 39*, 85-101.

Nation, K., Snowling, M. J., & Clarke, P. (2007). Dissecting the relationship between language skills and learning to read: Semantic and phonological contributions to new vocabulary learning in children with poor reading comprehension. *Advances in Speech Language Pathology, 9*, 131-139.

Neale, M. D. (1997). *Neale Analysis of Reading Ability - Revised (NARA-II)*. Windsor, UK: NFER.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers, 36*, 402-407.

Ngo, H.-V. V., Martinetz, T., Born, J., & Mölle, M. (2013). Auditory closed-loop stimulation of the sleep slow oscillation enhances memory. *Neuron, 78*, 545-553.

Ngo, H.-V. V., Miedema, A., Faude, I., Martinetz, T., Mölle, M., & Born, J. (2015). Driving Sleep Slow Oscillations by Auditory Closed-Loop Stimulation—A Self-Limiting Process. *The Journal of Neuroscience, 35*, 6630-6638.

Nicolas, A., Petit, D., Rompre, S., & Montplaisir, J. (2001). Sleep spindle characteristics in healthy subjects of different age groups. *Clinical Neurophysiology, 112*, 521-527.

Nieuwenhuis, I. L., Folia, V., Forkstam, C., Jensen, O., & Petersson, K. M. (2013). Sleep promotes the extraction of grammatical rules. *PloS One, 8*, e65046.

Oakhill, J., Hartt, J., & Samols, D. (2005). Levels of comprehension monitoring and working memory in good and poor comprehenders. *Reading and writing, 18*, 657-686.

Ohayon, M., Carskadon, M., Guilleminault, C., & Vitiello, M. (2004). Meta-analysis of quantitative sleep parameters from childhood to old age in healthy individuals: developing normative sleep values across the human lifespan. *Sleep, 27*, 1255.

Østby, Y., Tamnes, C. K., Fjell, A. M., & Walhovd, K. B. (2011). Dissociating memory processes in the developing brain: the role of hippocampal volume and cortical thickness in recall after minutes versus days. *Cerebral Cortex*, bhr116.

Østby, Y., Tamnes, C. K., Fjell, A. M., Westlye, L. T., Due-Tønnessen, P., & Walhovd, K. B. (2009). Heterogeneity in subcortical brain development: a structural magnetic resonance imaging study of brain maturation from 8 to 30 years. *The Journal of Neuroscience, 29*, 11772-11782.

Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology, 98*, 554.

Penno, J. F., Wilkinson, I. A., & Moore, D. W. (2002). Vocabulary acquisition from teacher explanation and repeated listening to stories: Do they overcome the Matthew effect? *Journal of Educational Psychology, 94*, 23.

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*, 357-383.

Perry, L. K., Axelsson, E. L., & Horst, J. S. (2015). Learning What to Remember: Vocabulary Knowledge and Children's Memory for Object Names and Features. *Infant and Child Development*.

Pexman, P. M., Hargreaves, I. S., Siakaluk, P. D., Bodner, G. E., & Pope, J. (2008). There are many ways to be rich: Effects of three measures of semantic richness on visual word recognition. *Psychonomic Bulletin & Review, 15*, 161-167.

Pexman, P. M., Lupker, S. J., & Hino, Y. (2002). The impact of feedback semantics in visual word recognition: Number-of-features effects in lexical decision and naming tasks. *Psychonomic Bulletin & Review, 9*, 542-549.

Phillips, N. (2017). yarrr: A Companion to the e-Book "YaRrr!: The Pirate's Guide to R". R package version 0.1.5. *https://CRAN.R-project.org/package=yarrr*.

Prehn-Kristensen, A., Göder, R., Fischer, J., Wilhelm, I., Seeck-Hirschner, M., Aldenhoff, J., & Baving, L. (2011). Reduced sleep-associated consolidation of declarative memory in attention-deficit/hyperactivity disorder. *Sleep Medicine, 12*, 672-679.

