Language Learning in Adults: A Role for Offline Memory Consolidation and Prior Knowledge

Isabelle Rose O’Halloran

Submitted in accordance with the requirements for the degree of

Doctor of Philosophy (PhD)

York St John University

School of Education, Language and Psychology

September 2025

Intellectual Property Statement

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others.

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr Jelena Mirković. These last six years have presented many challenges for me personally, challenges that at times made it difficult to envision reaching the finish line. Yet, your patience, kindness and unwavering support made this possible. You have been behind me every step of the way and your belief in me has been a constant source of strength and encouragement. Not only that, but you have made my PhD journey so enjoyable. Your passion, enthusiasm, and insight have been truly inspiring, making it both a joy and a privilege to work with and learn from you – an experience I will always treasure.

I would also like to thank my second supervisor, Dr Nicola Savill, for your advice, feedback and support over the years, all of which have been invaluable. Sincere thanks also go to the members of the Psychology department for creating such an inspiring and supportive environment. It has been such a wonderful community to be a part of and one that I will deeply miss. I am especially grateful to Debra and Lisa, whose friendship, humour and solidarity over the past few years made my PhD journey brighter and far less lonely. I am very glad to have shared this chapter with you both.

I am also very grateful to my friends outside academia. You have all been there for me in so many ways, from reminding me to take breaks to cheering me on. A special thank you to Liam for all the laughter along the way and for always being such a light in good times and in bad. Last but definitely not least, I would like to thank my parents and brother, Joe. You have been so patient and understanding, and in these last few months, voices of reason when I needed it most. Thank you for always believing in me.

Abstract

Sleep plays an active role in the long-term consolidation of new words (Schreiner & Rasch, 2015), but its role in grammar learning is unclear, with some studies reporting benefits (Batterink et al., 2014) and others not (Mirković & Gaskell, 2016). However, previous research has not considered how prior knowledge may influence offline consolidation in grammar learning. For instance, the Complementary Learning Systems model (CLS; McClelland, 2013) posits that existing knowledge should facilitate neocortical integration of related information at encoding. Conversely, for new unrelated knowledge, consolidation is predicted to be slower, requiring integration to happen offline. Across four experiments, we examined the role of prior knowledge in grammar learning and investigated to what extent it influences initial learning and consolidation. Participants were trained and tested on an artificial language consisting of novel two-word phrases (e.g., *tib viffeem*) referring to a familiar picture (e.g., waitress). Systematic phonological properties (*tib*, *eem*) cued grammatical categories, and varied in their relation to existing knowledge, while arbitrary properties (i.e., new words) were reflected in the form-meaning mappings (*viff-waitress*). In Experiment 1, a recall task showed evidence of a role for sleep in the consolidation of the arbitrary mappings but no benefit for the consolidation of the systematic mappings, even for those mappings that were unrelated to existing knowledge. Consistent with CLS, participants showed better immediate recall for systematic mappings related to existing knowledge compared to those that were unrelated. This advantage at initial recall was replicated in Experiments 2 and 3 but took longer to emerge when training involved a more complex artificial language (Experiment 4). Moreover, a prior knowledge advantage at initial learning was not observed in online measures (Experiments 1 & 4) and did not extend to generalisation (Experiments 1- 4). The findings are discussed in relation to models of memory consolidation.

List of Contents

List of Tables

List of Figures

Chapter 1. Literature Review

1.1 Introduction

The aim of this chapter is to provide a review of the literature to date on the role of memory consolidation in language learning. This chapter will discuss language learning within the context of a connectionist framework, in which the core components of language (grammatical and lexical knowledge) vary in their degrees of arbitrariness and systematicity (McClelland & Patterson, 2002; Seidenberg & Plaut, 2014). The review will explore the relative contributions of the declarative and procedural memory systems to the learning and consolidation of arbitrary and systematic form-meaning mappings by drawing on insights from artificial grammar learning paradigms, developmental and neuroimaging research. Furthermore, the role of sleep in the extraction of systematic regularities and strengthening of arbitrary mappings will be discussed before turning to more recent theoretical and empirical work on the role of prior knowledge and how it influences the encoding and, consequently, the consolidation of new linguistic knowledge.

1.2 Properties of Language: Arbitrary and Systematic Mappings

Broadly speaking, arbitrariness facilitates meaning individuation for individual words and systematicity supports the grouping of words into categories (Monaghan et al., 2011). Arbitrary aspects of language are often considered in reference to word learning as the association between the phonological or orthographic form of a word and its meaning is highly arbitrary in that its form (written or spoken) typically does not reflect its meaning. For example, although the nouns *'rose'* and *'hepatica'* derive from the same semantic category (e.g. flowers), they share no commonalities in the phonological or written form. This limited systematicity between word forms and their meaning is a well-established characteristic of languages around the world (Hockett, 1960).

In contrast to arbitrariness, systematicity refers to the statistical relationship between the phonological properties of groups of words and their meaning. For example, morphologically complex words such as *'teacher'* consist of both arbitrary and systematic components as the verb *'teach'* and morpheme *'–er'* combine to form a semantically related noun *'teacher'.* Such category-level systematicity is generally supported by language-specific phonological cues; for example, while the suffix -er does not carry meaning on its own, its consistent use in deriving nouns from verbs provides a cue that is thought to aid the acquisition of grammatical regularities (e.g., McClelland & Patterson, 2002). Certainly, insights from developmental research show that children are better at learning nouns and verbs if the phonological properties of the words systematically correspond to their respective grammatical categories (Cassidy & Kelly, 1991; Fitneva et al., 2009).

However, although there is evidently ample reason to distinguish between systematicity and arbitrariness and functions of language, this distinction is not entirely clear-cut in that there are also some exceptions to the arbitrariness of form-meaning mappings. Indeed, it has been argued that the link between the form of a word and its meaning is not entirely random but rather that there is often an element of iconicity (non-arbitrariness) amongst lexical items that share similar semantic properties (Perniss et al., 2010). One type of iconic mapping is referred to as 'phonesthesia' where word-initial or word-final consonant clusters tend to correlate with words of a similar meaning. As an example, the phonestheme *'gl'* is associated with light-emittance as in the words *'glisten', 'glimmer', 'glow', 'glitter',* while *'-ack'* as in *'whack'* and *'crack'* denotes forceful contact. The systematicity between vowel height and word meaning also highlights the pervasiveness of iconicity in the English language as expressions that refer to small sizes e.g. *'tiny/little'* tend to contain high vowels whereas those referring to large sizes e.g. *'large/humongous'* contain low vowels (Monaghan et al., 2011; Ohala, 1984). Similarly, onomatopoeic words such as the English words *'woof'* and *'drip'* are also exceptions to arbitrariness. Providing these examples, the strength of the arbitrary relation between word forms and their meaning is evidently a matter of degree or 'relative iconicity' as opposed to being a universal property of all form-meaning mappings. Nonetheless, at its broadest level, the form to meaning correspondences outlined above can be conceptualised into two parts and these two parts serve two fundamental functions in language acquisition: systematicity facilitates category membership through phonological cues and arbitrariness supports meaning individuation through distinctive forms (Monaghan et al., 2011).

1.3 The Contribution of Memory Processes to the Acquisition of Arbitrary and Systematic Mappings

At its broadest level, memory comprises two distinct domain general systems, each of which enable the encoding, storage and retrieval of different types of information. Declarative or explicit memory refers to the conscious recall and recognition of facts (e.g., the capital of France) and specific episodic events (e.g., one's first day at school) and is largely subserved by the Medial Temporal Lobe (MTL) which consists of structures such as the hippocampus, entorhinal, perirhinal and para-hippocampal cortices (Squire et al., 2004). Cortical regions including the pre-frontal cortex have also been implicated, along with the parietal lobes (Scoville & Milner, 1957; Squire, 2004). Conversely, non-declarative or procedural memory facilitates the unconscious recall and recognition of skill-based knowledge (e.g., riding a bike), habits, priming, habituation and sensitisation and unlike declarative memory, it is not governed by a unitary neural network but instead is dependent on the neural circuits that were activated at encoding (Reber, 2013). Processing areas associated with procedural memory include the basal ganglia/cortico-striatal system, neocortex and cerebellum (Gabrieli, 1998; McClelland et al., 1995; Ullman, 2001; 2004).

The specific anatomical regions in which these memory types reside have been used by neurocognitive frameworks to dichotomise memory into a declarative and procedural system and henceforth inform our understanding about the underlying memory mechanisms supporting aspects of language learning (e.g., McClelland & Patterson, 2002; Ullman, 2001; 2004). The demarcation of systematic and arbitrary properties of language has been useful in exploring the contribution of specific memory processes to the learning of distinct properties of language; namely the acquisition of vocabulary and grammar (e.g., Ullman, 2001; 2004). This next section will discuss how broader theoretical models of memory and learning have been applied to word and grammar acquisition specifically.

1.3.1 The Complementary Learning Systems Framework

The mapping of distinct declarative memory processes to specific structural properties of language is largely based on systems consolidation models which have proposed that the acquisition of new knowledge is supported by two key systems: a hippocampal and a neocortical system (Alvarez & Squire, 1994; Marr, 1971). According to this, the neocortex is primarily involved in the long-term retention of memories while the hippocampus is critical to memory encoding. The notion of a dual systems account is further discussed in the Complementary Learning Systems (CLS) model by McClelland et al., (1995), a computational framework for understanding the specific contributions that the hippocampus and neocortex make to learning and memory. By this account, the hippocampus is primarily involved in the rapid encoding of arbitrary mappings because of its ability to minimise interference between related experiences through pattern separation, a process whereby similar neural codes are decorrelated or orthogonalised to allow similar experiences to be encoded independently of one another (McClelland et al., 1995; Kumaran & McClelland, 2012). Conversely, the extraction of statistical structure (systematicity) from the environment is thought to be better suited to neocortical encoding. This is because the neocortical system uses an overlapping, distributed encoding scheme, meaning it is better able to extract shared features across ensembles of experiences (McClelland et al., 1995; O'Reilly & Rudy, 2001).

Davis and Gaskell (2009) explored the contribution of long-term memory processes in word learning within the context of a CLS framework. Based on a large-scale review of behavioural, neuropsychological and neuroimaging research, Davis and Gaskell (2009) suggested that complementary systems in the hippocampus and neocortex are involved in the ability to acquire new words and integrate them into the existing mental lexicon across two stages. The first stage involves rapid initial familiarisation of the word mediated by the hippocampus as the hippocampal system is thought to be involved in the rapid encoding of new arbitrary linguistic information (McClelland et al., 1995; i.e. form-meaning mappings as in vocabulary learning). The second stage is one of slow offline consolidation which is mediated by the neocortical structures. During this stage, newly encoded words that are initially represented as distinct episodic traces in the hippocampus are repeatedly reactivated (e.g., through hippocampal replay during sleep) and gradually integrated with existing lexical knowledge in the neocortex for long-term retention.

This neurocomputational account of word learning is consistent with both behavioural and neuroimaging research. In the behavioural literature, a discerning measure of whether a new word has been integrated into the mental lexicon is to assess whether the new word interacts with existing lexical items by competing for activation during word processing (i.e., engaging in lexical competition).

Indeed, it is well established that spoken word recognition involves the temporary activation of multiple lexical candidates that overlap with the incoming speech signal, and that these candidates compete as the signal unfolds over time (e.g., McClelland & Elman, 1986). For instance, when listeners hear the onset of a spoken word, several phonologically similar candidates may initially become activated, with these candidates competing until enough input is available to identify the intended word.

To assess whether new words engage in lexical competition, many studies have used pause detection paradigms (Mattys & Clark, 2002). In this task, participants monitor connected speech for a brief, artificially embedded silent pause. The speed with which listeners detect this pause is used as an online index of the overall amount of lexical activity evoked with slower pause detection interpreted as reflecting greater lexical activity. This is because when multiple candidates are activated, they draw on processing resources that would otherwise be available for detecting the pause. Crucially, studies using this paradigm have shown that, for novel words to compete with established lexical entries, an offline consolidation period is usually required.

For example, in a study by Gaskell and Dumay (2003) adult participants were taught novel words (e.g., *cathedruke*) that were selected to be close phonological competitors of established words in the lexicons of participants (e.g., *cathedral*). Participants were then tested using a speeded pause detection task, in which they had to detect 200 ms pauses embedded in existing words for which a novel competitor had been taught (*cathedral*) and in matched control words for which no novel competitor had been taught. Immediately after learning, participants were highly accurate in recognising the novel competitors in a two-alternative forced choice task, where they were presented alongside phonologically overlapping foils (e.g., *cathedruke* vs. *cathedruce*). However, no difference in pause detection latencies was observed between existing words with newly learned competitors and matched control words. It was only after a delay period of 1 week that participants became significantly slower to detect the pause inserted into the existing word form (*cathedral*) compared to the control words for which no competitor had been taught. This pattern of results is consistent with other research that has used similar paradigms and also detected greater lexical interference (as indexed by increased pause detection time) after a delay period (e.g. Bakker et al., 2014; Dumay & Gaskell, 2007; Tamminen & Gaskell, 2008). Thus, these findings offer support for the proposal that lexical integration is a protracted process.

Insights from neuroimaging research also substantiate Davis and Gaskell's (2009) account of word learning. For example, several studies have used functional magnetic resonance imaging (fMRI) to investigate the underlying neural mechanisms responsible for word encoding and consolidation. Consistent with the CLS inspired account of word learning, exposure to novel words initially elicited an elevated hippocampal response which declined over later presentations (Breitenstein et al., 2005; Davis et al., 2009). Moreover, relevant neocortical regions (e.g., Superior Temporal Gyrus) responded to words learned on the first day of scanning in the same way as to unfamiliar words, whereas words learned the previous day elicited activation patterns similar to known English words (Davis et al., 2009). Thus, like the behavioural evidence outlined above, the evidence from neuroimaging studies outlined here supports the CLS model of word learning by Davis and Gaskell (2009) that word consolidation via integration into the mental lexicon is a gradual process that is initially dependent on the hippocampus.

However, despite clear support for the CLS model in word learning, recent research has presented a challenge to the proposal that the hippocampus and neocortex are involved exclusively in the consolidation of their respective mappings. For instance, although considerable evidence supports the division of the two memory systems in the consolidation of form-meaning mappings and the gradual extraction of statistical regularities, the CLS overlooks an additional category of learning: our ability to acquire regularities rapidly – over minutes or hours (Schapiro et al., 2017). Indeed, several studies provide empirical support for the involvement of the hippocampus in the rapid extraction of regularities (e.g. Bornstein & Daw, 2012; Harrison et al., 2006; Schapiro et al., 2014). Furthermore, Schapiro et al. (2017), trained a neural network model of the hippocampus on sequences of items containing temporal regularities. Interestingly, the researchers found that two anatomical pathways, the monosynaptic pathway and the trisynaptic pathway, facilitate the learning of statistical regularities and individual episodes (item-specific memory) respectively. Thus, contrary to the strict demarcation proposed by the CLS, this evidence would suggest that the hippocampus is involved in *both* the associative binding of arbitrary information and aspects of rule extraction.

1.3.2 Declarative/Procedural Framework

An alternative to the CLS model that considers both the declarative and the procedural memory systems was proposed by Ullman (2001; 2004). The declarative/procedural (DP) model (Ullman, 2001; 2004) suggests that predictions can be made as to how properties of language are acquired based on their association with the two domain general memory systems. According to this, declarative memory and the mental lexicon are subserved by the same brain system, the MTL, as both are involved in the acquisition of arbitrary relationships (associative binding) including non-linguistic conceptual-semantic information and idiosyncratic, word-specific knowledge. This is consistent with insights from neuropsychology in which amnesic patients with MTL atrophy typically have difficulty acquiring information of an arbitrary nature (Gabrieli et al., 1998; Graf & Schacter, 1985; Postle & Corkin, 1998) and PET and fMRI research (Damasio et al., 1996; Newman et al., 2001). As an example, Breitenstein et al. (2005) used fMRI to investigate the neural mechanisms underpinning lexical acquisition. Inside the fMRI scanner, participants were presented with novel word-picture pairings and subsequently tested on their recognition of the novel words. The authors reported that learning of the novel items was associated with increased activation in the left hippocampus. Thus, as predicted by the DP model, the MTL appears to support both the acquisition of arbitrary form-meaning mappings in addition to domain-general declarative knowledge.

In contrast, the DP model predicts that grammar is linked to the procedural memory system, suggesting that grammar learning and procedural skills must be subserved by the same brain networks, namely those involved in the computation of rule-based structures that govern the regularities of language (Ullman, 2001; 2004). Specifically, procedural memory is thought to be involved in learning all aspects of grammar that conform to a hierarchical or sequential structure such as syntax, derivational and inflectional morphology, aspects of phonology as well as other complex forms, such as subordination routines and long-distance dependencies (e.g. in the sentence, *'David is easy to consider it impossible for anyone but a genius to try to talk to'* the noun *'David* is placed at the start of the sentence as opposed to the end, thus there is a gap: nothing following 'to' where an object would normally follow). Indeed, behavioural research investigating second language (L2) learning has found that individual differences in procedural memory ability predict performance on assessments measuring the acquisition of L2 syntax. For example, the study by Brill-Schuetz and Morgan-Short (2014) trained adult native English speakers on an artificial language with a grammatical gender system consisting of a subject-object-verb order across four either explicit or implicit training conditions. A Serial Reaction Time Task (SRTT) and the Weather Prediction task were used as measures of procedural memory. Grammar learning was assessed using a grammaticality judgement task in which participants had to judge whether the constituent order of an untrained item was correct based on items they had been exposed to during training. The results showed that higher procedural memory ability, particularly if the language was learnt implicitly, is positively related to performance on the L2 syntactic judgement task. Thus, these findings suggest an association between procedural memory and grammatical rule learning, the strength of which appears to depend on the learner's own procedural memory ability.

In addition, some research has directly compared the contributions of both declarative and procedural memory to the acquisition of grammatical regularities. Morgan-Short et al. (2014) assessed whether implicit L2 grammar learning is linked to the procedural or declarative memory system. To measure declarative memory, two tasks were administered: one verbal and one non-verbal. For the verbal task participants were assessed on their learning of pairs of novel words and their English translation. In the non-verbal task participants were assessed on their recognition of familiar abstract images. The Weather Prediction task and the Tower of London task made up the procedural measures and learning of the artificial language was assessed using a grammaticality judgement task (the same as in Brill-Schuetz & Morgan-Short, 2014) administered at both the start and end of training conditions. At the initial stages of learning, the declarative system was solely related to syntactic development whereas the later stages of learning were related exclusively to procedural learning ability. Moreover, individual differences in the two memory systems accounted for a substantial amount of variance at both early and late stages of development with declarative memory serving as a unique predictor at the initial stages of acquisition and procedural memory predicting later learning stages. Similarly, Yi et al. (2014) assessed speech-category learning in adults using fMRI and found the involvement of declarative and procedural memory to be contingent upon the stage of learning (i.e. early or late). Feedback-dependent learning of the speech categories was used to index acquisition of the categories whereby positive feedback indicated successful learning. Consistent with Morgan-Short et al. (2014), the researchers found that participants' use of explicit/declarative memory was associated with earlier stages of learning, with a gradual shift towards implicit/procedural strategies as learning progressed. This is in line with computational models of associative learning (Gluck & Myers, 1993). Thus, these findings support the contribution of *both* domain-general memory systems to grammar learning at different stages in learning and in doing so, challenge the simplicity of earlier frameworks' dissociation between the two memory systems and the type of information (i.e. arbitrary form-meaning mappings/systematic regularities) they encode.

Although the conceptual distinctions made by multiple-memory systems frameworks is a useful simplification (Shanks & Berry, 2012), it presents the assumption that single brain regions are involved in single functions. However, the experimental tasks used to measure procedural and declarative memory rarely recruit one of these systems in isolation (Dew & Cabeza, 2011; see Sawi & Rueckl, 2019 for a full review). This lack of one-to-one mapping is especially true of the hippocampus which has been shown to interact with components of other memory systems such as the striatum. For example, Poldrack et al. (2001) investigated the interaction of the declarative and procedural memory systems during classification learning in humans using fMRI. Differential engagement of the MTL and striatal systems was contingent upon the nature of the task and whether they indexed declarative or procedural memory. Furthermore, both domain-general systems were found to be negatively correlated across individuals, suggesting that the two networks compete during learning processes. This is consistent with previously mentioned findings that have indicated that engagement of the declarative and procedural memory systems is not just dependent on the type of learning (i.e. word learning versus grammar learning) but also the time-course (Morgan-Short et al., 2014; Yi et al., 2014). The neural substrates underpinning declarative and procedural memory have also been found to interact positively with each other, suggesting that the two systems may cooperate to guide behaviour in some tasks. For instance, as evidenced in various empirical studies, procedural memory has been shown to engage the MTL, specifically, the hippocampus during implicit procedural tasks such as sequence learning (Schendan et al., 2003).

A theoretical framework proposed by Shohamy and Turk-Browne (2013) attempts to account for the involvement of the MTL in procedural memory. By this account, the hippocampus' involvement in most behavioural functions combined with its extensive anatomical connections to other brain regions such as the temporal cortex, the dorsolateral pre-frontal cortex, the lateral parietal cortex and striatum (Goldman-Rakic et al., 1984; Rockland & Hoesen., 1999), suggests that this structure may be responsible for exerting direct control over cognitive functions. In line with this notion, some research has found the hippocampal system to be implicated in the acquisition of statistical learning, a process whereby regularities or co-occurrences are extracted from the environment (Aslin & Newport, 2012). For instance, in a study by Witt et al. (2013) the effect of task instructions on learning an artificial grammar in 5–8-year-old children was tested. In this particular study, children received implicit training on the artificial grammar and were then either assigned to an explicit or implicit test condition. In the explicit test condition, the children received instructions that made explicit reference to the regularities they had been exposed to in the training phase whereas in the implicit condition, the children were given neutral instructions that did not make any references to the regularities they had been implicitly presented with during training. Interestingly, it was reported that older children had significantly higher task performance than younger children when given the explicit task instructions. However, no differences were observed between the older and younger children's grammar learning when they were given implicit task instructions. This age difference was attributed to developmental differences in the processes underlying declarative memory (i.e. hippocampus and prefrontal cortex) as these connections would be more matured in older versus younger children (Gomez & Edgin, 2016) and thus they were better able to engage the declarative system to improve their learning. Overall, when taken from a computational perspective, statistical learning is arguably more systematic than arbitrary. Thus, this evidence presents a challenge to the core principles of the DP model as it suggests that the hippocampus is, at least in some cases, contributing to the acquisition of systematic linguistic knowledge.

1.4 Offline Consolidation in Word and Grammar Learning

The previous section has discussed the contribution of declarative and procedural memory to the acquisition of arbitrary and systematic mappings and how their structural properties may provide insight into the underlying memory mechanisms supporting word and grammar learning. This next section will review evidence which has investigated the contribution of offline consolidation (i.e. sleep versus wake) to word and grammar learning.

The notion that sleep supports memory consolidation is long-standing (e.g., Jenkins & Dallenbach, 1924). Sleep consolidation benefits were initially thought to be solely a protective factor in that sleep would limit interference of newly encoded memories. However, only recently has the physiology of sleep been directly associated with specific memory processes. For instance, based on empirical findings, the Dual Process Hypothesis by Plihal and Born (1997) posits that Slow Wave Sleep (SWS), a deep sleep stage characterised by large oscillations on polysomnography recordings (Achermann & Borbely, 2003), may be beneficial for the consolidation of hippocampally dependent, declarative memories (Rasch et al., 2007). Conversely, Rapid Eye Movement (REM) sleep which is characterised by low amplitude, desynchronised neural oscillations that resemble waking brain activity, is thought to be beneficial for the consolidation of the less hippocampally dependent, procedural memories (Wagner et al., 2003).

Due to its wake-like EEG and oculomotor activity, REM sleep was initially thought to be the sleep stage that supported memory consolidation (Rasch & Born, 2015). However, sleep deprivation studies that have selectively deprived individuals of REM sleep have led to a shift in focus on the role of SWS in memory consolidation (Vertes & Eastman, 2000). Consequently, this gave rise to several dominant theories in the field such as the Active System Consolidation (ASC; Diekelmann & Born, 2010; Born & Wilhelm, 2012) and the Information Overlap to Abstract (iOtA; Lewis & Durrant, 2011) model. The ASC framework theorises that the hippocampus is involved in binding the components of several declarative memories into one mnemonic representation. The newly encoded engram is then repeatedly reactivated and propagated across the associated memory network while components of the recently encoded memory are reactivated across the neocortex (Diekelmann & Born, 2010). From this perspective, sleep, in particular, is thought to support the integration of hippocampally dependent memories into the neocortex by encouraging cross-talk between the two systems (e.g., Diekelmann & Born, 2010).

Relatedly, the iOtA model (Lewis & Durrant, 2011) builds on the ASC but is predominately concerned with the formation of schematic memory (a collection of related, long-term memories that provide a knowledge-consistent representation of an experience based on previous, similar experiences). According to this, the connections between the related memory traces (those belonging to the same schema), are strengthened through hippocampal reactivation during sleep. This overlapping replay of associated memories selectively amplifies the response of the overlapping neocortical neurons and in doing so, integrates the newly encoded information with existing knowledge and in turn, progressively builds schematic representations. Thus, both the ASC and iOtA accounts posit that sleep plays an active role in the consolidation of new memories.

Indeed, a substantial body of empirical work has accumulated over the past decades, providing support for the active role of sleep in the consolidation of new memories (e.g., Rasch et al., 2007; Rasch & Born, 2013; Stickgold, 2005; Walker & Stickgold, 2004).

However, more recent reviews and findings have begun to call into question the replicability and generalisability of some of these sleep-related memory effects (Ackermann et al., 2015; Cordi & Rasch, 2021). In addition, a meta-analysis by Newbury et al. (2021) suggests that the effect sizes associated with post-learning sleep deprivation on memory performance may be smaller than previously assumed.

Nevertheless, within the language learning literature, the ASC model's prediction that sleep contributes to the consolidation of arbitrary knowledge has been used to account for findings of improved word learning after a period of sleep (e.g., Davis & Gaskell, 2009).Taken from this perspective, the hippocampus is hypothesised to play a fundamental role in the binding of word forms and their meaning through spontaneous re-activation during sleep. This replay/reactivation of the recently encoded phonological form-meaning mappings causes localised synaptic downscaling and in turn, the distribution and integration of word forms into the long-term neocortical network (Davis & Gaskell, 2009; Rasch & Born, 2013). Thus, in the context of language acquisition, sleep is thought to support the integration of recently encoded lexical representations into the existing long-term lexical network, a process that gives rise to the consolidation of arbitrary word form-meaning mappings (e.g. Gaskell et al., 2014).

1.4.1 Word Learning and Offline Memory Consolidation

Word learning is a complex, multi-faceted process that requires a variety of steps from associative binding (i.e. linking word-forms to their meaning), to the encoding and long-term storage of information about a new word so that it can be later recalled on demand. Within the language learning literature, *'word learning'* is often used to denote different aspects of learning and memory, including retention, recall, and generalisation. For the purposes of this review, the focus will be on research which has explored how offline consolidation may be beneficial for the retention, including the integration of individual words and referents into existing memory networks, of arbitrary form-meaning mappings.