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

Qualtrics, L. (2014). Qualtrics [Software]: Provo: Author.

R Core Team. (2015). R: A language and environment for statistical computing. *R Foundation for Statistical Computing, Vienna, Austria*.

Rasch, B., & Born, J. (2013). About sleep's role in memory. *Physiological Reviews, 93*, 681-766.

Rezlescu, C. (2015). *TESTABLE: The web-based platform for creating, running and sharing behavioural experiments.* Paper presented at the Association for Psychological Science Teaching Institute Conference, New York, 21--24 May.

Ricketts, J., Bishop, D. V., & Nation, K. (2008). Investigating orthographic and semantic aspects of word learning in poor comprehenders. *Journal of Research in Reading, 31*, 117-135.

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

Rodd, J. M., Cutrin, B. L., Kirsch, H., Millar, A., & Davis, M. H. (2013). Long-term priming of the meanings of ambiguous words. *Journal of Memory and Language, 68*, 180-198.

Roediger, H. L. I., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*, 249-255.

Rose, J. (2006). *Independent review of the teaching of early reading*. DfES Publications.

Rose, J. (2009). *Identifying and teaching children and young people with dyslexia and literacy difficulties*. DCSF Publications.

Ryherd, K., & Landi, N. (2019). Category Learning in Poor Comprehenders. *Scientific Studies of Reading*, 1-12.

Sadeh, A., Pergamin, L., & Bar-Haim, Y. (2006). Sleep in children with attention-deficit hyperactivity disorder: a meta-analysis of polysomnographic studies. *Sleep Medicine Reviews, 10*, 381-398.

Samuels, S. J., & Flor, R. F. (1997). The importance of automaticity for developing expertise in reading. *Reading & Writing Quarterly: Overcoming Learning Difficulties, 13*, 107-121.

Scarborough, H. S., Catts, H., & Kamhi, A. (2005). Developmental relationships between language and reading: Reconciling a beautiful hypothesis with some ugly facts. *The Connections between Language and Reading Disabilities*, 3-24.

Schlichting, M. L., Guarino, K. F., Schapiro, A. C., Turk-Browne, N. B., & Preston, A. R. (2016). Hippocampal Structure Predicts Statistical Learning and Associative Inference Abilities during Development. *Journal of Cognitive Neuroscience*, 1-15.

Schreiner, T., Göldi, M., & Rasch, B. (2015). Cueing vocabulary during sleep increases theta activity during later recognition testing. *Psychophysiology*, *52*, 1538-1543.

Schreiner, T., & Rasch, B. (2014). Boosting vocabulary learning by verbal cueing during sleep. *Cerebral Cortex*, bhu139.

Schreiner, T., & Rasch, B. (2015). Cueing vocabulary in awake subjects during the day has no effect on memory. *Somnologie, 19*, 133-140.

Schreiner, T., & Rasch, B. (2016). The beneficial role of memory reactivation for language learning during sleep: A review. *Brain and Language*.

Schreiner, T., & Rasch, B. (2018). To gain or not to gain–The complex role of sleep for memory. *Cortex, 101*, 282-287.

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery & Psychiatry, 20*, 11-21.

Sénéchal, M., Thomas, E., & Monker, J.-A. (1995). Individual differences in 4-year-old children's acquisition of vocabulary during storybook reading. *Journal of Educational Psychology*, *87*, 218.

Seress, L. (2001). Morphological changes of the human hippocampal formation from midgestation to early childhood. *Handbook of Developmental Cognitive Neuroscience*, 45-58.

Sirota, A., & Buzsáki, G. (2005). Interaction between neocortical and hippocampal networks via slow oscillations. *Thalamus & Related Systems*, *3*, 245-259.

Smith, F. R., Gaskell, M. G., Weighall, A. R., Warmington, M., Reid, A. M., & Henderson, L. M. (2018). Consolidation of vocabulary is associated with sleep in typically developing children, but not in children with dyslexia. *Developmental Science, 21*, e12639.

Smith, F. R. H., Gaskell, M. G., Weighall, A. R., Warmington, M., Reid, A. M., & Henderson, L. M. (2017). A different role for sleep in the consolidation of new vocabulary in children with and without dyslexia. *Developmental Science*.

Snowling, M. J., Stothard, S. E., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., . . . Hulme, C. (2009). *YARC York Assessment of Reading for Comprehension Passage Reading*: GL Publishers.

Sobel, H. S., Cepeda, N. J., & Kapler, I. V. (2011). Spacing effects in real-world classroom vocabulary learning. *Applied Cognitive Psychology, 25*, 763-767.

Sommer, T. (2016). The Emergence of Knowledge and How it Supports the Memory for Novel Related Information. *Cerebral Cortex*.

Spencer, M., Quinn, J. M., Wagner, R. K. J. L. D. R., & Practice. (2014). Specific reading comprehension disability: Major problem, myth, or misnomer? *Learning Disabilities Research & Practice, 29,* 3-9.

Spencer, S., Clegg, J., Stackhouse, J., & Rush, R. (2016). Contribution of spoken language and socio-economic background to adolescents' educational achievement at age 16 years. *International Journal of Language & Communication Disorders*.

Squire, L. R., & Alvarez, P. (1995). Retrograde amnesia and memory consolidation: a neurobiological perspective. *Current Opinion in Neurobiology, 5,* 169-177.

Squire, L. R., & Zola-Morgan, S. (1991). The Medial Temporal. *Science, 253,* 5026.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 360-407.

Stanovich, K. E. (1993). Does reading make you smarter? Literacy and the development of verbal intelligence. *Advances in Child Development and Behavior, 24*, 133-180.

Staresina, B. P., Bergmann, T. O., Bonnefond, M., van der Meij, R., Jensen, O., Deuker, L., . . . Fell, J. (2015). Hierarchical nesting of slow oscillations, spindles and ripples in the human hippocampus during sleep. *Nature Neuroscience*.

Steyvers, M., & Tenenbaum, J. B. (2005). The Large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science, 29,* 41-78.

Storkel, H. L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research, 44*, 1321-1337.

Storkel, H. L. (2009). Developmental differences in the effects of phonological, lexical and semantic variables on word learning by infants. *Journal of Child Language, 36,* 291-321.

Storkel, H. L., & Adlof, S. M. (2009). The effect of semantic set size on word learning by preschool children. *Journal of Speech, Language, and Hearing Research, 52*, 306-320.

Storkel, H. L., Armbrüster, J., & Hogan, T. P. (2006). Differentiating phonotactic probability and neighborhood density in adult word learning. *Journal of Speech, Language, and Hearing Research, 49*, 1175-1192.

Storkel, H. L., Bontempo, D. E., Aschenbrenner, A. J., Maekawa, J., & Lee, S.-Y. (2013). The effect of incremental changes in phonotactic probability and neighborhood density on word learning by preschool children. *Journal of Speech, Language, and Hearing Research, 56*, 1689-1700.

Storkel, H. L., Bontempo, D. E., & Pak, N. S. (2014). Online learning from input versus offline memory evolution in adult word learning: Effects of neighborhood density and phonologically related practice. *Journal of Speech, Language, and Hearing Research, 57*, 1708-1721.

Storkel, H. L., & Hoover, J. R. (2011). The influence of part-word phonotactic probability/neighborhood density on word learning by preschool children varying in expressive vocabulary. *Journal of Child Language, 38*, 628-643.

Storkel, H. L., & Lee, S.-Y. (2011). The independent effects of phonotactic probability and neighbourhood density on lexical acquisition by preschool children. *Language and Cognitive Processes, 26*, 191-211.