Several studies have demonstrated a role for sleep in the integration of word forms and their meanings into the mental lexicons of both adults (e.g., Davis et al., 2009; Dumay & Gaskell, 2007; Wang et al., 2017) and children (e.g., Henderson et al., 2012). As an example, Tamminen et al. (2010) presented two groups of adult participants with novel spoken words that overlapped phonologically with existing English words (e.g. *cathedruke*, derived from *cathedral*). The primary aim of this study was to explore the differential effects of sleep and wake on novel word consolidation. For the participants in the sleep group, training of the words was administered in the evening, followed by an initial test and a polysomnography recorded night of sleep. Conversely, the wake group were administered training on the novel items in the morning, followed by an initial test and remained awake for the same amount of time the sleep group slept. Lexical integration was assessed using a lexical decision task in which participants made speeded word/nonword judgements to familiar base-words (e.g., *cathedral*), with slower responses to these words taken as evidence that the newly learned nonwords had become integrated into the mental lexicon and were competing with the existing base-words. Following acquisition of the novel words, recognition of the familiar words was slowed in both groups but only after the retention period (sleep/wake), indicating that the new words had been integrated into the mental lexicon. However, for the sleep group, Stage 2 sleep spindle activity on the polysomnography recordings was associated with lexical integration. The role of sleep spindles in memory consolidation has been linked to their timing during NREM sleep. Specifically, spindles tend to increase during the up-state of slow oscillations (Mölle et al., 2002) and are temporally aligned with hippocampal ripple activity (Siapas & Wilson, 1998) and it is this temporal coordination that is thought to support communication between the hippocampus and neocortex, helping newly learned information become stabilised and integrated into existing long-term knowledge. The findings by Tamminen et al. (2010) thus show how sleep appears to be important for lexical acquisition, specifically, the integration of novel words into long-term neocortical networks.

However, not all findings are consistent with the view that sleep is required for lexical integration. For example, Szmalec et al. (2012) used a Hebb repetition learning paradigm, in which participants repeatedly recalled visually presented sequences of syllables, with one particular sequence repeated on every third trial. Crucially, these repeated sequences contained trisyllabic nonword forms (e.g., *sa-fa-ra* in the sequence *la-va-bu-sa-fa-ra-re-si-di*) that overlapped with existing Dutch words (e.g., safari). In a first experiment, Szmalec et al. (2012) tested whether the nonwords learned through Hebb repetition would later behave like lexical competitors for the existing Dutch words. When tested 24 hours after learning, participants were slower to respond to the overlapping real words than to matched control words in both a pause detection task and an auditory lexical decision task, suggesting that the learned nonwords had become integrated into the mental lexicon and were engaging in lexical competition with existing words. Interestingly, however, Szmalec et al. (2012) showed that there was no specific role for sleep in the emergence of these competition effects: in a follow-up experiment, competition effects were observed after a 12-hour delay, regardless of whether that delay included sleep.

Moreover, some research has shown that competition effects can also emerge without a delay (i.e., within a single day). For example, Lindsay and Gaskell (2013) observed same-day competition effects when training on novel words was interleaved with exposure to their existing competitors, such that participants were repeatedly exposed not only to the novel words (e.g., *cathedruke*), but also to their existing phonological neighbours (e.g., *cathedral*) through repeated lexical decision testing across the day. Along similar lines, Kapnoula et al. (2015) and Kapnoula and McMurray (2016) found evidence for lexical competition *immediately* after training. Taken together, although some research has observed a specific role for sleep in lexical integration via lexical competition effects (e.g., Dumay & Gaskell, 2007; Tamminen et al., 2010; Wang et al., 2017), the findings outlined above highlight that lexical integration can, in some circumstances, also occur without sleep.

Several suggestions have been put forward in attempt to explain some of the discrepancies in these findings. For instance, in their study, Lindsay and Gaskell, (2013) suggested that repeated exposure to existing words may induce plasticity and promote the integration of the novel, phonologically related representations into the mental lexicon. This is consistent with the CLS account that the integration of new knowledge into the neocortical network can also occur during wake as a consequence of interleaved exposure to new and existing knowledge (McClelland., 1995). Furthermore, differences in encoding conditions have also been considered. For instance, Szmalec et al. (2012), who found no specific role for sleep in the lexicalisation of novel words, used Hebb repetition learning, an implicit learning paradigm, to train participants. However, it has been suggested that the implicit encoding of words may rely on mechanisms that are less dependent on sleep-mediated consolidation (Palma & Titone, 2021; Schimke et al., 2021). On the other hand, it has been argued that the tests that are widely used to assess lexical competition effects (e.g., lexical decision, pause detection) may be too coarse to detect the presence of subtle or weaker immediate lexicalisation effects (Kapnoula et al. 2015). This could explain why studies that have used alternative measures have observed competition between novel and existing words immediately after learning (Kapnoula et al. 2015; Kapnoula & McMurray, 2016).

Alongside some of the above evidence from lexical competition paradigms, some research has evidenced a causal role for sleep in word consolidation using Targeted Memory Reactivation (TMR). TMR paradigms typically involve replaying sound cues during SWS that were previously associated with specific word items. It is thought that by playing associated sound cues during sleep, the associated memory trace (i.e. a word form) will be replayed and reactivated, thus facilitating consolidation. Schreiner and Rasch (2015) investigated whether verbal cueing of newly encoded Dutch words during sleep improved their acquisition in native speaking German participants. Participants were cued on the Dutch words either while they were awake or during non-rapid eye movement sleep (NREM). Consistent with previous research, reactivating the memory for the Dutch words via an associated cue during NREM sleep improved memory for the German translation of the words. In contrast, recall of the non-cued words (words that were not reactivated during NREM sleep) was no different to a control group who received no cue during sleep. Importantly, successful cues were often followed by slow oscillations and increased theta and spindle activity which is consistent with other research that has investigated the association between slow oscillations and the consolidation of arbitrary knowledge (Staresina et al., 2015). This suggests that slow oscillations during NREM sleep may provide a temporal framework for the transfer of hippocampally dependent memory traces into the neocortex. Taken together, this demonstrates a clear role for sleep in the learning of arbitrary linguistic components; namely word-form acquisition and the process of lexical integration.

1.4.2 Grammar Learning and Offline Memory Consolidation

A key aspect of proficient use of grammar is the learner's ability to abstract and apply grammatical regularities from previously learned experiences to new examples. In contrast to the encoding of arbitrary knowledge, generalisation requires a continual, adaptive response to stimuli that falls outside the conditions in which it was presented during training. In the context of language acquisition, language is open-ended such that its components are often encountered in varying contexts, thus, language learners must be able to generalise shared linguistic principles to new word combinations.

Similar to word learning, some research has demonstrated a role for sleep in the abstraction and generalisation of grammatical regularities. For example, using a TMR paradigm, Batterink and Paller (2017) assessed whether reactivating memories during sleep using cues would improve generalisation of grammatical regularities in an artificial language. Participants were presented with 54 phrases in an artificial language and using a trial and error based interactive learning paradigm, were instructed to place them in their correct order (i.e. in accordance with the hidden grammatical rules of the language). For one group of participants, the correct phrases were replayed during an afternoon nap, while the other group were played stimuli from an unrelated task. The authors found that those participants who were re-exposed to the phrases from the new language during sleep showed larger gains in grammatical generalisation compared to those who were administered the control cues. In line with this, Batterink et al. (2014) reported that adults' ability to extract hidden regularities in a speeded categorisation task improved following a period of sleep. The findings observed by Batterink and Paller (2017) and Batterink et al. (2014) are in line with some other research which has observed improvements in the generalisation of novel grammatical mappings following a period of offline consolidation (Tamminen et al., 2012, 2015), as well as developmental work suggesting that sleep can promote the abstraction of grammatical dependencies in infants (Gómez et al., 2006).

The role of sleep in the learning of systematic knowledge is consistent with the proposals made by the CLS (McClelland et al., 1995), that the transfer of newly encoded, hippocampally dependent memories to the neocortex should give rise to the abstraction of systematic knowledge. This is because once new memory traces have been integrated into the neocortex they are represented in an overlapping and distributed manner, emphasising their shared features. Thus, from this view, sleep should support the consolidation of systematic knowledge as hippocampal reactivation during sleep plays a key role in the integration of new memories into the neocortex.

However, not all studies have found evidence of a role for sleep-related consolidation in the generalisation of systematic knowledge. For example, Mirković and Gaskell (2016) explored the role of sleep-related memory processes in the acquisition of new form-meaning mappings. In this study, participants were administered an artificial language with a grammatical gender system that had varying arbitrary and systematic elements. The arbitrary component of the language was administered via stem-referent mappings (e.g., *scoiff*-ballerina; *mof*-butterfly) and a more systematic element provided in the mapping between determiners/suffixes and common semantic features such as natural gender (e.g., *ked…ool/aff* and male; *tib…esh/eem* and female). Training on the new language was administered using a word-picture matching task and participants were tested after 2-hour consolidation period which consisted of either sleep or wakefulness. Participants in the nap group saw larger gains in the consolidation of arbitrary mappings (i.e. individual vocabulary items) compared to the group who did not nap. Conversely, there was no evidence of a sleep benefit for the generalisation of systematic mappings. The study by Mirković and Gaskell (2016) was the first to examine the contribution of sleep to the consolidation of a fully-fledged artificial language that combined both arbitrary and systematic mappings. Therefore, it was suggested that when learned simultaneously, arbitrary and systematic knowledge likely compete for consolidation, with the arbitrary mappings which would have the strongest hippocampal representation at encoding being prioritised during post-learning sleep. However, the acquisition of the more systematic grammatical mappings may be strengthened across subsequent consolidation periods. Thus, the absence of a role for sleep in Mirković and Gaskell (2016) could reflect the fact that relative to arbitrary knowledge, systematic knowledge is less dependent on the hippocampus at encoding and therefore less likely to benefit from hippocampal reactivation during sleep.

1.5 How Prior Knowledge Influences Encoding and Consolidation

In addition to the distinctions between arbitrary and systematic knowledge outlined above, one other important aspect of the CLS model relevant to understanding the encoding and consolidation of new linguistic knowledge concerns how new information relates to existing knowledge. Recent revisions to the CLS model of memory (Kumaran et al., 2016; McClelland, 2013; McClelland et al., 2020) consider the role of prior knowledge and how it might influence learning and consolidation, suggesting that the rate of consolidation is not fast or slow but prior knowledge dependent. According to CLS, memories for new experiences are rapidly acquired by the hippocampus and over time are integrated with existing neocortical representations for long-term retention. However, the time course of the learning and consolidation process is thought to be dependent on the degree to which the new information is consistent with existing knowledge. This is because when new information is inconsistent with existing knowledge it poses a greater risk of interference to the neocortically based long-term representations. To prevent interference, neocortical integration is thought to advance more slowly via interleaved learning, the process by which new information is repeatedly presented to the neocortex and interleaved with known information during hippocampal reactivation. Since prior knowledge consistent information presents less risk of interference to existing representations, it is integrated more swiftly.

Although the underlying neural mechanisms are not entirely clear, CLS posits that the hippocampus does play a role during the encoding of new, prior knowledge consistent memories, at least under the conditions of spatial learning (Tse et al., 2007; 2011); however, such representations may only be hippocampally dependent for a short amount of time before they are integrated into the neocortex. Instead, the availability of prior knowledge at encoding may support learning as learners can capitalise on their existing knowledge to facilitate their learning of new, related information. Specifically, since prior knowledge is already represented in neocortical networks, new information that aligns closely with this existing knowledge can be rapidly integrated into these networks, potentially reducing the need for extensive hippocampal reactivation offline (Kumaran et al., 2016). In sum, by the CLS account, prior knowledge related information is less likely to cause interference to existing knowledge at encoding. Instead, the availability of prior knowledge at encoding may facilitate the learning and neocortical integration of new, related knowledge. By contrast, new knowledge that is unrelated to prior knowledge may be more dependent on the hippocampus at encoding and as a result subject to stronger offline consolidation.

One other possibility is that the integration of prior knowledge related information occurs directly within the neocortex, without engaging the hippocampus. Van Kesteren et al. (2012) present a framework called SLIMM (Schema-Linked Interactions between Medial prefrontal and Medial temporal regions) which proposes an alternate pathway via the medial prefrontal cortex (mPFC) through which information that is congruent to existing knowledge (or a schema) can be rapidly integrated with existing neocortical representations. According to SLIMM, new information that is highly congruent with a pre-existing schema activates the pre-existing schema. This congruency is then detected by the mPFC which is assumed to inhibit the medial temporal lobe (MTL: a set of regions containing the hippocampus) to prevent details unrelated to the schema being encoded. The mPFC then facilitates the encoding of the new schema-consistent representations by strengthening the neocortical connections between these representations and the pre-existing schema, resulting in fast learning, directly within the neocortex. By contrast, when information is incongruent to an existing schema, the mPFC will activate the MTL to encode that information.

Although SLIMM makes no predictions about sleep-dependent consolidation, standard consolidation theory predicts that sleep plays an active role in the consolidation of MTL based representations via hippocampal replay (e.g., Wilson & McNaughton, 1994). Given this, the increased dependence of these incongruent representations on the MTL at encoding would suggest that when information is unrelated to existing knowledge, it is more likely to be influenced by sleep-related consolidation. In contrast, because they are assumed to be encoded by the mPFC, prior knowledge related representations may be less influenced by consolidation during sleep. Thus, the predictions outlined in SLIMM are broadly in line with CLS: prior knowledge related information may be integrated more swiftly with neocortically based knowledge and the comparatively limited involvement of the hippocampus during encoding may mean it is less susceptible to hippocampal reactivation during sleep.

Prior knowledge dependence is also relevant for a third model of the relationship between new and existing knowledge: in line with CLS and SLIMM, the information Overlap to Abstract (iOtA) model (Lewis & Durrant, 2011) suggests that newly learned information is better assimilated if it is related to existing knowledge as representations that overlap with existing representations form richer connections during learning. However, unlike CLS and SLIMM, the iOtA model suggests that it is during offline consolidation that the connections between existing representations and new, related ones will benefit as these shared connections will be simultaneously reactivated during sleep. This reactivation is assumed to strengthen the connections formed between new representations and neocortically related ones during learning thereby facilitating their integration. Unlike new representations related to existing knowledge, new representations that have little overlap with existing knowledge will forge fewer connections during learning (because of the limited existing connections available) and will therefore benefit less from reactivation during sleep. Although iOtA, like SLIMM and CLS assumes that when new information is consistent with existing knowledge it will be integrated more efficiently than when it is inconsistent, it additionally suggests that new representations consistent with existing knowledge would be more likely to be strengthened during sleep. In contrast, the predictions made by SLIMM and CLS suggest that new consistent representations may be integrated within the neocortex immediately at encoding, reducing the need for them to be strengthened by offline processes during sleep.

Within the word learning literature, some research has explored the influence of prior knowledge and specifically set out to address *when* it is helpful for word learning: at initial encoding or after a period of offline consolidation. For example, James et al. (2019) investigated the contributions of prior knowledge to initial learning and over a consolidation period by manipulating the psycholinguistic properties of the to-be-learned words. In this study, adults and 7–9-year-old children were taught novel words (e.g., *ballow*) that varied in similarity to English word-form "neighbours" (e.g., *wallow, bellow*) such that the words were either phonologically distinct from English words (had no phonological neighbours) or shared similarity with English words (had phonological neighbours). Thus, to assess the influence of prior lexical knowledge, the researchers manipulated the individual properties of the novel words so that they were similar or dissimilar in form to existing words. James et al. (2019) refer to this property (number of phonological neighbours) as 'local prior knowledge' and distinguish it from other forms of prior knowledge such as the size and richness of the learner's vocabulary which they refer to as 'global prior knowledge.' To investigate when connections to local prior knowledge would benefit learning, memory recall of the novel word-forms was assessed immediately after training, the next day and 1 week later. Both children and adults were significantly better at recalling words *with* neighbours than words *without* neighbours immediately after training; however, words *without* neighbours showed greater improvements after a long retention interval of 1 week, where there would have been opportunities for offline consolidation. Interestingly, words with neighbours showed no improvement over the 1-week interval. The authors suggested that these findings are in line with the CLS account: the words with no connections to prior knowledge are more dependent on the hippocampus at encoding and are therefore more likely to show greater improvement across a retention interval that includes offline consolidation. Conversely, words with stronger connections to prior knowledge may engage with the neocortex immediately, reducing the need for this integration to happen offline.

However, prior knowledge may not be helpful for all aspects of word learning. Some evidence has demonstrated that related semantic knowledge may interfere with the learning of new word-forms and their meanings. For example, James et al. (2023) investigated the contributions of semantic prior knowledge to word learning. In this study, adults and school-aged children were taught novel words assigned to novel concepts with either sparse or dense connections to existing semantic knowledge. Each concept was made novel by taking an existing base concept (e.g., gorilla) and adding an additional feature (e.g., that has green skin). To manipulate the semantic richness of the novel words, novel concepts with dense connections to existing semantic knowledge (high semantic neighbourhood density) were created by selecting base concepts that shared a lot of features with other existing concepts. In contrast, novel concepts with sparser connections to existing semantic knowledge (low semantic neighbourhood density) were created using base concepts that shared fewer features with existing concepts. To assess the influence of prior knowledge on word learning before and after opportunities for consolidation, participants were tested on their recall of the novel word forms and their definitions immediately after training, the next day and 1 week later. At immediate test, both children and adults exhibited poorer performance for words assigned to high- versus low- semantic density concepts. The authors suggested that the novel words with high semantic density concepts likely co-activated multiple related concepts at encoding which led to interference in establishing a new semantic representation. Crucially, performance did not change across the week, suggesting that the interference participants experienced during learning was not remediated by opportunities for offline consolidation. In line with this, other research that has investigated the influence of semantic prior knowledge on new word acquisition in adults and pre-school children has also reported poorer performance for novel words with denser connections to existing semantic knowledge (Storkel & Adlof, 2009; Tamminen et al., 2013). Thus, in contrast to the benefits observed for novel words with phonological neighbours (James et al., 2019), related semantic knowledge may interfere with new word learning.

Overall, the experimental findings outlined above are broadly in line with the predictions by CLS and SLIMM: that when new information is related to existing knowledge, it engages the neocortex directly during encoding, with little engagement from the hippocampus and is therefore less influenced by offline consolidation processes.
Unlike the studies that assessed the role of local prior knowledge, research that assessed global prior knowledge (i.e., individual differences in vocabulary size or richness) has reported the opposite pattern: that prior knowledge is more strongly associated with overnight change following consolidation. For example, Henderson et al. (2015) reported that children's standardised vocabulary scores were positively related to overnight improvement in the recall of novel word-forms. In line with this, Henderson and James (2018) taught 10- to 11-year-old children novel words (e.g., *crocodol*) embedded in spoken stories. When tested with a stem completion task ("which word began with cro-?"), children with a larger English vocabulary showed greater overnight improvement in their knowledge of the new word-forms than children with a smaller vocabulary. In addition to this, Zion et al. (2019) reported

a positive association between individuals' phonological and morphological awareness on a standardised phoneme and word detection task and offline improvements in inflectional morphology in an artificial language. Taken together, these findings suggest an association between prior knowledge and offline consolidation, which is more consistent with the predictions outlined in the iOtA model: that prior knowledge may facilitate the offline integration of overlapping memories.

To summarise, in line with CLS and SLIMM, studies that have explored the influence of local prior knowledge (e.g., James et al., 2019) have found a benefit of prior knowledge immediately after learning and no influence of offline consolidation. In contrast, findings from studies on global prior knowledge and language learning are more in line with iOtA as these studies have reported a later emerging prior knowledge benefit, occurring over a period of offline consolidation (e.g., Henderson et al., 2015).

1.6 Summary and Thesis Overview

This review outlined broad theoretical perspectives and empirical findings on the memory mechanisms underpinning language learning and consolidation, with a particular focus on vocabulary and grammar viewed through the lens of arbitrariness and systematicity. Two key neurocognitive models were considered: the CLS (McClelland et al., 1995) and DP (Ullman, 2001; 2004) frameworks, and the predictions each makes regarding the neural mechanisms underlying the acquisition of arbitrary and systematic mappings. Additionally, this review considered similarities and differences in the evidence for sleep-related offline consolidation in vocabulary and grammar learning, examined primarily through the CLS and systems consolidation theory. More recent theoretical and empirical work on the role of prior knowledge was also considered. Specifically, this review outlined the predictions of the CLS, SLIMM, and iOtA models of memory consolidation, and relevant empirical findings suggesting that the relationship between existing knowledge and new knowledge influences encoding and, consequently, the role of offline consolidation.

Against this background, the key focus of this thesis is on the role of prior knowledge in grammar learning. Specifically, the key research question this thesis aims to address is as follows: What is the role of prior knowledge in grammar learning and to what extent might it influence initial learning and offline consolidation? To address this research question, across four experiments, we trained participants on an artificial language consisting of novel form-meaning mappings with both arbitrary and systematic components. In doing so, this thesis focuses on the influence of local prior knowledge, that is, the extent to which the systematic mappings were related or unrelated to learners' existing linguistic knowledge, rather than on global prior knowledge such as individual differences in learners' vocabulary size or linguistic knowledge. In each experiment, we employed a range of behavioural measures to track participants' learning of the arbitrary components as well as their learning and generalisation of the systematic components of the language.

Following this chapter is the first empirical chapter, Chapter 2, the key aim of which was to examine the contribution of offline processes over sleep and wake to the consolidation of novel systematic mappings that were either related or unrelated to participants' existing grammatical knowledge. Also of interest was how prior knowledge would influence the initial learning and generalisation of the systematic mappings. To address these questions, participants were trained on the artificial language either in the evening or in the morning and tested immediately, and then again after a 12-hour interval including sleep or wake, and then again 1 week later, assessing long-term consolidation. Recall, recognition, and online language use were measured at each time point to test learning of trained regularities and generalisation of the trained regularities to new items. In Experiment 2 and 3, Chapter 3 of this thesis, the same artificial language from Experiment 1 was used to further examine the role of prior knowledge in the initial learning and generalisation of the systematic mappings. Additionally, we explored the possibility of an explicit knowledge contribution to the initial learning of the prior knowledge related systematic mappings. The final empirical chapter, Chapter 4, presents an exploratory study designed to examine the contribution of implicit and explicit processes to the learning of the systematic mappings using an artificial language that more closely matched the complexity of natural languages.

Chapter 2. The Influence of Offline Consolidation (Sleep versus Wake) and Prior Knowledge on Word and Grammar Learning

2.1. Introduction

2.1.1. Encoding and Consolidation of Language Knowledge

An expanding body of literature has highlighted a central role for consolidation mechanisms in the long-term retention of new linguistic knowledge and crucially, emerging research is now beginning to uncover the specific cognitive and neural processes that contribute to and influence different aspects of language learning. For instance, the stabilisation and strengthening of newly learned words has been causally linked to the process of hippocampal replay that occurs during sleep (e.g., Batterink et al., 2018; Schreiner & Rasch, 2015). Sleep-related consolidation has also been implicated in grammar learning and generalisation- the ability to abstract regularities from specific examples and apply them to novel instances – in both children and adults (Batterink et al., 2014; Batterink & Paller, 2017; Friedrich, et al., 2015; Nieuwenhuis et al., 2013; Tamminen et al., 2012). However, not all studies have replicated the finding of a contribution of sleep to grammar learning (e.g., Mirković & Gaskell, 2016).

The Complementary Learning Systems (CLS) model (McClelland et al., 1995; McClelland, 2013) provides a useful framework for understanding the relative contributions of specific neural systems to the long-term retention of new linguistic knowledge. Davis and Gaskell (2009) applied the principles of CLS to word learning, suggesting that two structures, the hippocampus and neocortex, are recruited during the encoding of new words. Further, interaction between the neocortical and hippocampal structures is thought to play a crucial role in the integration of new words into the mental lexicon, providing a lasting basis for retention. Specifically, the CLS applied to word learning predicts that new words are initially more reliant on the hippocampus as this system uses pattern separation to encode new representations as distinct episodic traces and thereby prevents interference with existing word knowledge within the neocortex. It is then through processes such as hippocampal replay, that have been observed during post-learning sleep, that these newly encoded memories are re-activated. This reactivation is then thought to prime related neocortical representations and facilitate the integration of new and existing knowledge.

Mirković and Gaskell (2016) focused on the encoding schemes utilised by the hippocampus and neocortex to further examine the extent to which CLS applies to language learning and how it can be used to account for initial acquisition behaviour. The authors hypothesised that the initial reliance of arbitrary knowledge (i.e., phonological form-to-meaning mappings as in word learning) on the hippocampus would mean that it is more likely to be influenced by subsequent offline consolidation opportunities such as sleep, of which hippocampal re-activation is a hallmark (Born & Wilhelm, 2012). In contrast, when new information is more systematic like regularities in grammar (e.g., in English, nouns are pluralised by adding the morpheme '-s' to indicate that there is more than one something), it should be better learned via the neocortex as this system is thought to use an overlapping, distributed encoding scheme to allow the extraction of shared features across experiences (Alvarez & Squire, 1994; O'Reilly & Rudy, 2001). Thus, in comparison to arbitrary knowledge, the greater reliance of new systematic knowledge in the neocortex would mean it is less likely to be influenced by hippocampal re-activation during sleep.

Mirković and Gaskell (2016) tested this hypothesis using an artificial language that incorporated both arbitrary and systematic mappings. Stem-referent mappings (e.g., *scoiff* in tib *scoiff*esh = ballerina) provided the arbitrary component (as in vocabulary learning) while determiner/suffix mappings (e.g., *tib…esh/eem* and female; *ked…ool/aff* and male) provided the systematic (grammatical) components. As with grammatical morphemes in natural languages, the systematic determiner and suffix mappings served as cues to the gender of the depicted referent, similar to grammatical gender systems in natural languages. Each determiner was consistently presented with either a picture depicting a female or male character (e.g., *tib* = feminine, *ked* = masculine). The suffixes provided a second cue to gender, however, instead of the one-to-one mapping between the determiner and gender, there were two suffix cues associated with each gender (e.g., *esh/eem* = feminine, *aff/ool* = masculine). Training on the arbitrary and systematic components of the novel words was administered using a word-picture matching task after which memory for the words was tested after a 2-hour delay, including wake or sleep. Those who had a two-hour nap saw more considerable gains in the consolidation of arbitrary mappings (i.e., individual vocabulary items), compared to those who remained awake. Conversely, there was no evidence of a beneficial effect of sleep on the generalisation of systematic mappings which contrasts with previous findings (e.g., Batterink et al., 2014). This study was the first to assess the influence of sleep on the consolidation of a fully-fledged artificial language that combined both arbitrary and systematic mappings. Thus, it was suggested that both types of mappings competed for consolidation, with arbitrary mappings being prioritised during the post-learning sleep opportunity because of their greater reliance on the hippocampus at encoding. This may explain why some previous research that has examined novel grammar learning in isolation (i.e., not in parallel to word learning) has found a beneficial effect of sleep on generalisation (e.g., Batterink et al., 2017; Tamminen at el., 2012).

However, one other possible explanation for the inconsistent findings regarding the role of sleep in grammar learning is that whether a sleep effect is observed or not may be dependent on the generalisation test used. For instance, Tamminen et al. (2012) trained participants on a set of novel affixes (e.g., -*nule)* attached to existing words. Participants were instructed to learn the novel affixes along with their meaning (e.g., a *'buildnule',* is someone who is able to build furniture at remarkable speed; a '*climbnule*', is a person who climbs mountains with dangerous peaks). Of interest was whether and when the participants would demonstrate an ability to generalise the novel affixes presented in training to untrained words (e.g., '*sailnule')*. Tamminen et al. (2012) used a speeded and non-speeded task to assess generalisation immediately after training and two days later. Interestingly, participants were able to generalise the knowledge of the affix meanings in a non-speeded task immediately after learning but in a speeded task generalisation to untrained words only emerged at a delayed test two days later. Based on these findings, the authors suggested that the two tasks (speeded vs. non-speeded) likely reflect different forms of generalisation. Specifically, Tamminen et al. (2012; 2015) proposed that since generalisation in non-speeded tasks affords participants the opportunity to reflect on their knowledge with minimal time pressure, this form of generalisation can arise on the basis of episodic hippocampal representations (e.g., by combining multiple episodes at the time of testing). However, due to the limited storage capacity of the hippocampus, generalisation in this form is only likely to be supported over limited timescales and in situations in which all relevant information is hippocampally represented (Kumaran & McClelland, 2012). In contrast, the delayed emergence of generalisation in speeded tasks suggests it is supported by neocortical representations that develop gradually over time through consolidation. This is synonymous with neocortical learning which adopts a slow learning rate to integrate new information within existing cortical networks (McClelland et al., 1995). Thus, it is possible that two forms of generalisation arise during language learning and are supported by two different representations: one that is based on episodic memory and is present immediately after learning but for a limited time, and the other that is supported by overlapping neocortical representations and hence is gradual and is only available once consolidation has taken place.

Although Tamminen et al. (2012) did not isolate the specific effects of sleep on learning and memory, they applied the CLS framework (McClelland et al., 1995) to suggest how sleep might specifically facilitate the formation of abstract, context-independent representations necessary for generalisation in speeded tasks. By the CLS account, the hippocampus initially encodes new linguistic information as distinct, context-dependent episodic memories, thereby preventing interference to existing neocortical knowledge. Hippocampal reactivation during sleep then facilitates the gradual integration of these recent episodic memories into existing neocortical networks. Because the neocortex encodes new information in an overlapping, distributed manner, it naturally emphasises shared properties and commonalities across multiple episodes. In doing so, the neocortical system gradually shifts representations away from context-specific details and toward more abstract, context-independent features. This process enables the extraction of regularities across initially separate episodic experiences, thus providing a powerful basis for flexible generalisation to novel contexts. Tamminen et al. (2012) therefore proposed that sleep-driven consolidation processes might explain the delayed emergence of generalisation observed in their speeded tasks.

Nonetheless, research beyond grammar learning has also observed continued improvements in generalisation over time, though using non-speeded tasks. Importantly, this line of research suggests that sleep-related consolidation may not be the only offline mechanism capable of supporting the abstraction of new knowledge. Rather, sleep- and wake-related consolidation processes may each independently facilitate the emergence of context-independent representations necessary for generalisation, albeit via different mechanisms. For example, Sweegers and Talamini (2014) assessed learning and generalisation of face-location associations. In their study, for half of the associations, certain combinations of facial features were predictive of locations (e.g., slender faces with headwear appeared in a particular location). Conversely, the other half of associations were randomly assigned to a location, thus they were arbitrary in that these associations did not predict a particular location. Participants were assigned either to a nap group (two-hour nap), a wake group (quiet wakefulness), or an immediate-test control group. For the nap and wake group, generalisation was tested 4-hours after learning and again several weeks later. The authors found that generalisation performance was present immediately after learning, increased across the 4-hours and continued to improve across a period of several weeks.

Crucially, Sweegers and Talamini (2014) found no differential effects of sleep versus wakefulness on generalisation performance, suggesting that both may similarly contribute to post-learning improvements in generalisation, though through different mechanisms. For example, as outlined above, one possibility is that offline processes during sleep facilitate the formation of abstract, context-independent representations. On the other hand, Sweegers and Talamini (2014) suggested that wakefulness may similarly support the emergence of generalisation through a different mechanism, namely interference-based memory decay. Specifically, they suggest that interference-based memory decay during wakefulness could contribute to regularity extraction by eroding arbitrary episodic details, while preserving the more regular, systematic associations. Thus, in addition to sleep-related consolidation processes, the loss of irrelevant, context-specific details during wakefulness may enhance the context-independent features of representations relevant for generalisation.

Additional support for the beneficial role of wake-related forgetting comes from Werchan and Gomez (2015), who investigated the influence of sleep versus wakefulness on word learning and generalisation. In their study, 2.5-year-old children learned labels for three novel object categories, each presented across varied contexts (e.g., differing backgrounds and colours). Immediately after training, one group of children napped, another remained awake for a 4-hour interval, and a third group completed an immediate generalisation test. The generalisation test assessed infants' ability to apply the learned labels to novel exemplars presented in new contexts. Interestingly, only the wakefulness group showed successful generalisation at the delayed test, outperforming both the nap and immediate-test groups. Werchan and Gomez (2015) suggested that the relative benefit of wake versus sleep-related consolidation in their study was due to the way infants encode information. For example, infants, unlike adult learners, tend to encode both relevant (e.g., object shape, label) and irrelevant details (e.g., background colour). Consequently, the stabilisation of new memories during sleep preserves these irrelevant contextual details, thereby hindering generalisation. In contrast, wake-related forgetting facilitated generalisation by enabling the infants to discard irrelevant contextual details, enhancing abstraction and thus generalisation to novel contexts. Werchan and Gomez (2015) therefore argue that, compared with adults, wake-related forgetting may be particularly beneficial for generalisation in young children. Nonetheless, these findings align closely with those of Sweegers and Talamini (2014), highlighting a beneficial role for wake-related forgetting in generalisation.

However, wake-related forgetting may not always facilitate the abstraction of new knowledge. If systematic knowledge is more dependent on the neocortex at encoding (McClelland et al., 1995; 2013; Mirković & Gaskell, 2016), then it might be particularly vulnerable to interference-based forgetting (Sadeh et al., 2014). For instance, within the neocortex, similar representations are thought to be encoded in overlapping, distributed neural networks (O'Reilly & Norman, 2002) but the overlapping nature of these representations may make them susceptible to interference from related feature-specific information. In contrast, hippocampal representations tend to be sparse and non-overlapping, due to the hippocampus' capacity for pattern separation (Bakker et al., 2008; Berron et al., 2016). As a result, the greater reliance of systematic representations on the neocortex may make them more vulnerable to interference-based forgetting compared to hippocampal-dependent representations.

This greater susceptibility of neocortical representations to interference-based forgetting may explain why some studies have not observed beneficial effects of wakefulness on the generalisation of systematic knowledge. For example, a study by Mirković et al. (2019) explored the role of sleep and wake consolidation in learning a novel language with a quasi-regular morphological system that mimicked the English past tense (e.g., dominance of regular, systematic forms such as walk-walked or jump-jumped alongside a smaller number of high-frequency irregulars such as sit-sat or go-went). Performance was assessed based on the recall of trained items and critically, the ability to generalise the morphological patterns to previously unseen items. For the generalisation task, participants were presented with previously unseen stems (uninflected word-form) and were required to produce the inflected form within the appropriate linguistic context. Testing took place either immediately after learning, 12 hours later with or without sleep or 24 hours later. The authors reported a strengthening of irregular forms and weakening of regular (more systematic) forms at delay periods in both the sleep and wake conditions. Additional analyses revealed less evidence of forgetting of trained irregular forms in the sleep and greater forgetting for trained regular, systematic forms over the wake period. Mirković et al. (2019) suggested that *irregular* forms were more likely strengthened during sleep due to the greater reliance of the more arbitrary irregular forms on the hippocampus at encoding. Conversely, the more systematic *regular* items that were initially more strongly reliant on the neocortical system may have been subject to stronger interference based-forgetting (Sadeh et al., 2014). Thus, the strengthening effects of sleep on arbitrary knowledge combined with the increased susceptibility to interference for the systematic forms, led to the same outcome for both the wake and sleep groups: a greater influence of the irregular trained items when generalising to novel forms. Taken together, while some evidence suggests that wake-related forgetting can support the abstraction and generalisation of systematic knowledge (e.g., Sweegers & Talamini, 2014; Werchan & Gomez, 2014), findings from Mirković et al. (2019) indicate this may not always be the case. Crucially, the greater reliance of systematic representations on the neocortical system at encoding means that it is not only less likely to benefit from hippocampal reactivation during sleep but that it is also more vulnerable to interference based-forgetting during wake. Conversely, arbitrary representations, with their greater initial dependence on hippocampal encoding, are more likely to benefit from hippocampal reactivation during sleep and, relative to neocortically dependent systematic representations, be less susceptible to interference-based forgetting during wake.

2.1.2. The Role of Prior Knowledge in Language Learning and Consolidation

In addition to the distinctions between arbitrary and systematic knowledge, a further factor that may influence the learning and consolidation of new linguistic knowledge is the extent to which new information relates to existing knowledge. According to CLS (Kumaran et al., 2016; McClelland, 2013; McClelland et al., 2020), new experiences are initially encoded by the hippocampus and gradually integrated into existing neocortical representations for long-term retention. However, the rate at which this integration occurs is thought to depend on how consistent the new information is with existing neocortical knowledge. When new information is inconsistent or unrelated to existing knowledge, it poses a greater risk of interference to established neocortical representations. Therefore, to minimise this risk of interference, neocortical integration is assumed to proceed more slowly through interleaved learning. This is the process by which the new, unrelated information is repeatedly reactivated and interleaved with existing knowledge during hippocampal reactivation. By contrast, new knowledge that is consistent with existing knowledge presents less risk of interference and can therefore be integrated more rapidly. Specifically, because prior knowledge is already represented in neocortical networks, new information that aligns closely with it may be integrated more rapidly, potentially reducing the need for extensive hippocampal reactivation to occur offline (Kumaran et al., 2016).

Although the underlying neural mechanisms are not entirely clear in the CLS model, Van Kesteren et al. (2012) proposed a framework called SLIMM (Schema-Linked Interactions between Medial prefrontal and Medial temporal regions), which provides a more explicit account of how the relationship between new and existing knowledge may influence encoding and consequently, consolidation. According to SLIMM, the medial prefrontal cortex (mPFC) assesses congruency between new and existing schematic knowledge at encoding. When new congruent information is encountered, prior knowledge (i.e. an existing schema) becomes activated. This activation is then detected by the mPFC, which inhibits the MTL/hippocampal system and facilitates the encoding of the new, congruent information directly within existing neocortical knowledge. However, when new information is less congruent with existing knowledge, the mPFC's inhibition of the hippocampal system is reduced, resulting in greater engagement of this system during encoding. The greater engagement of the hippocampus during the encoding of less congruent information means that sleep-related consolidation may be more likely to play a role, given that sleep is thought to support the consolidation of MTL-based representations via hippocampal replay (Wilson & McNaughton, 1994). Taken together, both the CLS and SLIMM accounts predict that new prior knowledge related information is integrated directly within the neocortex during learning and consequently less dependent on hippocampal encoding, and thus less susceptible to sleep-related consolidation, than information that is unrelated to prior knowledge.

Support for the predictions outlined in CLS and SLIMM has been observed within the word-learning literature. For example, James et al. (2019) examined the influence of prior knowledge on the initial learning and consolidation of novel words by training participants on pronounceable English pseudowords that varied in their similarity to existing English word-forms. Adults and 7–9-year-old children were taught novel words (e.g., *ballow*) that differed in their similarity to existing English word-forms (e.g., *wallow, bellow*), such that some were phonologically distinct from English words (i.e., had no phonological neighbours), whereas others overlapped phonologically with existing English words (i.e., had phonological neighbours). To assess how word-form similarity influenced initial learning and offline consolidation, recall of the novel words was tested immediately after training, the next day, and 1 week later. Based on CLS, the authors hypothesised that words *with* neighbours would benefit from their connections to existing knowledge during learning and therefore show better recall immediately after training, whereas words *without* neighbours would benefit more from an offline consolidation opportunity. In line with their predictions, both children and adults showed significantly better recall for words *with* neighbours than words *without* neighbours immediately after training. By contrast, 1 week later, words *without* neighbours were recalled better, suggesting that they benefited from offline consolidation processes, whereas words *with* neighbours showed no improvement over the same interval. James et al. (2019) interpreted this pattern of findings within the CLS framework, suggesting that word-forms with no connections to prior knowledge, because of their greater reliance on hippocampal encoding, benefited more from an offline consolidation interval. By contrast, word-forms with connections to prior knowledge likely engaged the neocortex immediately during learning, reducing the need for further integration offline.

One way through which prior knowledge may benefit learning immediately is through learners being able to intentionally draw links between their existing knowledge and new, related knowledge during learning. In a subsequent study, James et al. (2021) hypothesised that intentional learning strategies may have driven the prior knowledge benefits observed early on in learning in the study presented above. In the James et al. (2019) study, direct vocabulary instruction was used to teach participants the novel words, which involved participants repeating the words aloud and two rounds of a multiple-choice quiz. In the first round of the quiz, one of the novel objects was presented on the screen and participants had to select which word out of three was its name. The second round presented one of the novel words with three of the novel objects to choose from. Afterwards, correct answers were provided. This explicit training on the novel words likely encouraged learners to strategically draw links between the novel word-forms and similar known word-forms, potentially leading to greater strengthening of these connections at encoding and reducing the need for further strengthening offline. However, when learners are unable to explicitly capitalise on their existing knowledge, connections to prior knowledge may be weaker and consequently may be more likely to strengthen when there is opportunity for offline consolidation.

James et al. (2021) tested this hypothesis by examining the contributions of prior knowledge following incidental (versus intentional) learning of novel words. In this study, adults and children were presented with novel words embedded within an illustrated spoken story. As in the study by James et al. (2019), the novel words varied in their connections to prior knowledge such that they were either phonologically similar or distinct from English word-forms. Recall and recognition of the words was tested immediately, the next day and 1 week later. In contrast to James et al. (2019), neither children nor adults showed an immediate benefit for recalling words that were similar to English word-forms, suggesting that prior knowledge may be less helpful under incidental learning conditions, when there is less opportunity for learners to capitalise on their existing knowledge. Interestingly, no improvement in recall performance was observed for words with neighbours 1 week later, suggesting that even when connections to prior knowledge are weak at encoding, they do not benefit from an offline consolidation opportunity. Taken together, the findings reported in James et al. (2019) and James et al. (2021) suggest that prior knowledge benefits the acquisition of new word-forms immediately, during training but only when learners are able to form direct connections between their existing knowledge and new, related information.

However, prior knowledge may not always be helpful, with some evidence suggesting that related semantic knowledge may interfere with the learning of new word-forms and their meanings. In a study by James et al. (2023), adults and children aged 7-10-years were taught novel words assigned to novel concepts with either sparse (i.e. fewer links to related existing concepts) or dense (more links to related existing concepts) connections to existing semantic knowledge. Their knowledge of the words (form recall, definition recall and semantic categorisation) was tested immediately after training, the next day and 1 week later.
At immediate test, both children and adults performed more poorly for words linked to concepts with denser semantic connections than for words linked to concepts with sparser semantic connections. Moreover, the interference that participants experienced during learning was not remediated by an offline consolidation opportunity, as no improvements in performance were observed after 1 week. James et al. (2023) suggested that, for words with denser connections to existing knowledge, the co-activation of related concepts at encoding may have interfered with the establishment of new semantic representations. In line with this, other research that has examined the role of semantic prior knowledge in new word learning in adults and pre-school children has also observed poorer performance for novel words with denser connections to existing semantic knowledge (Storkel & Adlof, 2009; Tamminen et al., 2013). Thus, unlike the benefits observed for novel words with phonological neighbours (James et al., 2019), related semantic knowledge appears to interfere with the learning of new words.

Overall, the experimental evidence outlined above indicates that when prior knowledge is influential during initial learning (e.g., James et al., 2019) its benefit is observed immediately after learning and is not further influenced by offline consolidation. Similarly, even when prior knowledge connections are only weakly encoded (e.g., during incidental learning conditions), or when they interfere with initial learning (e.g., when learning new words in dense semantic neighbourhoods) they do not benefit from additional strengthening offline. These findings are consistent with the predictions of CLS and SLIMM that, when new knowledge is related to existing knowledge, it is less reliant on hippocampal encoding and may therefore be less likely to be influenced by offline consolidation processes.

2.1.3. The Current Experiment

The theoretical frameworks and evidence outlined above suggest that the contribution of offline consolidation mechanisms to the long-term retention of new linguistic representations depends on both the nature of the representation (arbitrary vs. systematic) and its relationship with existing knowledge (related vs. unrelated). For instance, research guided by the CLS framework (Mirković & Gaskell, 2016) suggests that new representations that are more arbitrary (such as phonological form-meaning mappings in word learning) should be preferentially encoded by the hippocampus and, consequently, more strongly influenced by hippocampal reactivation during post-learning sleep. Conversely, the neocortical system should be better suited to encoding systematic representations (such as regularities in grammar) because this system employs overlapping, distributed representations that can facilitate the extraction of shared features across experiences. Thus, the greater reliance of systematic knowledge on the neocortex at encoding would mean it is less influenced by hippocampal reactivation during sleep. In line with these predictions, Mirković and Gaskell

(2016) observed a beneficial effect of sleep on the consolidation of arbitrary mappings but found no such effect on systematic mappings. The benefit of sleep for learning new arbitrary mappings is consistent with a body of evidence demonstrating the role of sleep in new word learning (e.g., Schimke et al., 2021). However, contrasting findings have been reported regarding systematic mappings. For example, Tamminen et al. (2012; 2015) observed improvements in the generalisation of systematic representations following an offline consolidation period, suggesting that systematic knowledge may also benefit from sleep-dependent consolidation.

As suggested above, one potential explanation for these conflicting findings is the nature of the tasks used to assess generalisation. For example, Mirković and Gaskell (2016) used a non-speeded task to measure the generalisation of systematic mappings in their study. However, Tamminen et al. (2012) showed that generalisation in non-speeded tasks is available immediately after learning since this form of generalisation relies primarily on episodic, hippocampal representations (e.g., combining multiple episodes at the time of testing). In contrast, generalisation assessed by speeded tasks is more likely supported by overlapping neocortical representations and therefore might only emerge once neocortical integration has taken place (e.g., via offline consolidation processes during sleep). Therefore, the conflicting findings between these studies could arise from differences in task sensitivity to sleep-related consolidation effects.

Additionally, previous research examining the contribution of sleep-related consolidation to the long-term retention and generalisation of systematic mappings has not considered the role of prior knowledge. As outlined above, whether offline processes such as sleep contribute to long-term retention also depends on the relationship between new representations and existing knowledge. According to both the CLS (McClelland, 2013) and SLIMM (VanKesteren et al., 2012), prior knowledge-related information can be integrated efficiently into existing neocortical networks, resulting in comparatively limited hippocampal involvement during encoding and consequently reduced susceptibility to hippocampal reactivation during sleep, relative to information that is unrelated to or incongruent with prior knowledge. Consistent with the predictions outlined by CLS and SLIMM, experimental research within the word-learning literature suggests when a prior knowledge benefit is observed such as in the acquisition of novel word-forms (e.g., James et al., 2019), it emerges immediately after learning and is not further influenced by offline consolidation. This is in contrast to novel word-forms with no connections to local prior knowledge, which show further strengthening after an offline consolidation period.

In light of the above, it is also worth considering whether the lack of sleep-related consolidation benefits for systematic mappings reported by Mirković and Gaskell (2016) might be influenced by the relationship between new and existing knowledge. For instance, in their study, the systematic mappings served as cues as to the gender of the depicted referent similar to how the English language encodes gender in pronouns (e.g., she/her and he/him). In line with cortical learning accounts of CLS, and the SLIMM framework, this overlap with existing knowledge may have facilitated neocortical integration and consequently reduced the influence of offline consolidation. Therefore, to further understand the contribution of sleep-related consolidation to the long-term retention and generalisation of systematic knowledge, it is important to directly compare how offline processes (e.g., sleep) influence systematic mappings that are related versus unrelated to learners' existing linguistic knowledge.

Finally, in addition to examining how the relatedness of new linguistic knowledge to existing representations influences the contribution of offline consolidation processes, it is also important to establish whether prior knowledge is helpful for the initial encoding of new grammatical knowledge. Although experimental research by James et al. (2019) has shown that prior knowledge can support the learning of novel word-forms, some research has reported that related semantic knowledge interferes with the acquisition of novel words and their meanings (James et al., 2023; Tamminen et al., 2013). However, it remains to be explored whether prior knowledge helps or hinders the learning of more systematic types of knowledge such as regularities in grammar.

Taking the above into consideration, the aim of the current experiment was to examine the influence of prior knowledge on the learning and consolidation of novel systematic mappings, and in doing so, further elucidate the relative contributions of sleep- and wake-related consolidation to grammar learning. Previous research examining the role of offline consolidation mechanisms in more systematic types of knowledge such as grammar learning have not directly assessed the role of existing linguistic knowledge and how it might influence learning outcomes. For instance, studies reporting no sleep-related consolidation benefits for systematic mappings have typically implemented artificial grammars with novel mappings closely overlapping learners' existing grammatical knowledge (e.g., Mirković & Gaskell, 2016). Conversely, studies that have reported a role for sleep-related consolidation in grammar learning and generalisation have implemented artificial grammars with existing words (e.g., Batterink et al., 2014; Tamminen et al., 2012).

To disentangle these conflicting findings and to assess the influence of prior knowledge on the learning and consolidation of systematic knowledge, the current study employed an artificial language with varying degrees of arbitrariness and systematicity in its sound-meaning mappings. We incorporated two types of novel systematic mappings, prior knowledge dependent and prior knowledge independent, within a novel grammatical system similar to that used by Mirković and Gaskell (2016). The arbitrary components of the artificial language were reflected in the individual stem-meaning mappings (e.g., *zeap*-mermaid, *garr*-builder), and the systematic components in the relationship between determiners and suffixes and common semantic features across groups of items. Specifically, for the prior knowledge dependent systematic mappings, pictures depicting female and male characters were assigned to the *"ked -ool"* and *"tib -eem"* words respectively. Thus, they are consistent with participants' prior linguistic knowledge in that the semantic category assignment of female and male characters was encoded in phonological properties of the novel words in the same way the pronouns "he" and "she" encode sex in animate referents in English. In contrast, the prior knowledge independent systematic mappings were not related to existing linguistic knowledge in that the *"ked -ool"* and *"tib –eem"* words were paired with pictures of sea animals and insects respectively, i.e. semantic categories that are not related to any existing grammatical categories in English.

To assess the contribution of sleep and wake to the learning and long-term retention of the arbitrary and systematic mappings outlined above, participants were assigned to either a sleep-first or wake-first condition. Those in the sleep-first group were trained on the artificial language in the evening, tested immediately after learning and then again 12-hours later (the following morning) whereas those in the wake-first group were trained in the morning, tested immediately after learning and then again 12 hours later (that same evening). A 3rd test point was also included to assess the influence of consolidation in both groups after 1 week.

The learning of the systematic components of the novel words (determiners and suffixes) was assessed using both the trained items and by assessing participants' ability to generalise the systematic knowledge to previously unseen picture-word pairs. Generalisation tests offer a purer measure of systematic knowledge, as successful performance cannot rely solely on memory for specific, previously learned items (i.e., individual item knowledge). Furthermore, given that speeded generalisation tests have been reported to be more sensitive to offline consolidation effects (Tamminen et al., 2012; 2015), a combination of speeded and non-speeded tasks was used. First, a speeded shadowing task, similar to the one used in Tamminen et al. (2012), was administered to assess speeded responses to the trained items and generalisation to untrained items. Second, for the non-speeded task, a picture naming task was administered to measure recall accuracy of the mappings. This picture naming task was also used to assess recall accuracy of the arbitrary components of the language (i.e., the individual form-meaning mappings), alongside a recognition task, which provided an additional measure of participants' learning and consolidation of the arbitrary mappings.

Using the paradigm and tasks outlined above, the first aim of the current experiment was to examine how offline processes during sleep versus wakefulness contribute to the consolidation of systematic mappings that were either related or unrelated to participants' existing linguistic knowledge. In line with predictions outlined in the CLS (McClelland, 2013) and SLIMM (Van Kesteren et al., 2012) models, we hypothesised that if sleep does play a role in consolidating the systematic mappings, we would only see evidence of a contribution in the consolidation of mappings unrelated to participants' existing knowledge since novel information unrelated to existing knowledge is thought to rely more heavily on the hippocampus at encoding. Furthermore, if sleep does play a role in consolidating these mappings, its contribution may be more clearly reflected in the speeded tasks, since performance on these tasks is more likely to rely on neocortical representations that only become available once hippocampal-neocortical integration has taken place, such as during post-learning sleep (Tamminen et al., 2012; 2015). Conversely, for the mappings related to prior knowledge, we hypothesised that their connections to existing knowledge would facilitate rapid neocortical integration during encoding, leading to better immediate learning compared to mappings independent of prior knowledge. However, this stronger neocortical involvement during learning would reduce hippocampal engagement, making these mappings less likely to benefit from hippocampal reactivation during sleep.

In addition to the systematic mappings, we also examined the contribution of offline consolidation processes during sleep and wakefulness to the arbitrary components of the language, specifically the stem-meaning mappings. In line with the predictions by Mirković and Gaskell (2016), we hypothesised that memory for the arbitrary components would be better when encoding was followed by sleep as opposed to wakefulness, due to the greater reliance of arbitrary knowledge on the hippocampus at encoding.

To summarise, the key aims of the current experiment were:

1) To examine the contributions of offline processes during sleep versus wake to the consolidation of two types of systematic mappings: those that are related to existing knowledge versus those that are less related to existing knowledge (using speeded and non-speeded tasks).
2) To examine the contributions of offline processes during sleep versus wake to the consolidation of the arbitrary mappings.

2.2 Methods

2.2.1. Participants

60 adult participants aged between 18 and 40 consented to participating in the study for course credit or monetary payment. Of the 60, 31 participants (Wake First = 16; Sleep First = 15) were recruited from the York St John University undergraduate population and 29 participants (Wake First = 14; Sleep First = 15) were recruited from the online recruitment platform Prolific. The target sample size (N = 60) was set a priori based on sample sizes used in previous studies that have employed a similar paradigm to the current study (Mirković & Gaskell, 2016).

All participants were monolingual, native speakers of English with no self-reported history of language, hearing, visual difficulties, psychiatric or sleep disorders. Ethical approval for the study was obtained from the York St John University Cross-School Ethics Committee.

2.2.2. Materials

2.2.2.1. Artificial Language

Training set. The training set consisted of 32 pronounceable English pseudowords designed using the ARC database (Rastle et al., 2002). These items contained 32 unique stems, each of which was combined with one of two suffixes (*-eem* or *-ool*). All words were then assigned to either one of two determiners, a *'ked'* or *'tib'* determiner with half of the words corresponding to the *'ked'* determiner and the other half corresponding to the *'tib'* determiner. Words ending in *'ool'* were assigned the *'ked'* determiner and those words ending in *'eem'* were assigned the *'tib*' determiner. All novel words were digitally recorded in a sound-proof booth by a native English speaker. All words were paired with pictures selected from the commercially available ClipArt image database (see below for more details). There was no phonological overlap between the English word for the depicted character/animal and the novel word.