Stothard, S., Hulme, C., Clarke, P., Barnby, P., & Snowling, M. J. L., England: GL Assessment. Google Scholar. (2010). The York Assessment of Reading for Comprehension (YARC): Passage Reading Secondary.

Stothard, S. E., & Hulme, C. (1995). A comparison of phonological skills in children with reading comprehension difficulties and children with decoding difficulties. *Journal of Child Psychology and Psychiatry, 36*, 399-408.

Takashima, A., Petersson, K. M., Rutters, F., Tendolkar, I., Jensen, O., Zwarts, M. J., . . . Fernández, G. (2006). Declarative memory consolidation in humans: A prospective functional magnetic resonance imaging study. *Proceedings of the National Academy of Sciences of the United States of America, 103*, 756-761.

Taki, Y., Hashizume, H., Thyreau, B., Sassa, Y., Takeuchi, H., Wu, K., . . . Asano, K. (2012). Sleep duration during weekdays affects hippocampal gray matter volume in healthy children. *NeuroImage, 60*, 471-475.

Tamminen, J., Davis, M. H., Merkx, M., & Rastle, K. (2012). The role of memory consolidation in generalisation of new linguistic information. *Cognition, 125*, 107-112.

Tamminen, J., & Gaskell, M. G. (2013). Novel word integration in the mental lexicon: Evidence from unmasked and masked semantic priming. *The Quarterly Journal of Experimental Psychology, 66*, 1001-1025.

Tamminen, J., Lambon Ralph, M. A., & Lewis, P. A. (2013). The role of sleep spindles and slow-wave activity in integrating new information in semantic memory. *The Journal of Neuroscience, 33*, 15376-15381.

Tamminen, J., Payne, J. D., Stickgold, R., Wamsley, E. J., & Gaskell, M. G. (2010). Sleep spindle activity is associated with the integration of new memories and existing knowledge. *The Journal of Neuroscience, 30*, 14356-14360.

Taylor, J., Plunkett, K., & 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*, 60.

Tham, E. K., Lindsay, S., & Gaskell, M. G. (2015). Markers of automaticity in sleep-associated consolidation of novel words. *Neuropsychologia, 71*, 146-157.

Tong, X., Deacon, S. H., Kirby, J. R., Cain, K., & Parrila, R. (2011). Morphological awareness: A key to understanding poor reading comprehension in English. *Journal of Educational Psychology, 103*, 523.

Toppino, T. C., & Gerbier, E. (2014). About practice: Repetition, spacing, and abstraction *Psychology of learning and motivation* (Vol. 60, pp. 113-189): Elsevier.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. J. A., TX: Pro-Ed. (2012). TOWRE-2 Examiner's Manual.

Tse, D., Langston, R. F., Kakeyama, M., Bethus, I., Spooner, P. A., Wood, E. R., . . . Morris, R. G. (2007). Schemas and memory consolidation. *Science, 316*, 76-82.

Urbain, C., De Tiège, X., De Beeck, M. O., Bourguignon, M., Wens, V., Verheulpen, D., . . . Peigneux, P. (2016). Sleep in children triggers rapid reorganization of memory-related brain processes. *NeuroImage, 134*, 213-222.

van der Kleij, S. W., Rispens, J. E., & Scheper, A. R. (2016). The effect of phonotactic probability and neighbourhood density on pseudoword learning in 6-and 7-year-old children. *First Language, 36*, 93-108.

van Kesteren, M. T., Rijpkema, M., Ruiter, D. J., & Fernández, G. (2013). Consolidation differentially modulates schema effects on memory for items and associations. *PloS One, 8*, e56155.

van Kesteren, M. T., Ruiter, D. J., Fernández, G., & Henson, R. N. (2012). How schema and novelty augment memory formation. *Trends in Neurosciences, 35*, 211-219.