To assess the influence of prior knowledge, two types of grammatical mappings were incorporated into the artificial language: prior knowledge dependent and prior knowledge independent (see Table 2.1 for examples). For the prior knowledge dependent mappings, there were 16 exemplars (8 *"tib -eem"* and 8 *"ked -ool"* words). The words were presented with pictures depicting stereotypical female or male characters, selected from those used in the study by Mirković and Gaskell (2016). Pictures depicting female characters were paired with *"tib -eem"* words and pictures depicting male characters were paired with *"ked -ool"* words. For prior knowledge independent mappings, there were 16 exemplars also (8 *"tib - eem"* and 8 *"ked -ool"* words). For this category, the words were presented with pictures depicting sea animals or insects and were newly sourced by the experimenter from the ClipArt database. Pictures depicting sea animals were assigned to the *"ked -ool"* category and pictures depicting insects were assigned to the *"tib -eem"* category.

Generalisation set. An additional 24 items were designed to assess participants' ability to generalise the mappings acquired in training to previously unseen items. All generalisation items were consistent with the regularities presented during training in that the mapping between the determiner, suffix and semantic category represented by the picture (e.g., female/male characters, insects/sea animals) remained the same. As with the training set, half of the items in this set comprised the prior knowledge dependent mappings and half comprised the prior knowledge independent mappings.

The generalisation items were divided into 3 lists so that participants were presented with a new set of items at each test session.

Table 2.1 Example items from prior knowledge dependent and prior knowledge independent mappings and their assigned semantic category.

Mapping	Semantic Category	Determiner	Stem + Suffix	Picture
Prior Knowledge Dependent:	female character	*tib*	viff*eem*	*waitress*
	male character	**ked**	torm**ool**	**mechanic**
Prior Knowledge Independent:	insects	*tib*	shem*eem*	*bee*
	sea animals	**ked**	gesh**ool**	**jellyfish**

Note. Cues associated with *tib* words are marked in *italics* and *ked* in **bold**.

2.3. Procedure

2.3.1. Testing Schedule

Participants were trained on the artificial language either in the morning or in the evening and then tested at two time points: 12 hours and 1 week after training. As depicted in Figure 2.1, the procedure was identical for all participants, except that the Wake-First group completed Session 1 (training + immediate test) in the morning and Session 2 (12 hr delay testing) that same evening, and the Sleep-First group completed Session 1 (training + immediate test) in the evening and Session 2 the following morning. Both groups were tested again 1-week later at Session 3 to assess long-term consolidation.

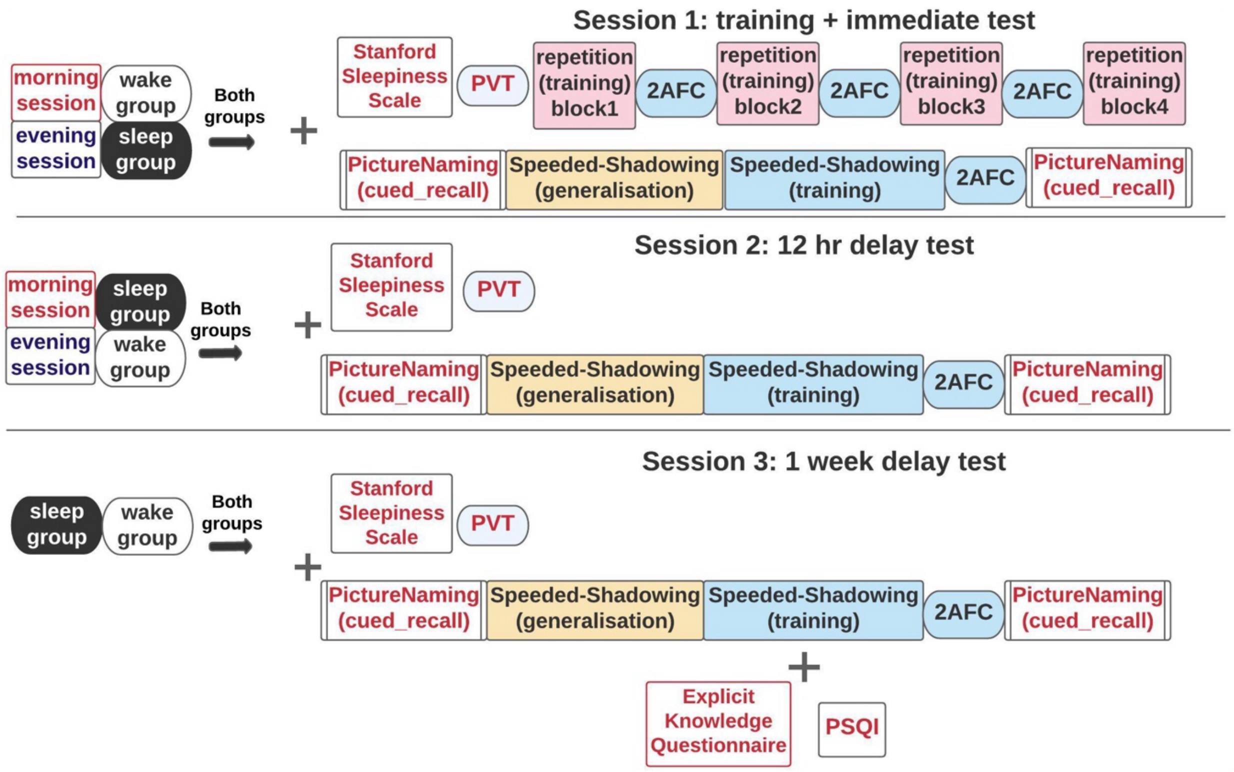

Figure 2.1 Overview of the experimental procedure including the order of tasks in each session. PVT = Psychomotor Vigilance Task; PSQI = Pittsburgh Sleep Quality Index; 2AFC = Two-Alternative Forced Choice.

Each session was scheduled to start at a specific time and with a pre-arranged Microsoft Teams call to ensure that the experiment would not be completed outside of the scheduled time window (between 8AM and 9AM in the morning and 8PM and 9PM in the evening for sessions 1 and 2). The call lasted approximately 5 minutes during which a member of the research team shared the URL link for the relevant session. All sessions were conducted remotely through the Gorilla Experiment Builder platform (Anwyl-Irvine et al., 2018).

At the start of Session 1, participants were presented with an information sheet and asked to provide written informed consent. No reference was made as to the existence of the underlying grammatical mappings at the initial briefing nor at any point during the study.

Before the experimental tasks commenced, participants were required to perform a sound and microphone test and following this, two tests to assess alertness/sleepiness (each task is described in more detail below). These tests were administered at the start of every session. The training phase, administered at the start of Session 1 consisted of 4 blocks of a word repetition task to provide participants with training on the novel words. To sustain engagement, participants completed a 2AFC word recognition task after each training block, with three in total across the training session.

The 3 test tasks comprised of tests assessing the initial learning and retention of the novel words and their systematic mappings, and a speeded shadowing test to assess speeded responses to the trained mappings and generalisation of the trained mappings to new items. The 2AFC word recognition task administered during the training phase of Session 1 was also administered at each test session, to provide a measure of participants' learning and retention of the novel words at each time point. At the end of Session 3, participants completed 2 additional tasks: a questionnaire designed to assess explicit awareness of the underlying systematic mappings and the Pittsburgh Sleep Quality Index (PSQI; Buysse et al., 1989) to assess subjective sleep quality and disturbances over 1 month.

A detailed description of each task is provided below.

2.3.2. Alertness Measures

To control for any circadian effects on task performance, participants' alertness was assessed at the beginning of each session using the Stanford Sleepiness Scale (SSS; Hoddes et al., 1973) and a 3-minute Psychomotor Vigilance Task (PVT; Dignes & Powell, 1985), the tasks typically used for these purposes in sleep studies. The SSS asked participants to rate their sleepiness on a scale of 1–9, with smaller numbers indicating higher alertness (less sleepiness) and larger numbers indicating increased sleepiness (lower alertness). The PVT provided an objective assessment of alertness. For the PVT, in each trial, a fixation cross was presented centrally on the screen. The duration of the fixation cross varied (randomly) on each trial, ranging from 2 to 10 seconds. The fixation cross was then replaced by a '1' and a '0' presented on either side of the screen. The task was to press the left arrow key when the '1' was presented on the left side of the screen and the right arrow key when the '1' was presented on the right side. The number of attentional lapses, defined as missed responses (failing to press a key) or incorrect responses (pressing the wrong key) was used to evaluate differences in alertness between the two groups at the start of each session.

2.3.3. Sleep Quality Measure

At the end of Session 3, all participants were asked to complete the Pittsburgh Sleep Quality Index (PSQI; Buysse et al., 1989) to assess general sleep quality over the last month. This questionnaire comprises 19 self-rated items, which are used to form seven component scores: subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep disturbances, use of sleeping medication and daytime dysfunction. The sum of scores for these seven components yields one global score with a maximum of 21, with higher scores indicating poorer sleep quality.

2.3.4. Language Training

2.3.4.1 Word Repetition

Participants were trained on the new words using a word repetition task. At the start of this task, participants were told that they would see an image and hear a made-up word and that their task was to repeat each word out loud after they had heard it.

Each trial began with a 500 ms centrally located fixation cross, followed by a picture. 500 ms after the picture was presented, the novel word associated with the picture was presented to participants auditorily. After the word had been presented, the text *'repeat the word'* appeared underneath the picture and participants' responses were recorded. Both the picture and the text then remained on the screen for 4000 ms to capture the participants responses before they could proceed to the next trial (see Figure 2.2 for an example trial). The order of presentation of items was randomised across participants. Two practice trials preceded the main task, using items that did not appear in the training set. No feedback was given.

There were 4 blocks of the repetition task, and each word was presented twice in one block. Overall, across the 4 blocks, participants had 8 repetitions of each word.

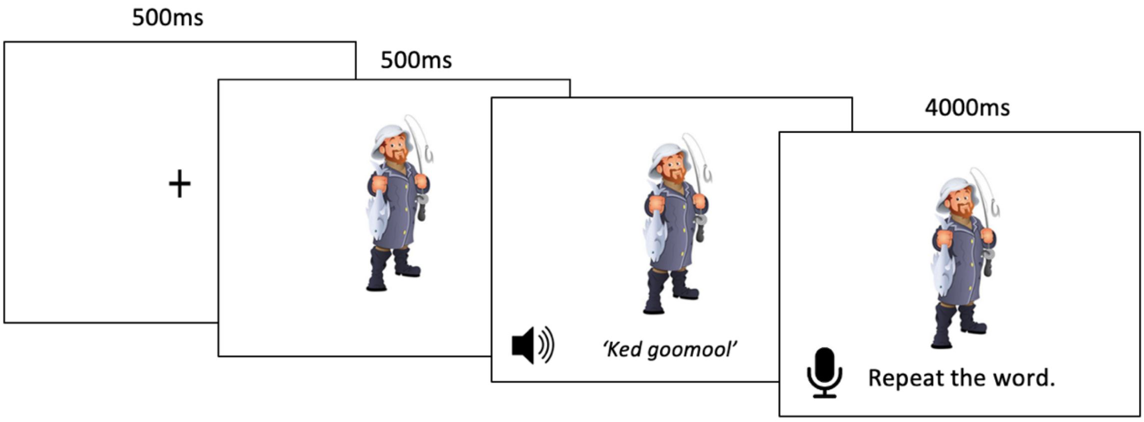

Figure 2.2. Example of a trial from the word repetition task. The novel word (*'ked goomol'*) was presented auditorily only.

2.3.5 Language Tests

2.3.5.1 Two Alternative Forced Choice Recognition (2AFC)

For this task, 2 pictures from training were presented side by side (right and left side of the screen) alongside an auditory presentation of one of the novel words from the artificial language. One of the two presented items was a correct match for the novel word, in that it was consistent with what had been presented during in training (target picture) and the other was an incorrect match (foil picture). Participants were instructed to click on the picture that they thought matched the word that they heard, and accuracy (proportion correct) was used as the outcome measure in this task.

In each trial, a centrally located fixation cross was presented for 500 ms, followed by 2 pictures from either the “tib” or the “ked” categories. Both pictures were shown for 800 ms, before the auditory presentation of the novel word (target sound). After the target sound was played, the 2 pictures remained on the screen until one was selected.

There were 32 trials in total and each training item was presented twice, once as a target and once as a foil. The position of the target picture on the screen was counterbalanced across trials so that in half the trials, the target picture was presented on the left side of the screen and for the other half it was presented on the right. The presentation of items was randomised across participants and no feedback was given. The aim of this task was to measure word learning, therefore, the two alternative pictures (the target and the foil) were always of the same “grammatical” category (e.g., both the target and the foil had been presented with a “ked…ool” word at training). This ensured that the participants would not be able to rely on the determiner and suffix cues to complete the task but instead would rely on their knowledge of the word itself (stem components).

2.3.5.2 Picture Naming (cued recall)

This task was used to provide a measure of word and grammar learning by assessing participants’ recall accuracy of the novel words.

Each trial started a fixation cross (500 ms), followed by a picture from the training set which remained on the screen until the maximum recording time of 8000 ms was reached. The participants were instructed to name each picture out loud using the words from the novel language. If they could not recall the exact word, they were encouraged to name any part of the word that they could remember. If they couldn't remember any part of the word, they were told to say pass. All responses were audio-recorded in Gorilla and transcribed verbatim by the experimenter offline.

Picture Naming was administered at the beginning and end of every test session (Figure 2.1) and the order of items was randomised for each participant and for each Picture Naming task. There were 32 trials, one trial for each of the 32 words presented during training. The main trials were always preceded by 2 practice trials using items that did not appear in the training set.

Accuracy of recall for the stem portion of the novel word (e.g., **ved** in **ved**eem) was used to measure learning of the words, i.e. the arbitrary components in the form-meaning mapping, while recall of the determiner (ked/tib) and suffix portions of the novel words (e.g., **eem** in ved**eem)** were used to measure learning of the systematic mappings. Recall accuracy of each word component was scored by assigning a score of 1 for correct recall and a score of 0 for incorrect recall or no recall. Scoring followed a lenient criterion such that a score of 1 was awarded to the same or phonologically similar word parts. For example, productions such as 'tip' and 'tee' for *'tib'*, along with 'ket' and 'ken' for *'ked'* were accepted as correct determiners, if the initial sound was correct. For the suffixes, productions such as 'een' and 'ee' for *'eem'*, along with 'oo' and 'all' for *'ool'* were accepted as correct. Like the determiners, stems had to start with the correct initial sound with productions such as 'mosh' for *'morch'* (*tib morch**eem*** – caterpillar) and 'pre' for *'pri'* (*tib prill**eem*** – nurse) being scored as correct. All data were transcribed and scored by the experimenter using the criteria outlined above.

2.3.5.3 Speeded Shadowing

This task was used to assess speeded responses to the systematic mappings. The task consisted of 2 blocks: one to assess speeded responses to the trained item systematic mappings (speeded shadowing training) and one to assess speeded responses to the mappings when they were applied to new items (speeded shadowing generalisation), providing a measure of participants' ability to generalise the underlying systematic regularities to items that they had not previously seen. Thus, the two different sets of stimuli described above were used in the two blocks.

For the training version of this task there were 32 trials in total (1 trial for each of the 32 items). For the generalisation block, there was a total of 8 trials at each session (Figure 2.1). Of the 8 trials, half were comprised of the prior knowledge dependent mappings (2 '*tib –eem'* + female mappings and 2 '*ked –ool'* + male mappings). The other half consisted of prior knowledge independent mappings (2 '*tib -eem'* + insect mappings and 2 '*ked -ool'* + sea creature mappings). A different set of items was presented at each test session to provide a measure of generalisation performance after opportunities for consolidation.

The first block of speeded shadowing began with 2 practice trials. The items used in practise trials did not appear in the training set. For both the training and generalisation blocks, each participant had the items presented to them in a random order.

For this task, a picture was presented alongside its corresponding novel word in the auditory format, and participants were asked to repeat each word aloud, as quickly and as accurately as possible. The speeded nature of this task was a key component, allowing us to assess online processing of the systematic mappings by reducing the opportunity for participants to reflect on their knowledge and engage explicit processes. Each trial began with a fixation cross, presented centrally on the screen for 500 ms, followed by a picture. 500 ms after the picture was shown, participants heard the novel word. The picture remained on the screen for 3000 ms during which time the participants repetition was recorded.

Speeded responses were assessed by measuring reaction times (RTs) via the voice onset latency from the onset of the auditory stimulus to the onset of the participant's repetition. Of key interest was how long it would take participants to repeat the word once the stimulus started to play. To account for the possibility that participants may begin speaking before the end of the stimulus presentation, both the spoken stimulus and the participants response were captured in a single uninterrupted recording of 3000 ms.

Each response was also coded for accuracy, with a score of 1 awarded for a correct response and a score of 0 for an incorrect response. As with Picture Naming, scoring for accuracy on this task followed the same lenient criterion such that a response was marked as correct if it was the same or phonologically similar to the spoken stimulus. Responses were scored as incorrect if no response was made or if a different or phonologically dissimilar word was spoken.

2.3.5.4 Explicit knowledge questionnaire

At the end of Session 3, participants in both groups were asked to complete a short questionnaire to assess any explicit awareness of the underlying systematic mappings that might have emerged during the course of the study. To probe their level of explicit knowledge of the mappings, participants were asked *1) 'What do you think the experiment you participated in was testing?'* and *2) 'Did you notice any patterns in the items?'*

For this task, participants were asked to type their responses and encouraged to provide as much detail as possible. All responses were then scored offline by the experimenter to measure participants' explicit awareness of the semantic-phonological mappings.

For scoring, awareness of the semantic-phonological mappings were administered separately for determiners and suffixes. However, in addition to this, participants were also scored separately on their knowledge of the 2 types of semantic-phonological mappings: prior knowledge dependent and prior knowledge independent. For the determiner prior knowledge dependent mappings, a score of 0 was given for no mention of the mappings, a score of 1 for a mention of one of the mappings (e.g., *tib* = female or *ked* = male) and a score of 2 for a mention of both mappings (e.g., *tib* = female and *ked* = male). For the determiner prior knowledge independent mappings, a score of 0 was given for no mention of the mappings, a score of 1 for a mention of one of the mappings (e.g., *tib* = insects or *ked* = sea creatures) and a score of 2 for a mention of both mappings (e.g., *tib* = insects and *ked* = sea creatures). Similarly, for the suffix prior knowledge dependent mappings, a score of 0 was given for no mention of the mappings, a score of 1 for a mention of one of the mappings (e.g., *eem* = female or *ool* = male) and a score of 2 for a mention of both mappings (e.g., *eem* = female and *ool* = male). For the suffix prior knowledge independent mappings, a score of 0 was given for no mention of the mappings, a score of 1 for a mention of one of the mappings (e.g., *eem* = insects or *ool* = sea creatures) and a score of 2 for a mention of both mappings (e.g., *eem* = insects and *ool* = sea creatures).

Altogether participants were scored on 2 measures of explicit knowledge for the determiners and suffixes each (4 in total).

2.4 Data Analyses

All analyses reported below were conducted in R (R Core Team, 2020). For all models, a maximal model that was able to converge was produced using the *buildmer* package (Voeten, 2021).

Our first research question concerned how offline processes during sleep versus wakefulness contribute to the consolidation of systematic mappings, specifically examining whether these contributions differed based on whether mappings were related or unrelated to participants' existing linguistic knowledge. Based on the CLS (McClelland, 2013) and SLIMM (Van Kesteren et al., 2012) frameworks, we hypothesised that if sleep does play a role in the consolidation of systematic mappings, this would only be evident for mappings independent of prior knowledge, since novel information with fewer connections to existing knowledge is thought to rely more heavily on the hippocampus at encoding. By contrast, for mappings related to prior knowledge, we hypothesised that their connections to existing knowledge would facilitate rapid neocortical integration during encoding, leading to better immediate learning compared to mappings independent of prior knowledge. However, their reduced reliance on the hippocampus during encoding would mean these prior knowledge related mappings would be less likely to benefit from hippocampal reactivation during sleep.

To test these hypotheses, we used both non-speeded and speeded tasks, as these tasks have been shown to have different sensitivity to sleep-related consolidation effects (Tamminen et al., 2012; 2015). First, participants' recall accuracy for the prior knowledge dependent and independent mappings was assessed using a non-speeded Picture Naming task at three time points: (1) immediately after training, (2) following a 12-hour consolidation interval involving either sleep or wake, and (3) after a 1-week delay. Additionally, participants' response times for the systematic mappings were assessed using a speeded shadowing task at the same three time points. In this task, we measured response times to the mappings both in their trained-item contexts and when generalised to new, previously unseen items.

In addition to examining systematic mappings, we also assessed the relative contribution of offline processes during sleep versus wakefulness to the consolidation of the arbitrary components of the novel language (i.e., stem-meaning mappings). We hypothesised that memory for these arbitrary mappings would be better when encoding was followed by sleep compared to wake, due to the greater reliance of arbitrary knowledge on the hippocampus at encoding. To test this hypothesis, participants' recall accuracy for the stem components of the novel words was assessed using the Picture Naming task at three time points: (1) immediately after training, (2) following a 12-hour consolidation interval involving either sleep or wake, and (3) after a 1-week delay. Additionally, participants completed the 2AFC recognition task measuring their recognition of stem-meaning mappings at these same three time points.

To assess whether any differences in performance between the sleep-first and wake-first groups on the language tasks could be attributed to differences in alertness, participants' scores on the Stanford Sleepiness Scale (SSS) and the number of lapses on the Psychomotor Vigilance Task (PVT) were analysed using independent-samples t-tests comparing the sleep-first and wake-first groups at each of the three sessions.

2.4.1 Picture Naming

As described above, participants' responses in the picture naming task were scored for accuracy as a binary measure (1, 0) for each component of the novel word (determiners, stem, suffix). Using these binary outcome variables, three mixed-effects binomial regressions (one for each component: determiner, stem, and suffix) were conducted using the glmer function in the lme4 package in R (Bates et al., 2015).

To assess the influence of sleep vs. wake-related consolidation and prior knowledge on recall of the arbitrary and systematic components of the novel words, group (sleep first vs. wake first), session (immediate, 12-hour delay, 1-week delay), and mapping (prior knowledge dependent vs. prior knowledge independent) were entered as fixed effects. The fixed effects of group and mapping were coded using sum contrasts (group: sleep first = 1; wake first = -1; mapping: prior knowledge dependent = 1; prior knowledge independent = -1). For session, custom pairwise contrasts were used to compare performance at Session 1 (immediate test) with performance at Session 2 (after a 12-hour consolidation interval of sleep or wake) (S2vsS1; S2 = 1; S1 = -1; S3 = 0) and to compare Session 2 performance with Session 3 (after a 1-week delay) (S3vsS2; S3 = 1; S2 = -1; S1 = 0).

For the random-effect structure, a maximal model (Barr et al., 2013) was initially fitted using the buildmer function in R, including random intercepts for participants and items, and random slopes for mapping and session contrasts (S2vsS1, S3vsS2) for participants, and for group and session contrasts for items. If this maximal model did not converge, buildmer simplified the random-effects structure until convergence was achieved. The final model that converged is reported in the Results section.

2.4.2 Speeded Shadowing

2.4.2.1 Speeded Shadowing (trained items)

Response Times. Performance on speeded shadowing training was used to assess online processing of the trained item mappings (determiner + suffix cues). For this task, of key interest was the time taken (RTs) for participants to repeat the novel word for a trained item aloud after it was presented auditorily. RTs were marked manually using CheckVocal (Protopapas, 2007). To account for the possibility that participants may begin speaking before the offset of the stimulus, RTs were measured from the onset of the spoken stimulus to the onset of the participants' response.

The data for 7 participants (4 in the sleep-first group and 3 in the wake-first group) were excluded for failing to repeat the stimuli correctly for at least 80% of the trials. Of the remaining data (n = 53), trials where participants failed to respond or responded incorrectly were excluded (3.01% of the data in Session 1 (sleep first group: 1.89%, wake first group: 1.12%); 3.01% of the data in Session 2 (sleep first group: 1.59%, wake first group: 1.42%); 3.24% of the data in Session 3 (sleep first group: 1.65%, wake first group: 1.59%)). To analyse RT data, a generalised linear mixed-effects model (GLMM) was implemented to better meet the assumption of normality (Lo & Andrews, 2015) using the glmer function with a gamma distribution in the lme4 package in R.

To examine the influence of sleep vs. wake-related consolidation and prior knowledge on speeded responses to the trained item systematic mappings group, mapping and session were used as fixed effects, coded as described for the Picture Naming task above. For the random effects structure, the maximal model included the same random effects as in Picture Naming. If the maximal model did not converge, buildmer simplified the random-effects structure until convergence was achieved. The final model that converged is reported in the Results section.

2.4.2.2 Speeded Shadowing (generalisation items)

Response Times. This task was used to measure generalisation by assessing speeded responses to the mappings when applied to new, untrained items. As with the trained item mappings of key interest was the time taken (RT) for participants to repeat the novel word for an untrained item aloud after it was presented auditorily. RTs for Session 1 and Session 2 were marked manually in CheckVocal (Protopapas, 2007); however, due to a technical error during data collection, Session 3 responses were not recorded for this task and therefore, the data for this session was not included in the analysis presented below.

As with the speeded shadowing training task, before any analysis on RTs was conducted, response accuracy was checked. Due to the generalisation items being seen for the first (and only) time in this task, a more lenient criterion of 75% accuracy was used. Thus the data for 9 participants (4 in the sleep-first group and 5 in the wake-first group) were excluded for failing to repeat the novel word correctly for at least 75% of the trials. Of the remaining data (n = 51), all inaccurate trials were removed (6.86% of the data in session 1 (sleep first group: 2.94%, wake first group: 3.92%); 9.07% of the data in session 2 (sleep first group: 5.15%, wake first group: 3.92%)). Accurate RTs were then analysed using a generalised linear mixed-effects regression with a gamma distribution (glmer function of the lme4 package in R), with group, mapping and session as fixed effects, all coded using sum contrasts.

The same approach was used for the random effects structure as in the analysis of the trained items described above.

2.4.2.3 Speeded Shadowing: Accuracy

In addition to response times, response accuracy on both the speeded shadowing training and generalisation tasks was analysed to assess whether offline consolidation or the nature of the mapping (whether or not it was related to prior knowledge) influenced the accuracy of responses. For this analysis, the data for all participants and all trials were included. Using accuracy as the binary outcome, two mixed-effects binomial regressions with the glmer function in the lme4 package in R were used to analyse the data. For the training items, group and mapping were entered as fixed effects using sum contrast coding while for session, Session 1 accuracy was compared with accuracy at Session 2 (S2vsS1: S2 = 1; S1 = -

1; S3 = 0) and accuracy at Session 2 compared with Session 3 accuracy (S3vsS2; S3 = 1; S2 = -1; S1 = 0). For the generalisation items, group, mapping and session were also entered as fixed effects; however, only Session 1 and Session 2 data were available so accuracy was compared between the two (session: S1 = - 1, S2 = 1). The same approach as above was used for the random effects structure.

2.4.3 2AFC Recognition

To assess the learning and consolidation of the arbitrary components of the novel words presented during training, a mixed-effects binomial regression was conducted on accuracy data (0,1) using sum contrast coding to test recognition of the novel words at Session 1 (immediate test), 2 (12-hr delay) and 3 (1-week delay). As above, group (sleep first, wake first), mapping (prior knowledge dependent, independent) and session (immediate, 12hr delay, 1-week delay) were coded as fixed effects. For session, Session 1 accuracy was compared with accuracy at Session 2 (S2vsS1: S2 = 1; S1 = -1; S3 = 0) and accuracy at Session 2 compared with Session 3 accuracy (S3vsS2; S3 = 1; S2 = -1; S1 = 0). The same approach as above was used for the random effects structure.

2.4.4 Explicit Knowledge Questionnaire

This questionnaire administered at the end of Session 3 was used to assess any explicit awareness of the underlying systematic mappings that might have emerged in the course of the study. For this, the coding scheme described above was used, with scores ranging from 0 to 2. The questionnaire yielded 4 outcome measures, assessing explicit awareness of the semantic-phonological mappings (determiner prior knowledge dependent, determiner prior knowledge independent, suffix prior knowledge dependent, and suffix prior knowledge independent).

To explore the influence of offline consolidation (sleep vs. wake) and mapping type (prior knowledge dependent vs. independent) on participants' explicit awareness of the prior knowledge dependent and independent determiner and suffix mappings (semantic-phonological awareness), we ran two linear mixed models: one with awareness scores for the prior knowledge dependent and independent *determiner* semantic-phonological mappings as the outcome and the second with awareness scores for the prior knowledge dependent and independent *suffix* semantic-phonological mappings as the outcome. For both models, group and mapping were entered as fixed effects with random intercepts included for participants.

2.5 Results

2.5.1 Post-Training Performance

The Picture Naming task administered at the end of the training session at Session 1 (after the final repetition block; see Figure 2.1) was used to assess how well participants in both groups had learned the novel language. Using this task, we measured knowledge of the systematic components of the novel language using accuracy of the recall of the determiner (e.g., *tib* in *tib* vedeem) and accuracy of the recall of the suffix (e.g., *eem* in tib ved*eem*). We also assessed accuracy in recalling the stem component (e.g., *ved* in *ved*eem) as a measure of the learning of the arbitrary components of the novel words.

To analyse the data, three mixed-effects binomial regressions (one for each component: determiner, suffix and stem) were conducted with mapping (prior knowledge dependent vs. prior knowledge independent) and group (sleep first vs. wake first) coded as fixed effects.

2.5.1.1 Systematic Mappings

Figures 2.3 and 2.4 illustrate the learning of the systematic mappings (determiners and suffixes) for the participants in both groups. On average, participants showed greater accuracy in their recall of the suffixes (Figure 2.4; prior knowledge dependent *suffix* mappings: $M_{\text{Sleep First}} = 0.64$, $M_{\text{Wake First}} = 0.67$; prior knowledge independent *suffix* mappings: $M_{\text{Sleep First}} = 0.52$, $M_{\text{Wake First}} = 0.57$) than the determiners (Figure 2.3; prior knowledge dependent *determiner* mappings: $M_{\text{Sleep First}} = 0.41$, $M_{\text{Wake First}} = 0.52$; prior knowledge independent *determiner* mappings: $M_{\text{Sleep First}} = 0.34$, $M_{\text{Wake First}} = 0.39$).

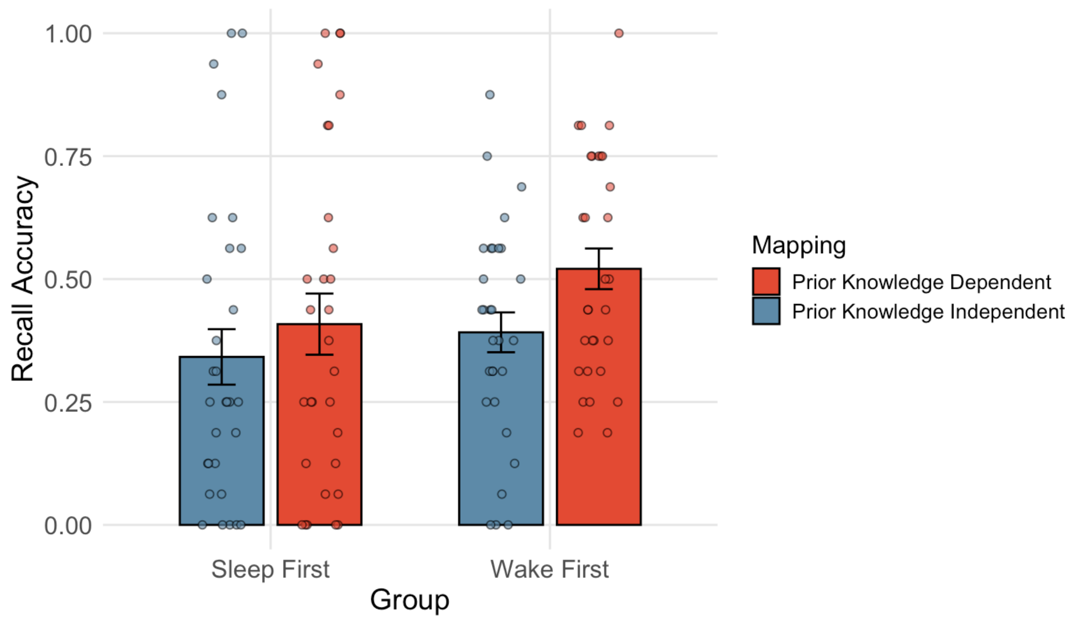

Figure 2.3. Determiner recall accuracy (proportion correct) on the first post training Picture Naming task at Session 1. Error bars represent standard error.

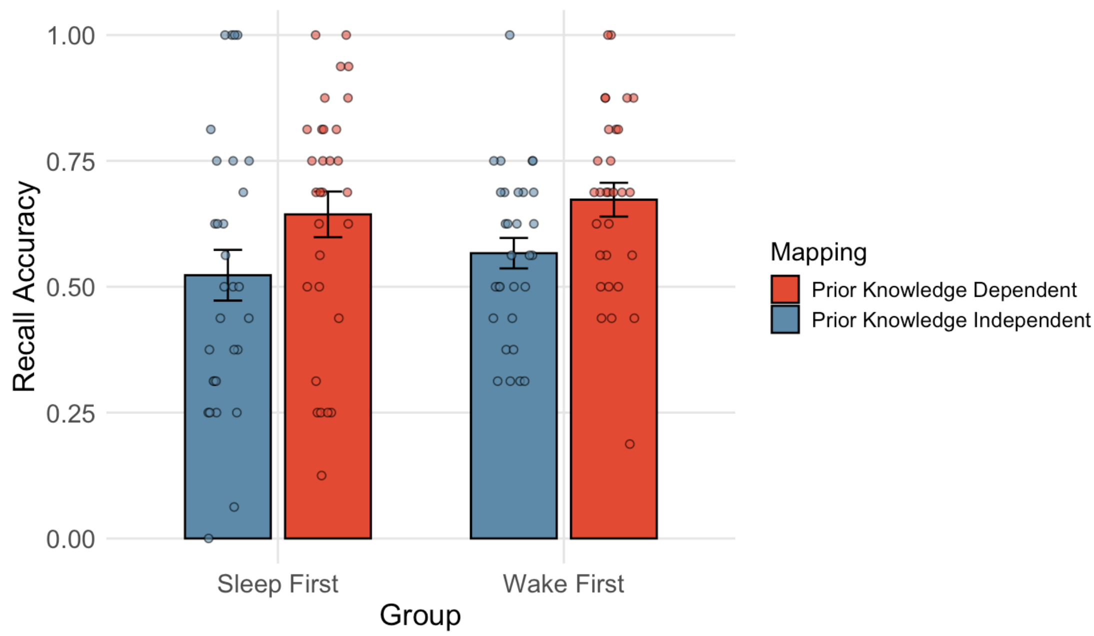

Figure 2.4. Suffix recall accuracy (proportion correct) on the first post training Picture Naming task at Session 1. Error bars represent standard error.

For the determiners, a linear mixed regression revealed a significant main effect of mapping, indicating that participants in both groups recalled determiners more accurately in the prior knowledge dependent condition compared to the prior knowledge independent condition immediately after training (Table 2.2; Figure 2.3). There was no significant main effect of group nor an interaction between group and mapping, suggesting that performance was comparable across both groups (Table 2.2).

Table 2.2 Regression coefficients for the fixed factors using determiner recall accuracy in the Picture Naming task immediately after training at Session 1 as the outcome.

	Estimate	Std. Error	z	p
Mapping	**0.274**	**0.084**	**3.267**	**.001**
Group	–0.255	0.210	–1.212	.226
Mapping x Group	–0.056	0.069	–0.811	.417

Note. Random effects included in the model: intercepts by participant and items. For participants, slopes were included for mapping. This structure represents the maximal model that successfully converged. Significant results are highlighted in bold.

For the suffixes, a linear mixed regression similarly revealed a significant main effect of mapping, indicating that immediately after training, participants in both groups recalled suffixes more accurately in the prior knowledge dependent condition compared to the prior knowledge independent condition (Table 2.3; Figure 2.4). Again, there was no significant main effect of group nor a significant interaction between group and mapping, indicating that performance was comparable across both groups for the suffix mappings as well (Table 2.3).

Table 2.3 Regression coefficients for the fixed factors using suffix recall accuracy in the Picture Naming task immediately after training at Session 1 as the outcome.

	Estimate	Std. Error	z	p
Mapping	**0.296**	**0.082**	**3.610**	**<.001**
Group	–0.050	0.146	–0.344	.731
Mapping x Group	0.034	0.053	0.646	.519

Note. Random effects included in the model: intercepts by participant and items. For participants, slopes were included for mapping. This structure represents the maximal model that successfully converged. Significant results are highlighted in bold.

2.5.1.2 Arbitrary Mappings

As shown in Figure 2.5, stem recall accuracy was lower relative to the recall of determiner and suffix mappings. Participants in both groups showed comparable recall accuracy for stems in the prior knowledge dependent and prior knowledge independent conditions, as indicated by the non-significant main effect of mapping and the absence of significant interaction between group and mapping (Table 2.4). Additionally, no significant main effect of group was observed, suggesting that, immediately after training, overall stem recall performance was similar for both groups (Table 2.4, Figure 2.5).

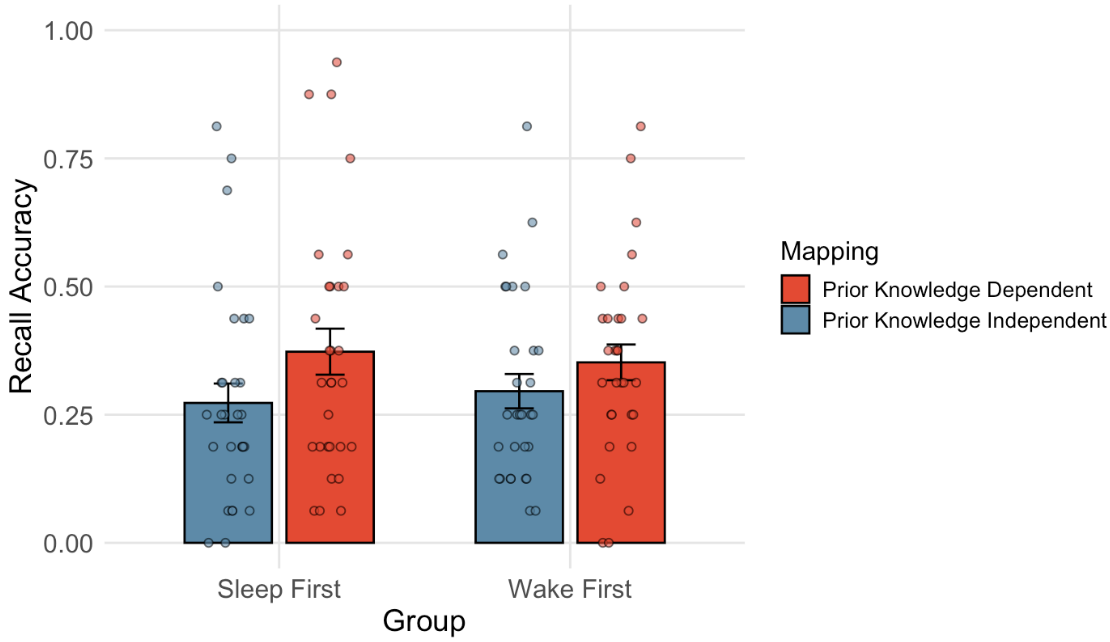

Figure 2.5. Stem recall accuracy (proportion correct) on the post training Picture Naming task. Error bars represent standard error.

Table 2.4 Regression coefficients for the fixed factors using stem recall accuracy in the Picture Naming task immediately after training at Session 1 as the outcome.

	Estimate	Std. Error	z	p
Mapping	0.252	0.165	1.524	.127
Group	-0.026	0.146	–0.182	.856
Mapping x Group	0.077	0.056	1.357	.175

Note. Random effects included in the model: intercepts by participant and items. This structure represents the maximal model that successfully converged. Significant results are highlighted in bold.

2.5.2 Post-Delay Performance

To assess how prior knowledge would influence the learning and consolidation of the systematic components of the novel language (determiner and suffix mappings), we used two types of tasks: non-speeded and speeded. First, the non-speeded Picture Naming task assessed recall accuracy for trained-item determiner and suffix mappings, both prior knowledge dependent and independent, immediately after training, following a 12-hour consolidation period involving either sleep or wake, and after a 1-week delay. Additionally, learning of these mappings was assessed through a speeded shadowing task, measuring response times to the novel words with the determiner and suffix mappings in the prior knowledge dependent and independent conditions across the three test points (immediately after training, after the 12-hour consolidation period, and after a 1-week delay). For this task, response times were assessed for the mappings in their trained-item contexts (paired with the original word stems encountered during training) as well as in novel contexts (paired with new, untrained word stems). Thus, the speeded task also provided a measure of participants' ability to generalise the mappings to novel items, thereby offering a purer test of systematic knowledge that did not rely on memory for specific item associations formed during training.

To examine how offline consolidation influenced the learning of the arbitrary mappings, we assessed recall accuracy for the stem components of the novel words in the Picture Naming task and recognition accuracy of the novel words in a 2AFC task immediately after training, following a 12-hour consolidation interval involving either sleep or wake, and again after a 1-week delay.

2.5.2.1 Systematic Mappings

Picture Naming: Determiners. Figure 2.6 illustrates the recall of the determiner mappings for both groups across the three sessions. The analysis showed that participants in both groups and across all three sessions recalled the determiners more accurately with the novel words in the prior knowledge dependent condition than the words in the prior knowledge independent condition, as indicated by the significant main effect of mapping, and no significant interactions with group or session (Table 2.5, Figure 2.6).

Participants in both groups showed a decrease in determiner accuracy over time, as shown by a significant main effect of the Session 2 vs. Session 1 and Session 3 vs. Session 2 contrasts (Table 2.5). The level of decrease in performance across the sessions did not interact with group, suggesting that consolidation was not influenced by brain state (wake or sleep), either for the first 12 hours, or over a week (see Table 2.5: no significant S2vsS1 x group or S3vS2 x group interactions). No other significant main effects or interactions were observed (see Table 2.5).

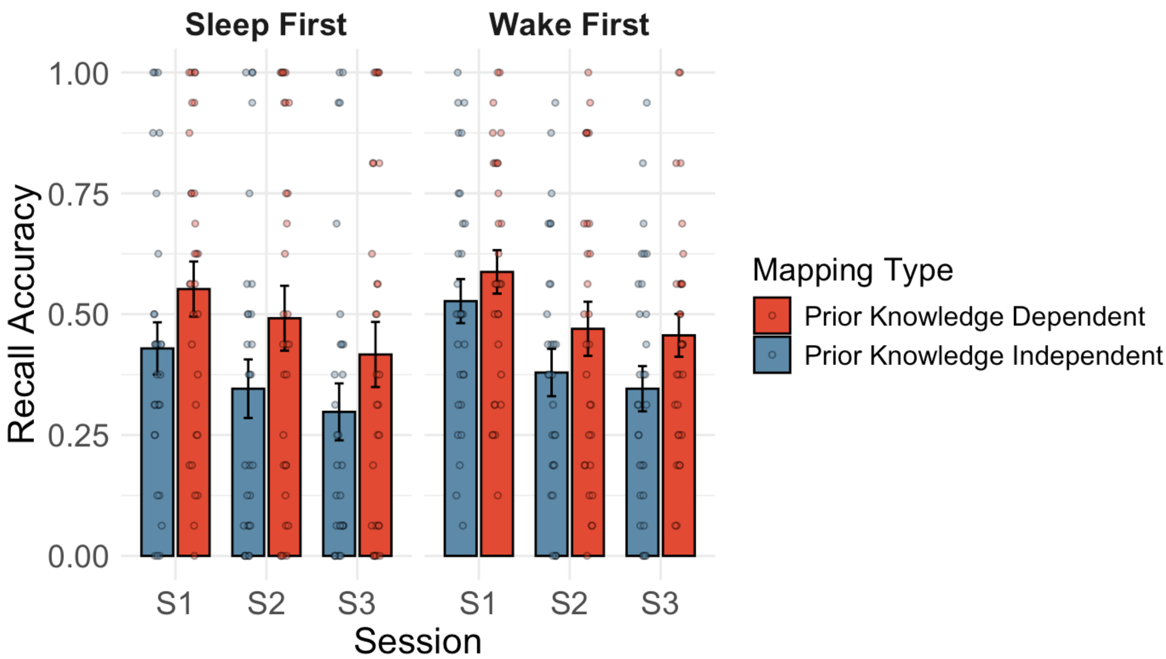

Figure 2.6. Determiner recall accuracy (proportion correct) on the Picture Naming task immediately after training (S1), after a 12-hour delay (S2), and after a week (S3). Error bars represent standard error.

Table 2.5 Regression coefficients for the fixed factors using determiner recall accuracy in the Picture Naming task as the outcome.

	Estimate	Std. Error	z	p
Mapping	**0.371**	**0.088**	**4.225**	**<.001**
Mapping x Group	0.106	0.061	1.746	.081
Mapping x S2vsS1	0.081	0.047	1.719	.086
Mapping x S3vsS2	0.029	0.050	0.571	.568
S2vsS1	**–0.523**	**0.082**	**–6.392**	**<.001**
S3vsS2	**–0.408**	**0.096**	**–4.256**	**<.001**
S2vsS1 x Group	–0.009	0.081	–0.106	.916
S3vsS2 x Group	–0.116	0.096	–1.214	.225
Group	–0.120	0.251	–0.477	.634
Mapping x S2vsS1 x Group	–0.009	0.047	–0.191	.848
Mapping x S3vsS2 x Group	–0.039	0.050	–0.780	.436

Note. Random effects included in the model: intercepts by participant and items. For participants, slopes were included for mapping and the session contrasts and for item, slopes for group. This structure represents the maximal model that successfully converged. Significant results are highlighted in bold.

Picture Naming: Suffixes. The results for the suffix mappings followed a similar pattern to those observed for the determiners. First, the analysis revealed a significant main effect of mapping on suffix recall accuracy, with higher accuracy for the suffix mappings in the prior knowledge dependent condition (Table 2.6; Figure 2.7). This greater accuracy for the prior knowledge dependent suffix mappings did not interact with group or session, suggesting that, like the determiner mappings, participants in both groups recalled these mappings with greater accuracy across all three sessions (Table 2.6).

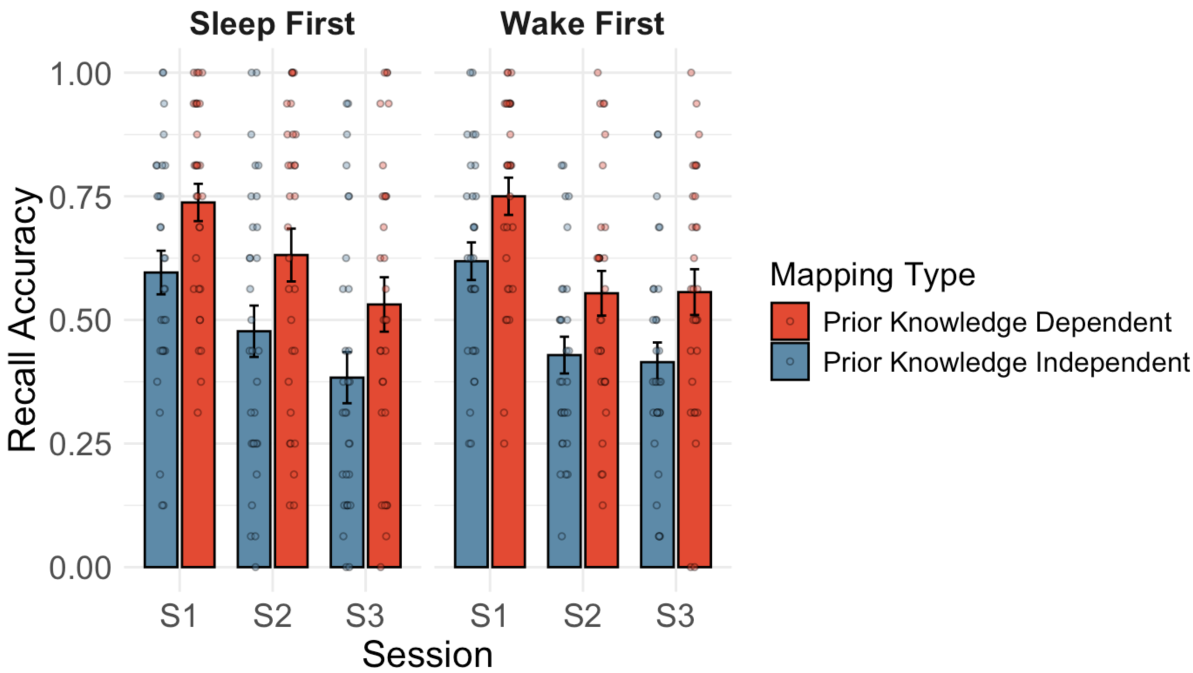

Figure 2.7. Suffix recall accuracy (proportion correct) on the Picture Naming task immediately after training (S1), after a 12-hour delay (S2), and after a week (S3). Error bars represent standard error.

As with the determiners, and as illustrated in Figure 2.7, participants demonstrated a decline in their recall accuracy of the suffixes over time as indicated by the significant main effect of the Session 2 vs Session 1 and Session 3 vs Session 2 contrasts (Table 2.6). The decline in performance across sessions did not significantly differ between groups, indicating that consolidation was not influenced by sleep or wake, both over the initial 12-hour period and across the subsequent week (see Table 2.6: no significant S2vsS1 x group or S3vS2 x group interactions). No other main effects or interactions were significant (see Table 2.6).

Table 2.6 Regression coefficients for the fixed factors using suffix recall accuracy in the Picture Naming task as the outcome.

	Estimate	Std. Error	z	p
Mapping	**0.406**	**0.086**	**4.697**	**<.001**
Mapping x Group	0.062	0.054	1.157	.247
Mapping x S2vsS1	–0.011	0.046	–0.241	.810
Mapping x S3vsS2	–0.001	0.045	–0.032	.974
S2vsS1	**–0.629**	**0.062**	**–10.128**	**<.001**
S3vsS2	**–0.495**	**0.077**	**–6.454**	**<.001**
S2vsS1 x Group	0.016	0.060	0.271	.786
S3vsS2 x Group	-0.111	0.076	–1.456	.145
Group	0.050	0.162	0.307	.759
Mapping x S2vsS1 x Group	0.016	0.045	0.347	.728
Mapping x S3vsS2 x Group	–0.010	0.045	–0.216	.829

Note. Random effects included in the model: intercepts by participant and items. For participants, slopes were included for mapping and the session contrasts and for item, slopes for group. This structure represents the maximal model that successfully converged. Significant results are highlighted in bold.

Overall, while we see evidence of a role for prior knowledge in the learning of the systematic mappings, the overall decline in recall of the determiner and suffixes for both the Sleep First and Wake First groups after a 12-hour and 1-week delay suggests that offline consolidation did not influence performance in the non-speeded Picture Naming task.

Speeded Shadowing. In addition to the non-speeded Picture Naming task, participants' learning and consolidation of the systematic mappings was also assessed using a speeded task. Here we assessed speeded responses to the trained items, as well as generalisation to previously unseen items.

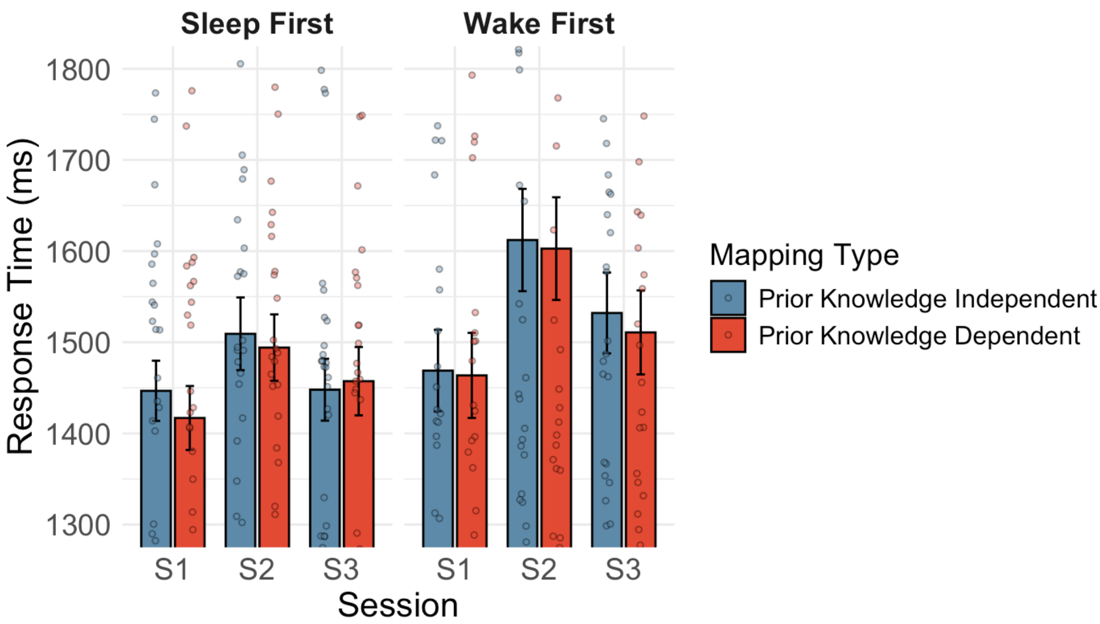

Figure 2.8. Response times (milliseconds) for the Speeded Shadowing task with the trained items immediately after training (S1), after a 12-hour delay (S2), and after a week (S3). Error bars represent standard error.

Trained Items (Response Times). Response times for the trained items in the speeded shadowing task are illustrated in Figure 2.8. The analysis showed no significant main effects of mapping and group (Table 2.7). There was a significant increase in response times at Session 2 relative to Session 1 (S2vsS1, Table 2.7), but there was no difference at Session 3 relative to Session 2 (S3vsS2, Table 2.7). There was a significant three-way interaction between mapping, group and S2vsS1 contrast (Table 2.7), showing that the increase in response times in the first 12 hours after training varied between groups and mappings. In the sleep-first group, the initial benefit in response times at Session 1 for the prior knowledge dependent mappings relative to the prior knowledge independent mappings was reduced at Session 2 (Figure 2.8), whereas in the wake-first group the response times for the two mappings were similar at both sessions and significantly longer at Session 2 relative to Session 1 than in the sleep-first group. There was also a significant three-way interaction between mapping, group, and S3vsS2 contrast (Table 2.7). Here, in the sleep-first group there was a similar decrease in response times for both types of mappings (Figure 2.8), but a somewhat greater reduction in the prior knowledge dependent mappings in the wake-first group. There were no other significant main effects or interactions (Table 2.7).