Veldkamp, C. L. S., Bakker, M., van Assen, M. A., Crompvoets, E. A. V., Ong, H. H., Nosek, B. A., . . . Wicherts, J. (2018). Ensuring the quality and specificity of preregistrations. *psyArxiv*.

Verhoeven, L., van Leeuwe, J., & Vermeer, A. (2011). Vocabulary growth and reading development across the elementary school years. *Scientific Studies of Reading, 15*, 8-25.

Vitevitch, M. S., Storkel, H. L., Francisco, A. C., Evans, K. J., & Goldstein, R. (2014). The influence of known-word frequency on the acquisition of new neighbours in adults: Evidence for exemplar representations in word learning. *Language, cognition and neuroscience, 29*, 1311-1316.

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., van der Maas, H. L., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science, 7*, 632-638.

Wagner, R. K., & Ridgewell, C. (2009). A large-scale study of specific reading comprehension disability. *Perspectives on Language and Literacy, 35*, 27.

Walker, S., Knowland, V. C. P., Fletcher, F. E., Cairney, S. A., Gaskell, M. G., & Henderson, L. M. (in prep). Growing up with interfering neighbours: The influence of time of learning and vocabulary knowledge on language learning in children.

Warren, D. E., & Duff, M. C. (2014). Not so fast: Hippocampal amnesia slows word learning despite successful fast mapping. *Hippocampus, 24*, 920-933.

Wechsler, D. (2011). *WASI-II: Wechsler Abbreviated Scale of Intelligence*: Pearson.

Weighall, A., Henderson, L., Barr, D., Cairney, S., & Gaskell, M. (2016). Eye-tracking the time-course of novel word learning and lexical competition in adults and children. *Brain and Language*.

Werchan, D. M., & Gómez, R. L. (2014). Wakefulness (not sleep) promotes generalization of word learning in 2.5-year-old children. *Child development, 85*, 429-436.

Wickham, H. (2009). *ggplot2: elegant graphics for data analysis*: Springer Science & Business Media.

Wickham, H. (2016). *ggplot2: elegant graphics for data analysis*: Springer.

Wilhelm, I., Diekelmann, S., & Born, J. (2008). Sleep in children improves memory performance on declarative but not procedural tasks. *Learning & Memory, 15*, 373-377.

Wilhelm, I., Kurth, S., Ringli, M., Mouthon, A.-L., Buchmann, A., Geiger, A., . . . Huber, R. (2014). Sleep slow-wave activity reveals developmental changes in experience-dependent plasticity. *The Journal of Neuroscience, 34*, 12568-12575.

Wilhelm, I., Prehn-Kristensen, A., & Born, J. (2012). Sleep-dependent memory consolidation–What can be learnt from children? *Neuroscience & Biobehavioral Reviews, 36*, 1718-1728.

Wilhelm, I., Rose, M., Imhof, K. I., Rasch, B., Büchel, C., & Born, J. (2013). The sleeping child outplays the adult's capacity to convert implicit into explicit knowledge. *Nature Neuroscience*, *16*, 391-393.

Wilkinson, K. S., & Houston-Price, C. (2013). Once upon a time, there was a pulchritudinous princess...: The role of word definitions and multiple story contexts in children's learning of difficult vocabulary. *Applied Psycholinguistics, 34*, 591-613.

Williams, S. E., & Horst, J. S. (2014). Goodnight book: Sleep consolidation improves word learning via storybooks. *Frontiers in Psychology, 5*, 1-12.

Wojcik, E. H. (2013). Remembering new words: integrating early memory development into word learning. *Frontiers in Psychology*, *4*.

Yates, M., Locker, L., & Simpson, G. B. (2003). Semantic and phonological influences on the processing of words and pseudohomophones. *Memory & cognition*, *31*, 856-866.

Yuill, N. (2009). The relation between ambiguity understanding and metalinguistic discussion of joking riddles in good and poor comprehenders: Potential for intervention and possible processes of change. *First Language*, *29*, 65-79.