Table 2.7 Regression coefficients for the fixed factors using accurate response times in the speeded shadowing task with the trained items as the outcome.

	Estimate	Std. Error	t	p
Mapping	−0.004	0.013	−0.333	.739
Mapping × Group	0.000	0.002	0.029	.977
Mapping × S2vsS1	0.002	0.002	1.172	.241
Mapping × S3vsS2	0.002	0.002	0.751	.453
S2vsS1	**0.032**	**0.011**	**2.783**	**.005**
S3vsS2	−0.005	0.013	−0.400	.683
S2vsS1 × Group	−0.011	0.011	−0.998	.318
S3vsS2 × Group	−0.001	0.013	−0.115	.908
Group	−0.020	0.024	−0.800	.422
Mapping × S2vsS1 × Group	**0.004**	**0.002**	**2.142**	**.032**
Mapping × S3vsS2 × Group	**0.005**	**0.002**	**−2.174**	**.030**

Note. Random effects included in the model: intercepts by participant and items. For participants, slopes were included for mapping, both session contrasts (S2vsS1, S3vsS2), and the interaction between mapping and the S3vsS2 contrast. For items, slopes were included for group and the S3vsS2 contrast. This structure represents the maximal model that successfully converged. Significant results are highlighted in bold.

To further explore the three-way interactions, we conducted an exploratory analysis in which we controlled for any baseline differences between the two groups at Session 1 by examining the changes in response times between sessions for both mapping types and groups. Specifically, for each participant and each individual item, difference scores were calculated by subtracting response times at Session 1 from Session 2 (S2–S1) and subtracting response times at Session 2 from Session 3 (S3–S2). These difference scores were then analysed in two separate mixed-effects linear regressions: one with the S2–S1 difference scores as the outcome, and one with the S3–S2 difference scores as the outcome. For both models, mapping, group and their interaction were entered as fixed factors. The results of the analysis with the S2-S1 scores revealed no significant effect of group, mapping or their interaction on the change in response times between sessions (Table 2.8). Similarly, no significant effect of group, mapping or their interaction was observed for the change in response times between Session 2 and Session 3 (Table 2.9). Therefore, these findings suggest that the three-way interactions observed for the analysis on raw response times were likely driven by group differences in performance at Session 1.

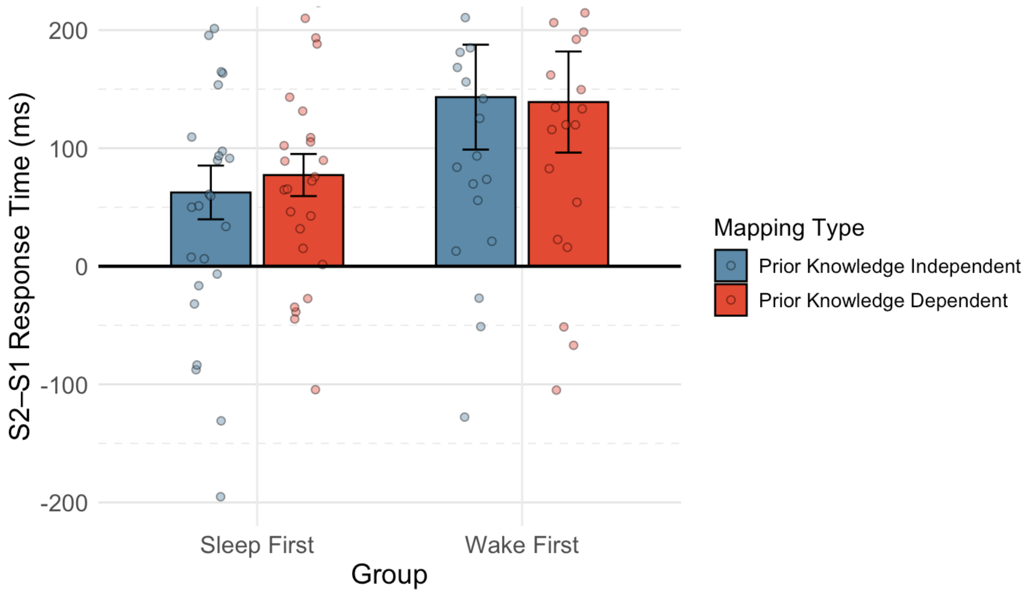

Figure 2.9(a). Difference in response times (milliseconds) between sessions 1 and 2 for the Speeded Shadowing task with the trained items. Positive values indicate increases in response times, while negative values indicate a decrease in response times. Error bars represent standard error.

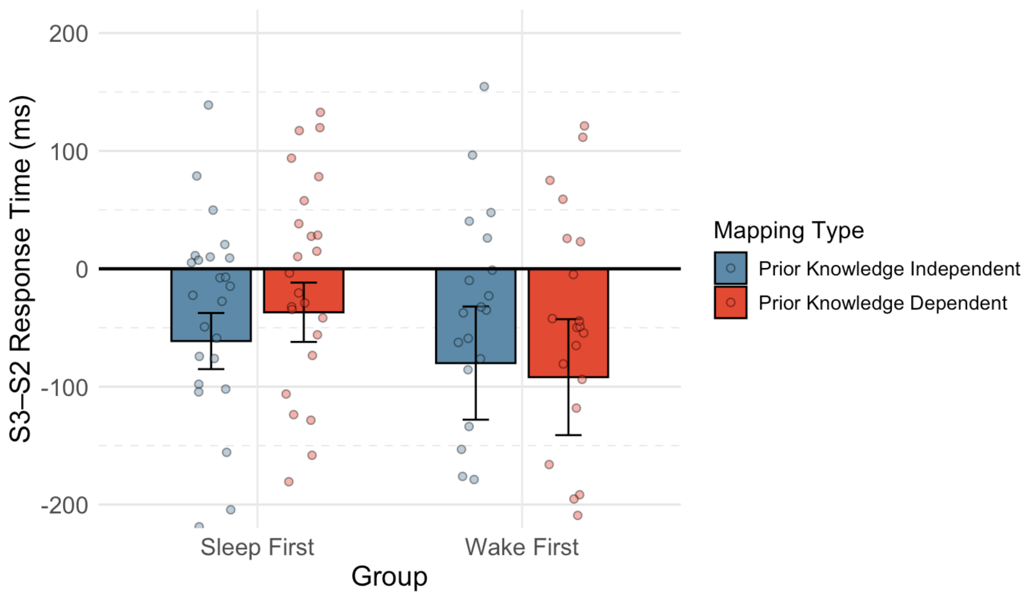

Figure 2.9(b). Difference in response times (milliseconds) between sessions 2 and 3 for the Speeded Shadowing task with the trained items. Positive values indicate increases in response times, while negative values indicate a decrease in response times. Error bars represent standard error.

Table 2.8 Regression coefficients for the fixed factors using differences in response times between Session 1 and 2 in the speeded shadowing task with the trained items as the outcome.

	Estimate	Std. Error	t	p
Mapping	2.64	4.66	0.57	.570
Group	-35.64	23.94	-1.49	.137
Group × Mapping	4.73	4.66	1.02	.310

Note. Random effects included in the model: intercepts by participants. This structure represents the maximal model that successfully converged. Significant results are highlighted in bold.

Table 2.9 Regression coefficients for the fixed factors using differences in response times between Session 2 and 3 in the speeded shadowing task with the trained items as the outcome.

	Estimate	Std. Error	t	p
Mapping	3.13	4.72	0.66	.508
Group	18.44	27.12	0.68	.497
Group × Mapping	9.09	4.72	1.93	.054

Note. Random effects included in the model: intercepts by participants. This structure represents the maximal model that successfully converged. Significant results are highlighted in bold.

Generalisation Items (Response Times). We also assessed speeded responses to the mappings when applied to new, untrained items in the speeded shadowing task. Here we only had data available for Sessions 1 and 2.

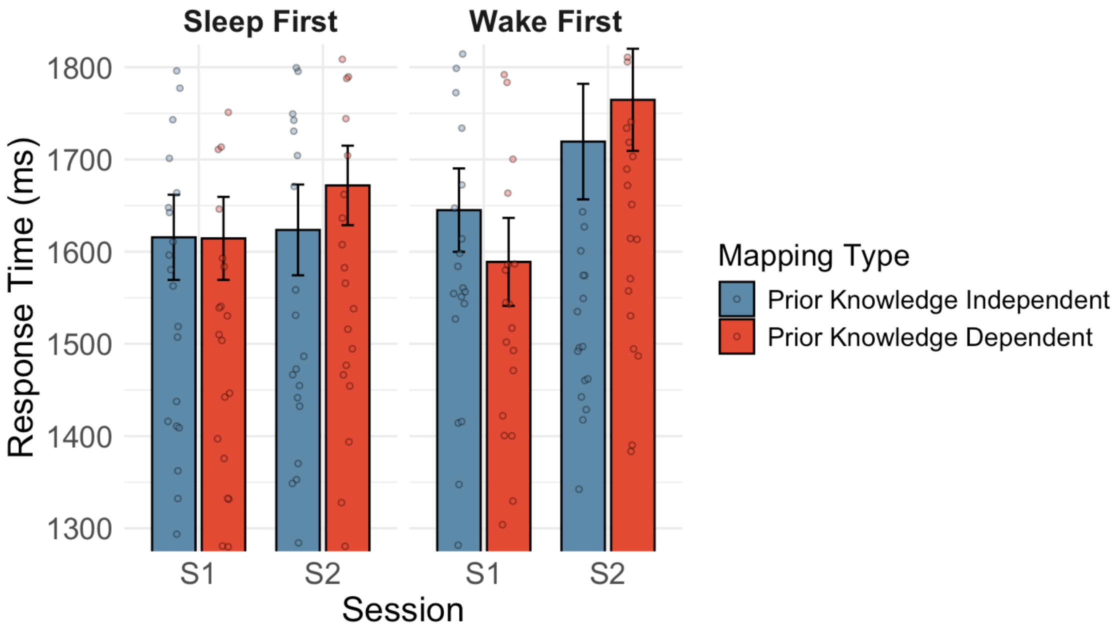

Figure 2.10. Response times (milliseconds) for the Speeded Shadowing task with the generalisation items. Error bars represent standard error

As shown in Table 2.10, none of the main effects or interactions were statistically significant. These findings show that response times for the previously unseen items did not vary between the two types of mappings, across the two sessions, or across the two groups.

Table 2.10 Regression coefficients for the fixed factors using accurate response times in the speeded shadowing task with the generalisation items as the outcome.

	Estimate	Std. Error	t	p
Mapping	0.007	0.019	0.351	.726
Mapping × Session	0.014	0.018	0.776	.438
Mapping x Group	0.005	0.006	0.880	.379
Session	0.022	0.024	0.899	.369
Session × Group	−0.013	0.012	−1.148	.251
Group	−0.014	0.028	−0.508	.612
Mapping × Session × Group	−0.003	0.005	−0.657	.511

Note. Random effects included in the model: intercepts by participant and item. For participants, slopes were included for mapping contrasts, session contrasts, and their interaction. For items, slopes were included for group contrasts. This structure represents the maximal model that successfully converged. Significant results are highlighted in bold.

Trained Items (Accuracy). In addition to the analysis of response times, we also assessed the accuracy of participants' responses. As illustrated in Figure 2.11, overall accuracy for the trained items on this task was high across all three sessions.

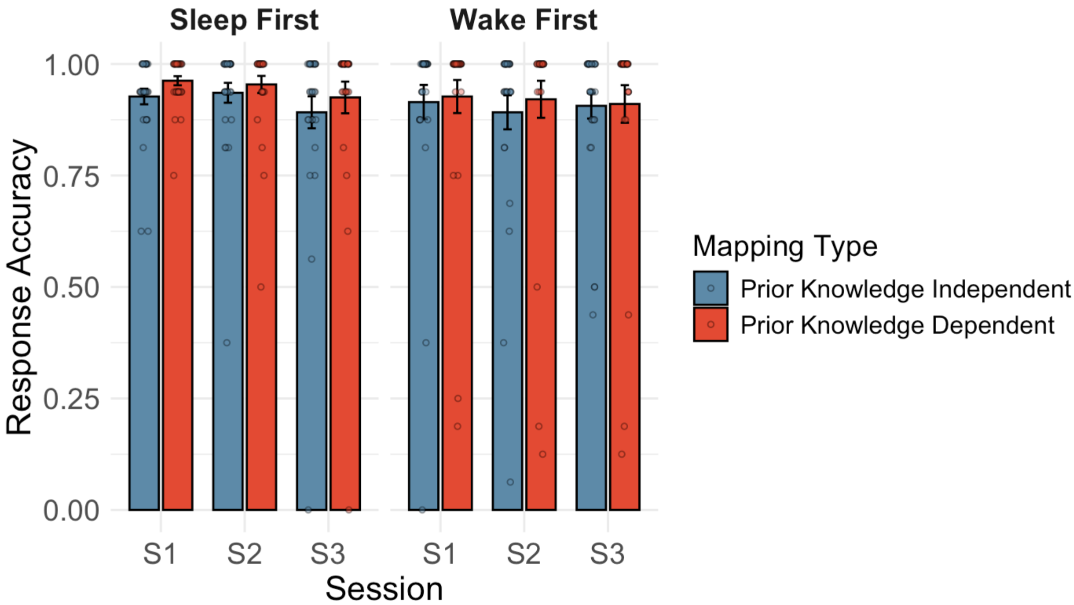

Figure 2.11. Response accuracy (proportion correct) for the trained items in the Speeded Shadowing task. Error bars represent standard error.

For the analysis of the trained items, a logistic mixed regression with mapping, group, S2vsS1 and S3vsS2 as fixed effects was conducted using accuracy as the binary outcome measure. The analysis revealed that, for both groups and across all three sessions, accuracy was greater for the prior knowledge dependent relative to prior knowledge independent mappings (see Figure 2.11), as indicated by the main effect of mapping and the non-significant interaction with group or session (Table 2.11). No other main effects or interactions were significant.

Table 2.11 Regression coefficients for the fixed factors using accuracy in the speeded shadowing task with the trained items as the outcome.

	Estimate	Std. Error	z	p
Mapping	0.628	0.201	3.120	.002
Mapping × Group	-0.117	0.123	-0.951	.342
Mapping × S2vsS1	-0.115	0.109	-1.055	.289
Mapping × S3vsS2	-0.120	0.102	-1.176	.240
S2vsS1	-0.169	0.193	-0.877	.380
S3vsS2	-0.336	0.205	-1.639	.101
S2vsS1 × Group	0.130	0.149	0.866	.385
S3vsS2 × Group	0.000	0.176	0.003	.998
Group	-0.113	0.286	-0.395	.693
Mapping × S2vsS1 × Group	-0.023	0.101	-0.226	.821
Mapping × S3vsS2 × Group	0.104	0.097	1.074	.283

Note. Random effects included in the model: intercepts by participant and items. For participants, slopes were included for mapping and session contrasts and for item, slopes for group. This structure represents the maximal model that successfully converged. Significant results are highlighted in bold.

Generalisation Items (Accuracy). As illustrated in Figure 2.12, the overall accuracy with the previously unseen items was slightly lower relative to the trained items.

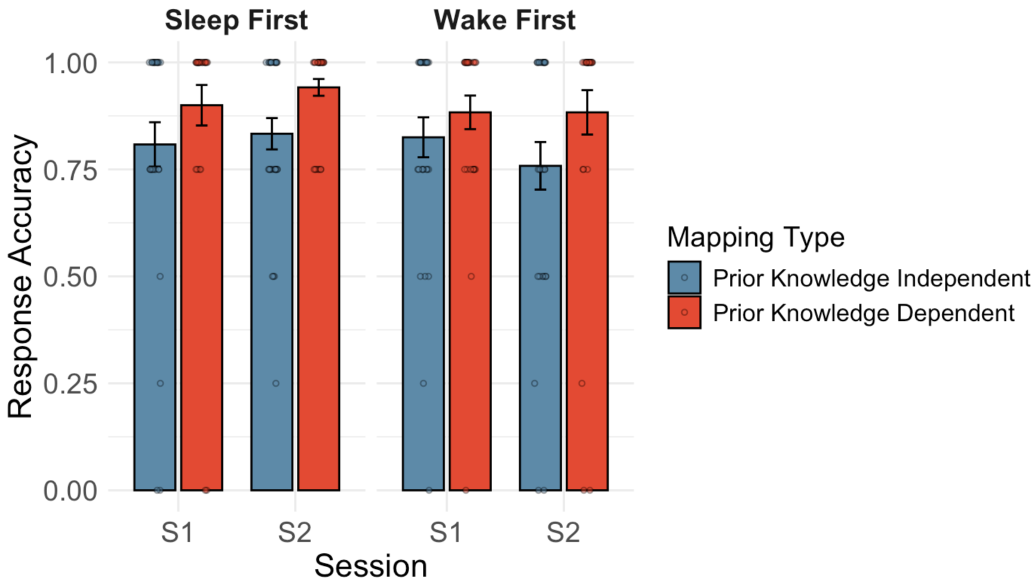

Figure 2.12. Response accuracy (proportion correct) for the generalisation items in the Speeded Shadowing task. Error bars represent standard error.

For the analysis of the generalisation items, a logistic mixed regression with mapping, group and session (S2vsS1) as fixed effects was conducted using accuracy as the binary outcome measure. Like the trained items, this analysis revealed a significant main effect of mapping, with higher accuracy overall for the prior knowledge dependent mappings relative to the prior knowledge independent mappings (Table 2.12). No other main effects or interactions were significant.

Table 2.12 Regression coefficients for the fixed factors using accuracy in the speeded shadowing task with the generalisation items as the outcome.

	Estimate	Std. Error	z	p
Mapping	0.839	0.255	3.293	.001
Mapping x Group	0.090	0.150	0.604	.546
Mapping × Session	-0.018	0.199	-0.093	.926
Session	-0.144	0.285	-0.505	.613
Session × Group	0.060	0.197	0.303	.762
Group	0.202	0.270	0.748	.455
Mapping × Session × Group	-0.098	0.130	-0.750	.453

Note. Random effects included in the model: intercepts by participant and items. For participants, slopes were included for mapping and session contrasts and for item, slopes for session. This structure represents the maximal model that successfully converged. Significant results are highlighted in bold.

Overall, the results of the speeded shadowing task provide some convergent evidence with findings from the non-speeded Picture Naming task, in that when there is a benefit of prior knowledge at initial learning (Session 1) neither sleep- nor wake-related consolidation processes influence the long-term retention of the determiner and suffix mappings, even for those mappings that are unrelated to existing neocortical knowledge. There was no strong evidence for the role of prior knowledge in the response times in the speeded shadowing task either for the trained or the generalisation items. However, accuracy analyses revealed consistently higher response accuracy for the mappings related to prior knowledge compared to those unrelated, across both trained and generalisation items and irrespective of session or group. Thus, response times in the speeded task were not sensitive to prior knowledge, unlike accuracy.

2.5.2.2 Arbitrary Mappings

The previous section presented the results for the systematic components of the novel words (determiner and suffix mappings). This section now focuses on the arbitrary components (i.e., stem-meaning mappings). Of key interest was how sleep versus wake-related consolidation would influence the learning of the arbitrary mappings. However, we also considered the possibility that the prior knowledge manipulation of the systematic mappings might influence the learning of the stems. Therefore, to address this, we examined the influence of sleep- and wake-related consolidation and prior knowledge on learning these arbitrary mappings by measuring participants' recall accuracy for the stem components of the novel words with the mappings in the prior knowledge dependent and prior knowledge independent condition in the Picture Naming task immediately after training, after a 12-hour consolidation interval of sleep or wake and after a 1-week delay. In addition to this, we assessed participants' word knowledge using a 2AFC task across the three time points.

Picture Naming: Stems. Participants' accuracy in recalling the stem component (e.g., *ved* in *ved*eem) in the Picture Naming task was used as a measure of learning and consolidation of the arbitrary components of the novel words.

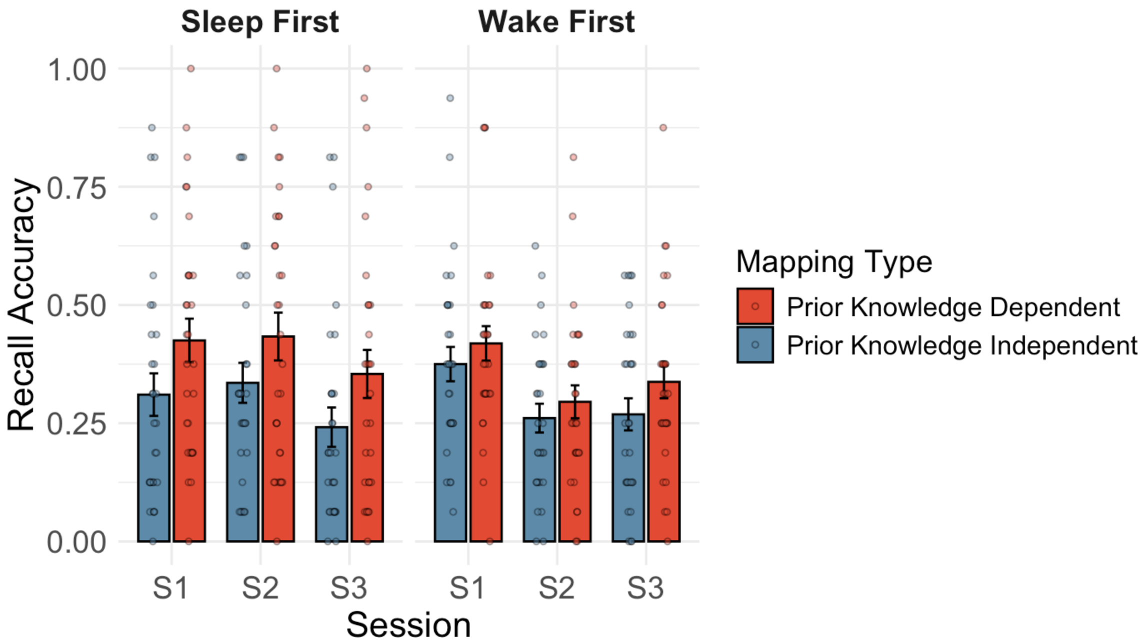

Figure 2.13. Stem recall accuracy (proportion correct) on the Picture Naming task. Error bars represent standard error.

As shown in Table 2.13, there was a significant change in stem accuracy between both Sessions 1 and 2, and Sessions 2 and 3 (S2vsS1, and S3vsS2 contrasts respectively). Stem accuracy for the participants in the Wake First group decreased at Session 2 relative to Session 1, whereas for participants in the Sleep First group accuracy was similar across the two sessions, as shown by a significant S2vsS1 x Group interaction (Table 2.13, Figure 2.13). For participants in both groups, there was an overall decline in stem recall accuracy across the week, as indicated by the significant main effect of the Session 3 vs. Session 2 contrast, and this decline was similar across the two groups (no significant S3vsS2 x Group interaction, Table 2.13).

The main effect of mapping on stem recall accuracy was non-significant, however, a significant interaction between mapping and group was observed (Table 2.13), indicating that the prior knowledge manipulation that involved the systematic mappings in the determiners and suffixes influenced the learning of the stems, and this influence was stronger for the

Sleep First group (Figure 2.13). This influence of mapping was consistent across all three sessions as indicated by the non-significant interactions between mapping, group, and the Session 2 vs. Session1 contrasts and mapping and the Session 3 vs. Session 2 contrasts. No other main effects or interactions were observed (Table 2.13).

Table 2.13 Regression coefficients for the fixed factors using stem recall accuracy in the Picture Naming task as the outcome.

	Estimate	Std. Error	z	p
S2vsS1	**−0.217**	**0.041**	**−5.292**	**<.001**
S3vsS2	**−0.185**	**0.042**	**−4.384**	**<.001**
S2vsS1 × Group	**0.127**	**0.041**	**3.096**	**.002**
S3vsS2 × Group	−0.077	0.042	−1.814	.070
Group	0.053	0.034	1.553	.121
Mapping	0.199	0.112	1.762	.077
Group × Mapping	**0.071**	**0.034**	**2.102**	**.036**
Mapping × S2vsS1	0.010	0.041	0.238	.812
Mapping × S3vsS2	0.044	0.042	1.035	.301
Mapping × S2vsS1 × Group	−0.015	0.041	−0.375	.708
Mapping × S3vsS2 × Group	−0.010	0.042	−0.245	.806

Random effects included in the model: intercepts by items with random slopes for group. This structure represents the maximal model that successfully converged. Significant results are highlighted in bold.

2AFC Recognition. Participants' learning and consolidation of the novel words was also assessed in a recognition task. As illustrated in Figure 2.14, recognition accuracy was high across all three sessions.

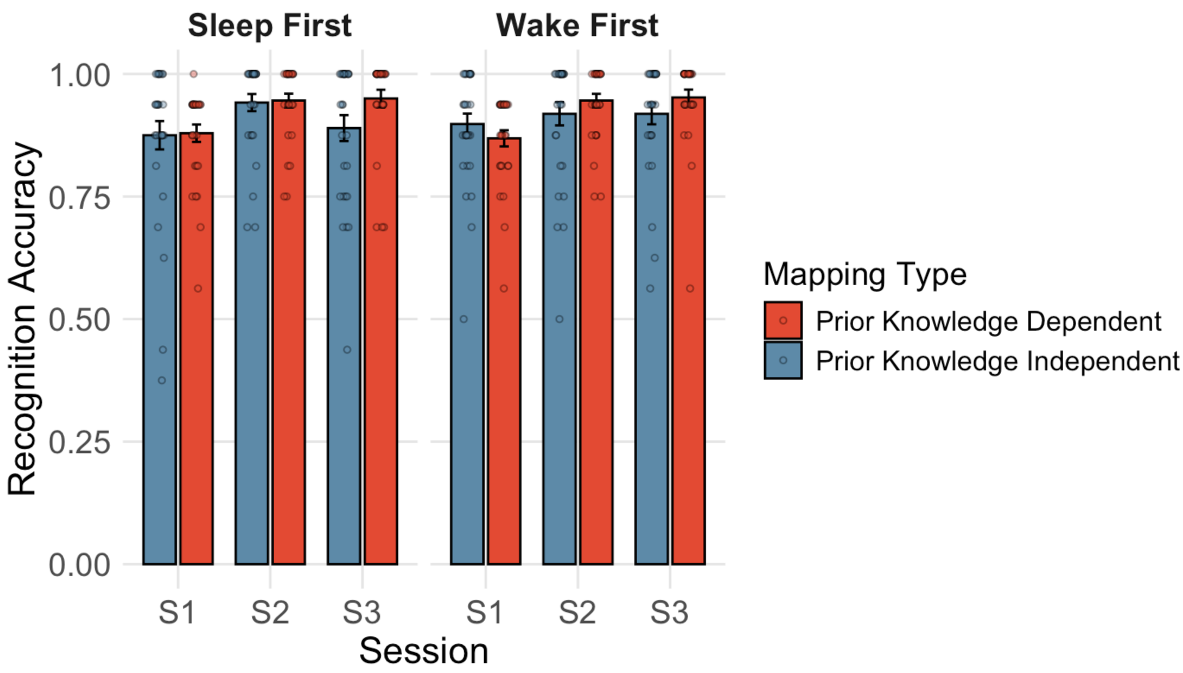

Figure 2.14. Recognition Accuracy on the 2AFC task. Error bars represent standard error.

Participants in both groups and across both mappings showed significant improvements in their word recognition accuracy after a 12-hour consolidation interval, as indicated by the significant main effect of the Session 2 vs. Session 1 contrast (Table 2.14). No other main effects or interactions were significant, indicating that these improvements did not differ between groups or mapping types (Table 2.14).

Table 2.14 Regression coefficients for the fixed factors using recognition accuracy in the 2AFC task as the outcome.

	Estimate	Std. Error	z	p
S2vsS1	**0.407**	**0.183**	**2.224**	**.026**
S3vsS2	0.041	0.118	0.350	.726
S2vsS1 × Group	0.020	0.092	0.220	.826
S3vsS2 × Group	-0.095	0.077	-1.241	.214
Group	0.005	0.180	0.030	.976
Mapping	0.032	0.134	0.236	.814
Mapping × S2vsS1	0.148	0.161	0.919	.358
Mapping × S3vsS2	0.189	0.114	1.660	.097
Mapping × Group	0.030	0.077	0.392	.695
Mapping × S2vsS1 × Group	-0.079	0.072	-1.096	.273
Mapping × S3vsS2 × Group	0.056	0.076	0.729	.466

Note. Random effects included in the model: intercepts by participant and items. For participants, slopes were included for mapping and the S2vsS1 contrasts and for item, slopes for the S2vsS1 and the S3vsS2 contrasts and their interaction were included. This structure represents the maximal model that successfully converged. Significant results are highlighted in bold.

In summary, the analyses examining the influence of sleep and wake-related consolidation and prior knowledge on the learning and long-term retention of the arbitrary components of the novel language revealed several key findings, albeit with some inconsistencies across the two tasks. In the Picture Naming task, participants who slept after training maintained their recall accuracy for the stem components of novel words over the 12-hour consolidation interval, whereas participants who remained awake showed a decrease in recall accuracy of the stems. Both groups then showed an overall decline in recall accuracy across the week. Additionally, the prior knowledge manipulation of the systematic mappings seemed to benefit in the learning of the stems in this task, and particularly in the Sleep-first group. Results from the 2AFC recognition task demonstrated a general improvement in word recognition after a 12-hour consolidation interval, with performance gains observed regardless of whether participants slept or remained awake. Furthermore, this recognition performance was unaffected by the prior knowledge manipulation of the systematic mappings. Taken together, these findings suggest that while sleep-related consolidation and prior knowledge influenced recall accuracy of the stem-meaning (arbitrary) mappings, recognition memory for these mappings did not show a preferential role for sleep, instead both the sleep and the wake groups showed some improvements in their performance on this task after a 12-hour consolidation interval and then no change across the week.

2.5.2.3 Explicit Knowledge

Retrospective explicit knowledge of the underlying systematic mappings (semantic-phonological awareness) was measured using the Explicit Knowledge Questionnaire, administered at the end of Session 3.

As illustrated in Figure 2.15, self-reported explicit knowledge was low overall but numerically higher for the determiners than for suffixes. Determiner and suffix knowledge was assessed in separate linear mixed-effects regressions. For the determiners, there was a significant main effect of mapping, with greater explicit knowledge for the prior knowledge dependent mappings (main effect of mapping in Table 2.15), that did not differ between the groups (no main effect of group, or mapping x group interaction, Table 2.15).

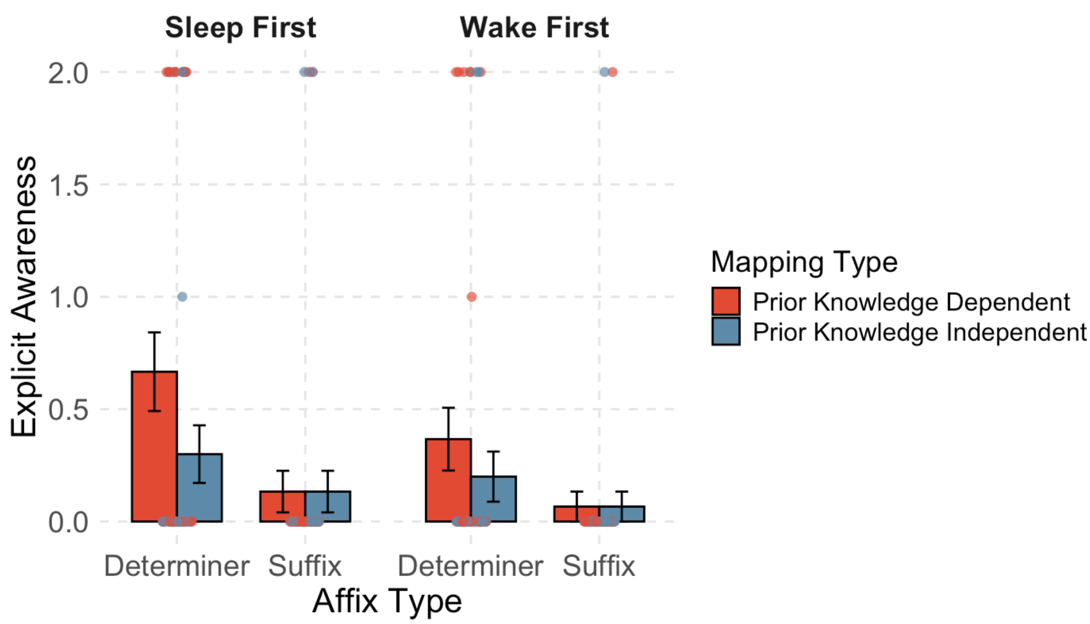

Figure 2.15. Explicit Knowledge Questionnaire: semantic-phonological awareness of the determiner and suffix mappings. Error bars represent standard error.

Table 2.15 Regression coefficients for the fixed factors using self-reported explicit knowledge (semantic-phonological awareness) of the determiner mappings as the outcome.

	Estimate	Std. Error	t	p
Mapping	**0.133**	**0.046**	**2.914**	**.004**
Group	0.100	0.088	1.134	.257
Group x Mapping	0.050	0.046	1.093	.274

Note. Random effects included in the model: intercepts by participant. This structure represents the maximal model that successfully converged. Significant results are highlighted in bold.

For the suffixes, explicit knowledge was very low overall and neither of the main effects or their interaction were significant (Table 2.16).

Table 2.16 Regression coefficients for the fixed factors using self-reported explicit knowledge (semantic-phonological awareness) of the suffix mappings as the outcome.

	Estimate	Std. Error	t	p
Mapping	0.000	0.040	0.000	1.000
Group	0.033	0.040	0.826	.411
Group x Mapping	0.000	0.040	0.000	1.000

Note. No random effects were included in the model. This structure represents the maximal model that successfully converged. Significant results are highlighted in bold.

2.5.2.4 Alertness

To assess any circadian differences in alertness between the groups at each session two measures were used: the Stanford Sleepiness scale (SSS) and a psychomotor vigilance task (PVT). Both were administered at the start of each session.

Using SSS, there was no significant difference in alertness between the two groups at the start of Session 1 ($M_{\text{Sleep First}} = 2.6$, $M_{\text{Wake First}} = 2.37$, $t(46.7) = 1.14$, $p = .258$). At the start of Session 2, participants in the Sleep First group had significantly higher scores than the Wake First group ($M_{\text{Sleep First}} = 2.63$, $M_{\text{Wake First}} = 2.10$, $t(53.9) = 2.19$, $p = .033$), suggesting decreased alertness (higher scores indicate lower levels of alertness). This indicates that the significant group differences in the test tasks outlined above (e.g., higher recall accuracy of stems for the Sleep First group at Session 2), were not likely due to reductions in alertness in the Wake First group relative to the Sleep First group. There were no significant differences in SSS between the two groups at Session 3 ($M_{\text{Sleep First}} = 2.37$, $M_{\text{Wake First}} = 2.20$, $t(57.2) = 0.683$, $p = .497$).

A 3-minute Psychomotor Vigilance Task (PVT; Dignes & Powell, 1985) was also administered at the start of each session to provide a more objective measure of potential differences in fatigue or alertness between the two groups. Although all participants completed the 3-minute PVT, due to a task programming error, the data for 16 participants from the undergraduate sample were lost. Therefore, initial analyses only focused on the 29 participants recruited from Prolific (Sleep-First: n = 15, Wake-First: n = 14). Attentional lapses (e.g., missed or incorrect responses) were used to assess group differences in alertness at the start of each session. No significant differences in the number of attentional lapses between the two groups were found at the start of Session 1 ($M_{Sleep\text{-}First} = 0.80$, $M_{Wake\text{-}First} = 0.92$, $t(26.8) = -0.434$, $p = .667$), Session 2 ($M_{Sleep\text{-}First} = 0.93$, $M_{Wake\text{-}First} = 0.93$, $t(22.0) = 0.017$, $p = .987$) or Session 3 ($M_{Sleep\text{-}First} = 0.73$, $M_{Wake\text{-}First} = 0.64$, $t(23.2) = 0.332$, $p = .743$).

Additional analyses conducted on the entire sample of 44 participants (Sleep-First: n = 28, Wake-First: n = 16), which included all participants recruited from Prolific and the available data from participants recruited from the undergraduate pool similarly revealed no significant differences between groups in attentional lapses at Session 1 ($M_{Sleep\text{-}First} = 0.61$, $M_{Wake\text{-}First} = 0.81$, $t(36.7) = -0.812$, $p = .426$), Session 2 ($M_{Sleep\text{-}First} = 0.79$, $M_{Wake\text{-}First} = 0.88$, $t(23.8) = -0.355$, $p = .725$), or Session 3 ($M_{Sleep\text{-}First} = 0.71$, $M_{Wake\text{-}First} = 0.56$, $t(28.0) = 0.622$, $p = .539$).

2.5.2.5 Sleep Quality

At the end of Session 3, all participants were asked to complete the PSQI (Buysse et al., 1989) to assess general sleep quality over the last month. Global PSQI scores were used to assess differences in sleep quality between the two groups and found no significant differences ($M_{Sleep\ First} = 6.27$, $M_{Wake\ First} = 6.00$, $t(58) = 0.47$, $p = .644$).

2.6 Discussion

We explored the role of prior knowledge and offline consolidation (sleep versus wake) in the learning of arbitrary and systematic components of novel words. Specifically, we examined the role of prior knowledge in the learning and consolidation of the systematic aspects of novel word-referent mappings, which are typically reflected in the grammatical properties of words. In line with predictions outlined in the CLS (McClelland, 2013) and SLIMM (Van Kesteren et al., 2012) models of memory consolidation, we hypothesised that if sleep does play a role in consolidating the systematic components of the novel words, its strongest contribution would be observed for the systematic mappings that were unrelated to prior linguistic knowledge, since novel information unrelated to existing knowledge is thought to rely more heavily on the hippocampus at encoding. Conversely, for mappings related to prior knowledge, we hypothesised that the presence of an existing schema at encoding would facilitate neocortical integration. This would result in better learning for these mappings immediately after training; however, their greater reliance on the neocortex during learning would limit hippocampal engagement, meaning they would be less likely to benefit from hippocampal reactivation during sleep. In addition to the systematic mappings, we also examined the contribution of offline processes over sleep and wake to the consolidation of the arbitrary mappings, reflected in the stem-meaning mappings. For the arbitrary mappings, we predicted a greater benefit of sleep-related consolidation relative to wake, due to the greater reliance of arbitrary knowledge on the hippocampus at encoding and confirmed by a significant body of evidence for the role of sleep in learning novel words (e.g., Schimke et al., 2021).

First, using a non-speeded task, we found that prior knowledge benefitted the initial learning of the systematic mappings. In the Picture Naming task administered at the end of training at Session 1, participants in both groups had better recall accuracy for the systematic mappings (determiners and suffixes) for the items in the prior knowledge dependent condition relative to the prior knowledge independent condition. Next, using both speeded and non-speeded tasks, we examined how offline processes during sleep versus wakefulness might contribute to the consolidation of the systematic mappings. Based on the predictions outlined by the CLS and SLIMM frameworks, we hypothesised that, if sleep does play a role in consolidating the systematic mappings, we would only see evidence of a contribution in the consolidation of mappings unrelated to participants' existing knowledge. To address this, in the non-speeded Picture Naming task, recall accuracy for trained-item determiner and suffix mappings (prior knowledge dependent and independent) was compared between the tests immediately after training, following a 12-hour consolidation period involving either sleep or wake, and again after a 1-week delay. For both the determiners and suffixes, we observed a decline in recall accuracy following the initial 12-hour consolidation interval, and this was true for both types of determiner and suffix mappings: those with the novel words in the prior knowledge dependent condition and those with the novel words in the prior knowledge independent condition. Crucially, this decline in recall accuracy occurred regardless of whether participants slept or remained awake during the interval, suggesting that sleep was not influential in the consolidation of the systematic mappings. When performance was assessed again after one week, recall accuracy continued to decline, and this reduction in performance was again observed for both types of mappings and groups. Thus, in a non-speeded task, we found no evidence supporting a role for sleep in the consolidation of systematic mappings irrespective of their relationship to participants' existing linguistic knowledge.

To further examine the role of offline processes in consolidating the prior knowledge dependent and independent systematic mappings, we also employed a speeded task that required participants to repeat the novel words as quickly and accurately as possible (speeded shadowing). Response times in this task were measured immediately after training, following the 12-hour consolidation period involving either sleep or wake, and again after a 1-week delay, and for the mappings presented in their trained-item contexts (paired with the original word stems encountered during training) as well as in novel contexts (paired with new, untrained word stems). For the speeded shadowing task with the trained items, we observed an overall increase in response times (participants responded more slowly) after a 12-hour consolidation interval, though the extent of this slowing varied depending on group (Sleep First vs. Wake First) and mapping type (prior knowledge dependent vs. independent), that seemed driven by the group difference for the two mappings at Session 1. Unlike the slowing observed after the initial 12-hour consolidation interval, there was some evidence that response times decreased (participants responded faster) across the subsequent week; yet the magnitude of this decrease varied depending on group (Sleep First vs. Wake First) and mapping type (prior knowledge dependent vs. independent). Importantly, however, a follow-up analysis in which we controlled for baseline differences in performance between the sleep and wake groups revealed no differences in response times between the two groups or mapping types, suggesting that the observed differences across sessions were likely driven by group differences in performance at Session 1. Additionally, for mappings presented in novel contexts, where data were only available for Sessions 1 and 2, no differences in response times emerged between groups or mapping types after the 12-hour interval. Taken together, the results of the speeded task corroborate those of the non-speeded Picture Naming task in that we found no evidence of a role for sleep in the consolidation of the systematic mappings, even for those mappings that were unrelated to existing knowledge.

The fact that we used both speeded and non-speeded tasks to measure consolidation of the systematic mappings and found no role for sleep is crucial. One key difference between previous studies that have found a contribution of sleep to grammar learning and those that have not is in the generalisation test used. For example, a study by Mirković and Gaskell (2016), found no beneficial role for sleep relative to wake in the generalisation of novel determiner and suffix mappings in a non-speeded task. By contrast, a study by Tamminen et al. (2012), did find evidence for a role of offline consolidation in the generalisation of novel affixes but only in a speeded task. In a non-speeded task, participants were able to generalise their knowledge of the affix meanings immediately after learning and performance was not further influenced by offline consolidation. As highlighted by Tamminen et al. (2012; 2015), a key distinction between generalisation in speeded and non-speeded tasks is the nature of the representations they test. For instance, when there is opportunity to reflect on knowledge, as in non-speeded tasks, participants can retrieve and combine multiple episodic (i.e., context dependent) memory traces to generate an abstract representation that supports generalisation. Thus, since this form of generalisation is supported by episodic representations it will be available immediately after learning. Conversely, in speeded tasks, generalisation cannot be supported by episodic memory as there is limited opportunity to retrieve and combine multiple episodic memory traces. Consequently, generalisation in speeded tasks is more likely to be supported by context-independent representations (i.e., neocortical knowledge) and therefore, if available, only likely to emerge once neocortical integration has taken place (e.g., after an opportunity for offline consolidation). Thus, the speeded tasks used in the current study, particularly the speeded task with the generalisation items, is not only a purer test of systematic knowledge (compared to the Picture Naming task which could reflect both item-level knowledge as well as generalised knowledge), but also likely more sensitive to sleep-related consolidation changes. Overall, when taking the above into consideration, the fact that we did not find evidence of a role for sleep in our speeded generalisation task further strengthens the conclusion that in the current study it was not influential.

Adding further weight to this conclusion is the fact sleep did not influence consolidation of the prior knowledge independent mappings. According to CLS (McClelland, 2013) and SLIMM (Van Kesteren et al., 2012) models of memory consolidation, when new knowledge is unrelated to or incongruent to existing knowledge it is more likely to be hippocampally dependent at encoding and consequently subject to stronger offline consolidation (i.e., hippocampal reactivation during sleep). By comparison, since prior knowledge related information may be integrated more swiftly with existing, neocortical representations, it is likely to be less susceptible to hippocampal reactivation during sleep. Thus, in the context of the present study, if the hippocampus is involved in the encoding of novel systematic mappings, we should expect to see an influence of offline consolidation, specifically, sleep-related consolidation for the prior knowledge independent mappings. However, this is not what we observed. As outlined above, in the non-speeded Picture Naming task, there was an overall decline in recall accuracy for the determiner and suffix mappings, with no evidence of a beneficial role for sleep even for the mappings that were unrelated to existing knowledge. This was also reflected in the speeded tasks, where we saw no improvements (no speeding up) in response times, for either the prior knowledge dependent or independent mappings after overnight consolidation.

One possible interpretation of the absence of a sleep-related benefit for the consolidation of the prior knowledge independent mappings is that the hippocampus may be generally less involved in the learning of systematic knowledge, particularly in a fully-fledged novel language that combines arbitrary and systematic mappings that might compete for consolidation (Mirković & Gaskell, 2016). This perspective aligns with recent proposals made by Gaskell (2024), which suggest a more graded view of hippocampal and neocortical involvement in word learning, rather than a strictly dichotomous one as proposed in the original Davis and Gaskell (2009) model. According to this updated account, both the hippocampus and neocortex are involved in encoding new words, but their relative engagement depends on the nature of the representations being learned. Specifically, Gaskell (2024) proposes that systematic properties of novel words are particularly well-suited to cortical learning because they exhibit consistency across many learning episodes. Assuming each trial contributes a small amount of cortical learning, the recurrence of systematic properties across training trials allows structure to gradually accumulate in the cortex, ultimately resulting in the formation of a stable, internally coherent schema for these systematic mappings. By contrast, arbitrary properties of new words do not exhibit consistency across learning trials and are therefore schema inconsistent. Drawing on the framework proposed by SLIMM, which predicts that schema-inconsistent information will depend more on hippocampal mediation at encoding, Gaskell (2024) proposed that arbitrary components should have the strongest hippocampal representation after learning and consequently be more likely to show a benefit of sleep. However, for the schema-consistent, systematic components, the bulk of learning will be cortical and therefore they will be less likely to benefit from post-learning sleep. Thus, the lack of a role for sleep in consolidating the prior knowledge independent mappings observed in the current experiment may reflect the generally limited involvement of the hippocampus in encoding systematic properties of novel words.

Nonetheless, an alternative explanation for these findings is that the prior knowledge independent mappings themselves presented suboptimal conditions for hippocampal encoding. For instance, the framework proposed by SLIMM (Van Kesteren et al., 2012) distinguishes between two different types of novelty in learning and makes different predictions about the role of the MTL and by extension, the hippocampus in each. According to SLIMM, the key factor that drives learning by the MTL is the amount of prediction error evoked at the time of encoding. For example, when new information is incongruent to an existing schema, it will produce a strong novelty effect (i.e., high prediction error) that will activate the MTL regions to encode. Conversely, because new, unrelated information does not match any existing schema, a prediction error cannot be evoked, so the MTL is not activated during encoding. Thus, under these conditions, memory is not only expected to be poor, but the limited involvement of the MTL would also mean it is not likely to benefit from hippocampal reactivation during sleep. In the present study, the prior knowledge independent mappings more closely align with the latter type of novelty (unrelated type) as they were not associated with any existing grammatical categories in English and thereby had no connections (whether congruent or incongruent) to existing knowledge. Given this, by SLIMM's account, it is perhaps not surprising that these mappings did not benefit from sleep-related consolidation.

Moreover, the absence of a role for sleep in the consolidation of the prior knowledge independent mappings may also be explained by the fact that participants in the present study learned both arbitrary and systematic mappings within a single artificial language. Previous research that has reported sleep-related benefits for systematic knowledge has typically examined it in isolation (e.g., Batterink et al., 2017; Tamminen et al., 2012; 2015). By contrast, studies that have found no such role have investigated systematic knowledge in the context of a full artificial language that also included arbitrary mappings (e.g., Mirković & Gaskell, 2016). As Mirković and Gaskell (2016) suggest, when learned simultaneously, arbitrary and systematic mappings may compete for consolidation, with arbitrary mappings being prioritised due to their greater reliance on the hippocampus at encoding. Thus, the lack of consolidation benefits for prior knowledge independent systematic mappings in the present study may not reflect an absence of hippocampal involvement altogether, but rather weaker hippocampal encoding relative to arbitrary mappings.

In addition to examining the role of offline processes during sleep in consolidating the systematic mappings, we also investigated whether the prior knowledge related mappings would benefit from their connections to existing knowledge during training. Specifically, based on predictions outlined by the CLS and SLIMM models of memory consolidation, we hypothesised that prior knowledge related mappings would benefit from existing schemas at encoding, facilitating their neocortical integration and resulting in better immediate learning compared to the mappings without connections to existing knowledge. In line with these predictions, and as summarised above, we did observe an initial learning benefit for the determiner and suffix mappings that were related to existing linguistic knowledge in the non-speeded Picture Naming task. Moreover, this prior knowledge benefit was not further influenced by offline consolidation, suggesting that connections to existing knowledge supported the recall of the novel, systematic mappings immediately after learning. These findings are consistent with some previous studies of word learning that have experimentally manipulated a reliance on existing knowledge (e.g., James et al., 2019) and the predictions made by SLIMM, which highlights the benefits of schematic knowledge in learning new, related information.

In contrast to the Picture Naming task, we did not see evidence of a prior knowledge benefit in the speeded shadowing tasks as participants did not show a processing advantage for the prior knowledge dependent mappings, either immediately after learning or after a consolidation opportunity. One possible explanation for this discrepancy between the two tasks may reflect differences in the retrieval processes they engage. For example, as outlined above, a key distinction between the speeded tasks and the non-speeded Picture Naming task is that the speeded tasks limit opportunity for reflection and thereby the engagement of explicit memory processes. By contrast, in the Picture Naming task, participants are better able to use explicit knowledge to support their recall of the mappings. If the superior recall for the prior knowledge related mappings is due to greater engagement of explicit processes at encoding, this may explain why the prior knowledge benefit is not reflected in the speeded tasks. Indeed, although the SLIMM and CLS accounts of memory processing make no predictions about the role of explicit processes in supporting prior knowledge related learning, some experimental findings within the word learning literature suggests that it may play a role. For instance, when trained under explicit instruction, participants in a study by James et al. (2019) showed an immediate benefit for recalling novel words that had connections to prior knowledge. However, when learning was incidental (versus intentional), participants did not show better recall for prior knowledge related words either immediately after learning or after opportunity for offline consolidation (James et al., 2021). The authors suggested that prior knowledge may be less helpful under incidental learning conditions as there is less opportunity to capitalise on existing knowledge. By contrast, in the experiment with explicit training, participants were likely able to draw links between their existing knowledge and new, related knowledge. Such a distinction may explain why we observed a prior knowledge benefit in the Picture Naming task, as this task, unlike the speeded tasks, afforded participants the opportunity to recall connections between the prior knowledge related mappings and their existing knowledge.

In addition to the systematic mappings, we also examined the role of sleep in the consolidation of the arbitrary mappings (i.e., stem-meaning mappings). For these mappings, we hypothesised that their likely greater reliance on the hippocampus during encoding would result in a greater benefit from sleep-related consolidation compared to consolidation during wakefulness. Consistent with this hypothesis, we found evidence supporting a role for sleep in the consolidation of arbitrary mappings in the Picture Naming task. Specifically, the sleep group maintained their knowledge of the stems, whereas the wake group showed a decline in stem recall following the 12-hour consolidation interval. This finding aligns with previous research demonstrating a sleep-related advantage in the recall of novel words compared to an equivalent interval spent awake (e.g., Dumay & Gaskell, 2007; Gais et al., 2006; Kurdziel & Spencer, 2016; see Schimke et al., 2021 for a review). Contrary to what we observed in the Picture Naming task, we did not find evidence of a role for sleep in the recognition task. Instead, both the sleep and the wake groups showed some improvements in their word knowledge after a 12-hour delay. The most likely explanation for these results is that re-exposure to the items during some of the test tasks led to improvements in word recognition for both groups. For instance, participants completed the speeded shadowing task with the trained items before they were tested on their word knowledge in the recognition task. Crucially, this task not only involved re-exposure to the novel words and their semantic referents but also required that participants repeat back the words. Hence, the speeded shadowing served as an additional learning opportunity before word knowledge on the recognition task was assessed.

Although we observed an influence of sleep on the consolidation of arbitrary mappings in the Picture Naming task, this effect appeared short-lived as both the sleep and the wake groups showed an overall decline in their recall of the stem component after a 1-week delay, resulting in similar levels of recall accuracy across groups. This decline in memory for the sleep group contrasts with previous studies that have demonstrated a role for sleep-associated consolidation in novel word learning with continued improvements in recall and recognition over the course of a week (e.g., Henderson et al., 2012; 2013). However, in line with our results, some research has also reported a decline in word recall over time. For example, Tamminen and Gaskell (2013) taught adults new nonwords and their meanings and tested both the explicit recall of the nonword meanings and their integration with existing semantic knowledge after a period of offline consolidation. Similarly to the current findings, it was reported that recall of the nonword meanings declined rather than improved over the course of a week. Despite this, familiar words that were not included in the training were primed by the newly related nonwords, with the strongest priming effects observed one week after training. The authors suggested that the nonwords were integrated with semantic memory, as evidenced by the priming effects but that the episodic representations of their meanings, formed during training, weakened over time, as reflected in the decline in explicit recall. Evidence of an inverse relationship between explicit knowledge of a word form-meaning and its integration into semantic memory was also highlighted in a study by Bakker et al. (2014), who reported delayed lexical competition effects between novel and existing words, which were associated with decreases in episodic recognition. Thus, one possibility is that, in the current study, the decline in recall accuracy after one week reflects a similar weakening of hippocampal-based episodic representations formed during training. However, since we only tested recall and not integration, we cannot determine whether the words were integrated with existing lexical knowledge nor whether sleep-related consolidation influenced this integration.

Nonetheless, the findings presented here provide some evidence in support of the CLS model in that learning of novel arbitrary form-meaning mappings thought to depend on the hippocampal system benefited from the sleep-associated memory consolidation. This was unlike the systematic aspects of the novel form-meaning mappings which did not benefit from an overnight consolidation opportunity. Furthermore, these results provide complementary evidence with the findings by Mirković and Gaskell (2016) who found a benefit of an afternoon nap on the consolidation of arbitrary components of an artificial language but no such benefit for the systematic components. In their study, like the current one, stem-referent mappings provided the arbitrary component while determiner and suffix mappings provided the systematic components. However, those systematic mappings overlapped with participants' existing linguistic knowledge, potentially making them less susceptible to sleep-related consolidation effects. Thus, one possibility was that the null effect of sleep on the systematic components was due to their consistency with prior linguistic knowledge which may have made them less susceptible to sleep-related consolidation effects (McClelland, 2013; VanKesteren et al., 2012). Here, we extended the findings by Mirković and Gaskell (2016) by showing that, even when systematic mappings are unrelated to an existing schema the more arbitrary components still take priority during post-learning sleep and we see no evidence for the role of sleep in the consolidation of systematic mappings.

Nevertheless, interpretation of the findings outlined above should also be considered in light of several limitations of the current experiment. For instance, no a priori power analysis was conducted, which should be considered when interpreting some of the null effects reported above. It is not clear whether the current study was sufficiently powered to detect potentially small or moderate effects of sleep on the consolidation of the systematic mappings.

A second consideration is that running the study online meant that we had little control over participants' activities between testing sessions (i.e., during the offline consolidation intervals). For example, while the scheduled session times and pre-arranged video calls ensured that participants completed their first and second sessions at the intended times, activities during the 12-hour sleep or wake interval could not be monitored directly. Although some checks were implemented at the start of the second session, for instance, participants in the wake group were asked whether they had napped since their first session, and participants in the sleep group were asked about their bedtime the previous night and their sleep duration, these measures relied on self-report. Thus, the lack of an objective measure of participants' behaviour during the 12-hour interval meant that we could not verify that participants in the wake group remained awake throughout the duration of the interval or whether and how participants in the sleep group slept.

In addition, testing participants online meant that there was limited control over the conditions under which they completed the study. For example, although participants were instructed to complete the study in a quiet room free from distractions, there was no way to ensure that they had done so. There was also a risk that participants could have written the words down during training and used this to support their performance in the subsequent test tasks. In an attempt to mitigate some of these limitations, participants' audio-recorded verbal responses on the repetition task were checked by the experimenter to monitor engagement during training. In addition, at the end of the third and final session, participants completed a post-experiment questionnaire in which they were asked whether they had written any of the novel words down during any of the sessions and whether they had attempted to avoid completing any of the tasks. However, a key limitation of these self-report checks is that they relied on participants' honesty.

One other limitation that ought to be considered is the potential weakness of the prior knowledge manipulation of the systematic mappings. Although the prior knowledge dependent mappings were designed to be consistent with existing grammatical knowledge, in that the semantic category assignment of male and female was encoded in the phonological properties of the novel words in a way intended to parallel the encoding of biological sex in pronouns in English, the extent to which these mappings genuinely reflected prior grammatical knowledge may be open to question. For instance, although English encodes biological sex in pronouns, it does not mark gender systematically across determiners or noun endings. As such, the current manipulation may provide only a relatively weak test of whether prior grammatical knowledge influences the initial learning and consolidation of novel systematic mappings. Relatedly, the possibility that the semantic categories associated with the prior knowledge dependent mappings (biological sex) were more salient to participants than those associated with the prior knowledge independent mappings (sea animals vs insects) should also be considered. If so, the prior knowledge advantage observed in the current experiment could also reflect this difference in category salience between the mapping conditions, rather than consistency with existing grammatical knowledge alone.

A final limitation of the current experiment that ought to be considered is the fact that trained novel words were not counterbalanced across the prior knowledge dependent and independent conditions. As a result, we cannot rule out the possibility that any differences observed between the two conditions in the current experiment may, in part, have reflected item-specific properties of particular word-picture pairings, resulting in some pairings in one condition being easier to learn or recall than others. This issue was addressed in the experiments outlined in the following chapter.

In sum, the current study found a beneficial role for sleep in the consolidation of arbitrary form-meaning mappings but no evidence of a role for sleep in the consolidation of the systematic aspects of the mappings, even when these aspects were less reliant on existing cortical knowledge. These findings are consistent with previous studies (Mirković & Gaskell, 2016) and suggest that when learners encounter novel mappings that have both arbitrary and systematic properties, sleep-associated consolidation prioritises the most arbitrary components. One possibility is that the systematic aspects can only benefit from hippocampal reactivation during sleep once arbitrary knowledge has been consolidated. However, this is unlikely to be the case for all types of systematic knowledge. For instance, our results showed better immediate memory for mappings related to prior knowledge. This is in line with theoretical models of memory consolidation such as SLIMM (VanKesteren et al., 2012) which suggest that when new learning overlaps with existing knowledge, it is more likely to be encoded directly within the neocortex, and therefore, not likely to benefit from hippocampal reactivation during sleep.

Chapter 3. The Role of Prior Knowledge and Explicit Awareness in Grammar Learning and Generalisation

3.1 Introduction

Two key theoretical models of memory consolidation, the Complementary Learning Systems Model (CLS; McClelland, 2013) and the Schema-Linked Interactions between Medial prefrontal and Medial temporal regions framework (SLIMM; Van Kesteren et al., 2012), posit that the time course for the consolidation of new knowledge may be influenced by its relation to existing knowledge. According to CLS and SLIMM, prior knowledge related information is likely to benefit from faster consolidation into long term knowledge than when information is unrelated or inconsistent with existing knowledge. Support for this prediction has been evidenced in the word learning literature which has shown that novel word forms that are related to existing word forms are better recalled immediately after learning whereas novel words without such connections to existing word knowledge are better recalled after an offline consolidation opportunity (James et al., 2019). In line with this, Experiment 1, Chapter 2 of this thesis, found greater memory recall immediately after training for novel grammatical mappings that were related to existing knowledge compared to those mappings that were unrelated.

However, whether prior knowledge facilitates learning may depend on the conditions under which it is encoded and retrieved. Although CLS and SLIMM make no predictions about the role of encoding or retrieval conditions in prior knowledge-related learning, experimental evidence suggests they may be a contributing factor. For example, in the study by James et al. (2019), where an immediate recall advantage was observed for prior knowledge related novel words, participants were trained under explicit instruction, which involved repeating novel words aloud and completing multiple-choice quizzes with corrective feedback. In contrast, when learning was incidental rather than intentional, no such benefit emerged, with participants showing no recall advantage for prior knowledge related words either immediately after learning or following a consolidation period (James et al., 2021). Consistent with this, in Experiment 1 we found evidence of a prior knowledge benefit in grammar learning in a cued recall task but no processing advantage for prior knowledge related mappings in a speeded shadowing task. One possible explanation is that, unlike the speeded task, the cued recall task provided participants with the opportunity to explicitly retrieve connections between their existing knowledge and new, related knowledge. Taken together with previous findings (James et al., 2019; 2021), the fact that we only observed evidence of a prior knowledge benefit in a task more likely to support explicit retrieval processes suggests that explicit knowledge may play a crucial role in facilitating the use of prior knowledge during learning.

This interpretation is supported by the broader literature on grammar learning which suggests that both implicit (unconscious) and explicit (conscious) knowledge can develop in parallel and contribute to performance on different tasks, with “direct” tasks, such as recollection-based tasks, reflecting explicit knowledge, and “indirect” tasks, such as speeded tasks, reflecting implicit knowledge (e.g., Batterink et al., 2015; Christiansen, 2019). For instance, one mechanism thought to underpin the acquisition of grammatical regularities is statistical learning (e.g., Aslin & Newport, 2012; Christiansen, 2019; Seidenberg & MacDonald, 2018), which refers to the ability to extract regularities from sensory input, including patterns in how elements co-occur, how frequently they occur, and how reliably one element predicts another (e.g., Batterink et al., 2015; Frost et al., 2015). While statistical learning has traditionally been viewed as an implicit process, recent evidence suggests that explicit knowledge representations may also emerge alongside implicit ones.

Batterink et al. (2015) demonstrated that both implicit and explicit knowledge support the statistical learning of phonological regularities contributing to grammar learning in adults. In this study, participants were exposed to a continuous speech stream composed of six trisyllabic nonsense words (*babupu, bupada, dutaba, patubi, pidabu, tutibu*). These nonsense words were presented in a pseudorandom order, with no pauses or other acoustic cues to word boundaries. As a result, word segmentation had to rely solely on statistical cues in the stream, including transitional probabilities between syllables. Transitional probabilities refer to the likelihood that one element in a sequence will be followed by another, and in continuous speech, these probabilities are typically higher *within* words than *across* word boundaries, making them a useful cue for segmenting words from the speech stream. For example, in the word *babupu*, the syllable *ba* would reliably be followed by *bu*, whereas the transitional probability between *pu* (the last syllable of that word) and the first syllable of the next word in the stream would be much lower, since *pu* could be followed by the first syllable of any of the other five words in the stream (*bu, du, pa, pi,* or *tu*).

After exposure to the speech stream, word segmentation was assessed using two tasks: a recognition task (a direct task) to assess explicit memory for the learned words, and a speeded target detection task (an indirect task) to capture real-time processing of the statistical regularities. To determine whether performance on these tasks reflected implicit or explicit knowledge of the statistical regularities, three different measures were employed across two experiments. To measure explicit knowledge, in Experiment 1 participants were asked to provide a confidence judgment upon completion of the recognition task, and in Experiment 2 an integrated judgement task was used which required participants to classify their recognition responses as ‘remember’, ‘familiar’ or ‘guess’. The confidence judgment was a key indicator of explicit knowledge, following the framework of Dienes and Berry (1997). According to this framework, knowledge is implicit if participants lack awareness (meta-knowledge) of what they have learned, and it is explicit if they show evidence of awareness. For instance, using a direct test of memory such as a recognition tasks, knowledge is inferred to be implicit if participants either report they are guessing but still perform above chance (the guessing criterion), or if their confidence is unrelated to their accuracy (the zero-correlation criterion). Conversely, if participants express greater confidence for accurate responses compared to inaccurate ones thus failing the zero-correlation criterion, knowledge is inferred to be explicit. To assess implicit knowledge additionally, in both experiments reaction times on a speeded target detection task were used to capture real-time processing of the statistical regularities. In this task, faster RTs to more predictable syllables provided a more indirect test of the learning of the regularities. Moreover, in Experiment 2 EEG was recorded during both the recognition and target detection tasks to examine the contribution of explicit and implicit neural processes to performance on these tasks. Specifically, the Late Positive Component (LPC) was measured during the recognition task to assess whether recognition judgments were supported by explicit memory retrieval as larger LPC amplitudes are typically linked to the recollection of previously learned information. For the target detection task, the P300 EEG component, typically associated with the allocation of attentional resources, was examined to assess the predictability of syllable targets as an index of implicit processing.

Using these different measures, Batterink et al. (2015) found a clear dissociation between explicit and implicit contributions to statistical learning. Accuracy on the recognition task positively correlated with confidence ratings in Experiment 1 and ‘remember’ responses in Experiment 2, indicating the role of explicit knowledge. ERP analyses further supported this, showing an enhanced LPC for learned words compared to nonword foils in the recognition task. In addition, evidence of learning was also observed in the target detection task, as reaction times were faster for predictable syllables, suggesting participants had learned the word boundaries. Furthermore, these reaction times correlated with the P300 component. However, recognition accuracy did not correlate with reaction times or the P300 component. Based on these findings, Batterink et al. (2015) suggested that explicit knowledge may develop unprompted during statistical learning, alongside dissociable implicit representations, with implicit learning reflected in real-time processing and explicit knowledge reflected in tasks that allow for conscious retrieval. These findings align with other research which has shown that alongside implicit learning, explicit knowledge may develop and support the learning and generalisation of regularities within a statistical learning paradigm (e.g., Franco et al., 2011; Mirković et al., 2021; Smalle et al., 2018).

The findings reported by Batterink et al. (2015) are particularly relevant to the interpretation of Experiment 1 results, as they demonstrate that explicit and implicit knowledge can be dissociated through different tasks and measures. Their observation that explicit knowledge is primarily captured in direct recognition tasks, while implicit knowledge is reflected in speeded processing measures, parallels our pattern of results where prior knowledge benefits emerged in a cued recall task (a direct measure) but not in a speeded task (an indirect measure). This alignment suggests that the prior knowledge advantage we observed in our previous experiment may reflect enhanced explicit awareness of the grammatical mappings that were consistent with existing knowledge. In line with this, several other studies suggest that explicit knowledge contributions to the acquisition of novel regularities might be more likely to be observed in paradigms that incorporate existing knowledge. For example, Brown et al. (2022) examined how six-year-old children and adults learn and generalise semantic regularities. The study used a semi-artificial language learning paradigm where novel particles co-occurred with real nouns (known to participants), with particle usage linked to the semantic category of the noun: vehicles and animals. For instance, participants heard sentences such as *'glim dog dak'* or *'glim car pag'*, where *dak* and *pag* were novel particles linked specifically to each semantic category (animals and vehicles respectively; glim was a meaningless carrier word). The authors found that successful generalisation of the novel particles to new, untrained nouns was dependent on explicit awareness of the particle-semantic category mappings. Similarly, other studies that have reported an explicit knowledge contribution to the generalisation of novel semantic regularities have used referents and semantic categories that were known to participants (e.g., animals and artifacts, Hickey, 2022; Mirković et al., 2021). In contrast to this, a study by Vujović et al. (2021) which trained adult participants on a fully artificial language with fully novel referents did not find an explicit knowledge contribution to the generalisation of semantic regularities. Based on these findings, Hickey (2022) suggested that the availability of prior knowledge during learning may prompt the engagement of explicit processes; however, when little or no prior knowledge is available, learning may rely more on implicit processes.

3.1.1. The Current Experiments

The two experiments reported in Chapter 3 build on the findings from Experiment 1, using the same artificial language, which incorporates two types of novel grammatical mappings: prior knowledge dependent and prior knowledge independent. Experiment 2 had three primary aims. The first was to replicate our previous finding of a prior knowledge benefit at initial learning, as assessed by the Picture Naming cued recall task administered immediately after training. In line with CLS (McClelland, 2013) and SLIMM (Van Kesteren et al., 2012), and based on the findings from Experiment 1, we hypothesised that there will be a difference in recall accuracy of the two mappings immediately after training, with greater recall accuracy for the prior knowledge dependent mappings.

In addition to this, Experiment 2 also explored the possibility of an explicit knowledge contribution to the observed prior knowledge benefit. As outlined above, previous research that has explored the role of prior knowledge in novel word learning suggests that it may only be useful if learners are able to use explicit strategies to draw links between their existing knowledge and new, related knowledge at encoding (James et al., 2019; 2021). Consistent with this, our findings in Experiment 1 showed a prior knowledge benefit in a cued recall task, which allowed for deliberate retrieval (i.e., a direct measure), but not in a speeded task, which relied on more automatic processing (i.e., an indirect measure). This dissociation raises the possibility that the prior knowledge advantage may be supported by explicit knowledge, which is more likely to influence performance when retrieval conditions allow for conscious access to learned information.

We explore the possibility that explicit knowledge supports the learning of the prior knowledge dependent mappings by assessing participants' explicit knowledge of the novel grammatical mappings and whether awareness is greater for those mappings that are related to existing knowledge. To investigate this, we assessed participants' explicit knowledge of the grammatical mappings using two measures. As in Experiment 1, explicit knowledge of the grammatical mappings was assessed using a retrospective verbal report administered at the end of the experiment. However, in addition to this, the current study introduced a Word-Picture Matching task with an incorporated confidence rating measure (e.g., Batterink et al., 2015; Hickey, 2022) to further examine explicit knowledge contributions. This measure was introduced because verbal reports may not be sensitive enough to capture the full range of explicit knowledge. For example, although informative, their open-ended nature means participants may provide responses that are not always relevant to the knowledge being assessed or may not respond at all (Dienes & Berry, 1997). Moreover, the assumption that knowledge is unconscious simply because learners fail to express what they know has been widely criticised (Rebuschat, 2013), as explicit knowledge is not always verbalisable, making these reports an incomplete measure of awareness (Dienes & Perner, 1999; Shanks & St. John, 1994). Thus, to overcome the limitations of the verbal report, we also employed the Word-Picture Matching task with an integrated confidence rating measure. Confidence ratings provide an alternative way to measure awareness of knowledge (i.e., explicit knowledge) without requiring learners to verbalise what they know. For example, confidence ratings may be used on a grammaticality judgement task by asking a person to judge whether an item is 'grammatical' or 'ungrammatical', and to indicate, for each judgement, how confident they were in their judgement (e.g., guess, somewhat confident, very confident). In the Word-Picture Matching task used in the current experiments, participants were presented with word-picture pairings that either matched (were consistent with) or mismatched (were inconsistent with) the grammatical mappings learned during training and were then asked to make a single integrated judgment, simultaneously indicating how well they thought each word and picture went together and their confidence in that judgment (e.g., 'I guess they go well together', 'I am certain they do not go well together'). These confidence judgements were then used to assess awareness of knowledge using the zero-correlation criterion (Dienes et al., 1995). As described above, according to the zero-correlation criterion, if participants lack explicit knowledge of what they have learned, their confidence will be unrelated to their accuracy; however, if explicit knowledge is present, participants will express greater confidence in their accurate responses compared to inaccurate ones. In other words, there will be a relationship between confidence and accuracy. Following this logic, it was hypothesised that if the prior knowledge benefit is supported by explicit knowledge, then the relationship between confidence and accuracy (i.e., the difference in mean confidence for correct vs.

incorrect responses) will be greater for the prior knowledge dependent mappings compared to the prior knowledge independent mappings.

Finally, Experiment 2 also examined whether participants could generalise their learning of the mappings to new, untrained items, and whether generalisation would be greater for the prior knowledge dependent mappings. In Experiment 1, we found no evidence of a prior knowledge benefit in generalisation using a speeded task. However, as outlined above, if participants are drawing on explicit knowledge to support their learning of the prior knowledge dependent mappings, greater generalisation may be more likely to emerge under conditions that allow for explicit retrieval. To test this, Experiment 2 introduced a non-speeded generalisation task to assess whether participants could generalise their learning when given the opportunity to consciously retrieve and apply that knowledge. Specifically, the same Word-Picture Matching task used to assess explicit knowledge of the mappings was also employed to assess generalisation. Unlike the speeded generalisation task used in Experiment 1, this task included two types of generalisation items: those consistent with the mappings presented during training, and those inconsistent with the trained mappings. The inconsistent items were included to assess whether participants could discriminate between grammatically correct (consistent) and incorrect (inconsistent) mappings. Greater endorsement of consistent items relative to inconsistent items would indicate that participants had successfully generalised the learned grammatical patterns. If the prior knowledge benefit observed in Experiment 1 indeed reflects greater explicit knowledge for these mappings, we hypothesised that participants would demonstrate greater generalisation for the prior knowledge dependent mappings compared to the independent mappings on this test.

Experiment 3 further focused on the generalisation of the prior knowledge dependent mappings. Here, we used the same artificial language and test tasks as in Experiment 2, with one key modification: we increased the number of exemplars presented during training. This change was motivated by previous evidence which has shown that a greater number of exemplars embedding the same regularities in training can promote generalisation by enhancing learners' ability to detect shared features across individual items (e.g., Gomez 2002; Tamminen et al., 2015).

To summarise, Experiment 2 was, in the first instance, concerned with replicating our previous finding of greater recall for prior knowledge dependent grammatical mappings at initial learning. In addition to this, we also sought to explore the possibility of an explicit knowledge contribution to the learning of these prior knowledge dependent mappings. Finally, we investigated whether the prior knowledge advantage extends beyond recall of the trained items to support greater generalisation. Given that the current experiments specifically focused on the initial learning of the mappings, participants in Experiments 2 and 3 completed all training and testing within a single session, and no delayed tests were included.

The key research question these experiments aimed to address is as follows:

1) Is there a prior knowledge benefit in grammar learning and generalisation at initial learning and if so, what is the role of explicit knowledge?

3.2 Experiment 2

3.2.1 Methods

3.2.1.1. Participants

65 adult participants aged between 18 and 40 were recruited from the York St John University undergraduate population and consented to take part in exchange for course credit. All participants were monolingual, native speakers of English with no self-reported history of language, hearing or visual difficulties. Of the 65 recruited, the data of 3 participants were excluded from the analyses entirely after reporting a diagnosis of dyslexia. Thus, the final sample comprised 62 participants. Ethical approval for the study was obtained from the York St John University Cross-School Ethics Committee.

Our target sample size was derived by performing simulations in R using the mixedpower (Kumle, Vo, & Draschkow, 2021) and simr (Green & MacLeod, 2016) packages, with the objective of obtaining 80% power to replicate the prior knowledge benefit to grammar learning observed in Experiment 1. The simulations were based on effect sizes obtained from the Picture Naming task (specifically determiner and suffix recall accuracy) in Experiment 1 ($\beta = 0.262$ for determiner recall; $\beta = 0.409$ for suffix recall). The simulation outputs indicated that determiner recall accuracy provided a more conservative estimate, suggesting that at least 60 participants were required to achieve 80% power.

3.2.1.2. Materials

Training set. The artificial language used in the current experiment was identical to the one used in Experiment 1, incorporating the same two types of mappings: prior knowledge dependent and prior knowledge independent. As in Experiment 1, for the prior knowledge dependent mappings, there were 16 exemplars (8 *tib* and 8 *ked* words) with all *tib...eem* words presented with pictures of female characters and all *ked...ool* words presented alongside pictures of male characters. The prior knowledge independent mappings were also identical to Experiment 1, consisting of 16 exemplars (8 *tib* and 8 *ked* words), with *tib...eem* words presented with pictures of insects and *ked...ool* words with pictures of sea creatures.

However, to control for any unintended associations between the novel words and the picture stimuli, in the current experiment, across the prior knowledge dependent and independent conditions, the trained words were counterbalanced. As an example, for one half of the participants, the word *tib vedeem* was paired with the semantic referent 'airhostess', while for the other half, the word *tib vedeem* was paired with the semantic referent 'ant.'

Generalisation set. The same generalisation set as in Experiment 1 was used, consisting of 24 items evenly divided between prior knowledge dependent and prior knowledge independent mappings. Unlike Experiment 1, in which all generalisation items were consistent with the mappings presented during training, in this experiment, only half of the generalisation items conformed to the trained grammatical patterns (consistent items). The remaining items violated these mappings (inconsistent items). For example, during training, pictures of female characters were always paired with the determiner *tib* and the suffix *eem*, while pictures of male characters were paired with the determiner *ked* and the suffix *ool*. On consistent trials, new images of female characters were also paired with *tib* and *eem*, whereas on inconsistent trials, they were paired with the opposite category's mapping (e.g., *ked* and *ool*). Each of the four semantic categories (female characters, male characters, insects, and sea animals) contained six generalisation items, with three consistent and three inconsistent items per category.

3.2.1.3. Procedure

Experiment 2 followed a similar procedure to Experiment 1, but with three key differences. First, all participants were trained and tested in one session and there were no delayed tests. Second, rather than participating remotely, participants completed the study in a quiet room in the University's Psychology labs. Third, an additional, Word-Picture Matching task was introduced at test, in place of speeded shadowing. As in Experiment 1, all training and test tasks were implemented in the experiment builder platform Gorilla (Anwyl-Irvine et al., 2020). The session lasted approximately 90 minutes overall and participants were given an opportunity to take a break halfway through training (before the third repetition block). The overview of the Procedure is provided in Figure 3.1

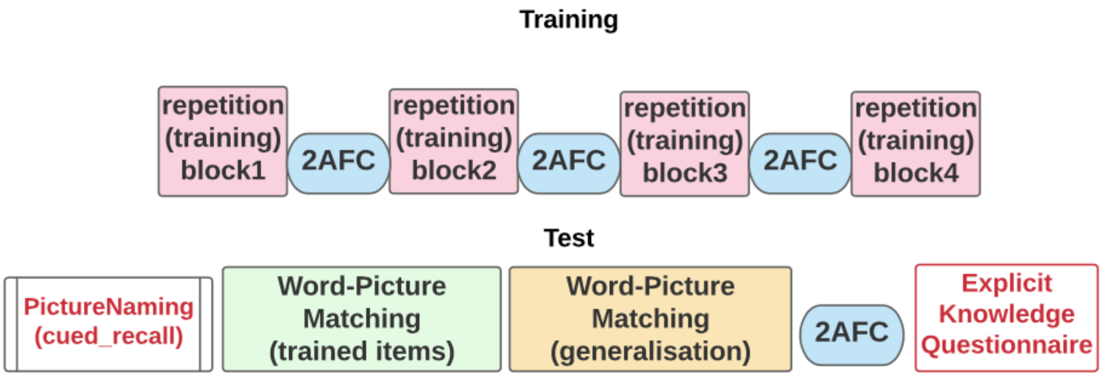

Figure 3.1. Experiment 2 procedure.

3.2.1.4. Tasks

All tasks in the current experiment were the same as the ones used in Experiment 1 with the exception of the Word-Picture Matching task which was introduced to assess explicit knowledge of the underlying grammatical mappings. This task also served to measure participants' ability to generalise the determiner and suffix mappings to new items.

Language Training: Word Repetition

This task was identical to Experiment 1, consisting of four repetition blocks, with each new word presented twice per block (eight repetitions per item in total). The task followed the same procedure as Experiment 1, viewing an image, hearing a made-up word, and repeating the word aloud. As in Experiment 1, three recognition tasks (2AFC) were also administered to increase engagement with the training, one after each training block. As in Experiment 1, these tasks are not considered further. The training phase took approximately 1 hour to complete.

Language Tests: Word Learning

2AFC. This task followed the same procedure as in Experiment 1 in that the block administered as the very last language task before the Explicit Knowledge Questionnaire (Figure 3.1) was used to assess word learning. Accuracy (0, 1) was the outcome measure and was recorded automatically in Gorilla.

Picture Naming. The procedure for this task was identical to that in Experiment 1. Recall accuracy (0,1) for the stem was the outcome measure and responses were scored offline by the experimenter using the same method described in Chapter 2.

Language Tests: Grammar Learning

Picture Naming. As in Experiment 1, Picture Naming was also used to assess grammar learning by measuring participants' recall accuracy of the determiners and suffixes, coded using the same criteria as in Experiment 1.

Word-Picture Matching: a' generalisation. The same Word-Picture Matching task used to test explicit knowledge of the underlying grammatical mappings (see below for details) was also used to assess generalisation. A' (Pallier, 2002) was used as the outcome measure for this task, derived from the endorsement rates for consistent versus inconsistent items. Endorsing all consistent trials (‘hits’) and rejecting all inconsistent trials would yield an A’ of 1.0, while the opposite pattern (endorsing all inconsistent trials and rejecting all consistent ones) would yield an A’ of 0. An equal endorsement rate for consistent and inconsistent trials would produce an A’ score of 0.5. Thus, A’ scores above 0.5 indicate reliable endorsement of consistent items relative to inconsistent items, providing evidence of successful generalisation of the mappings.

Language Tests: Explicit Knowledge

The possibility of an explicit awareness contribution to the learning of the mappings was explored in two ways. First, we used a within-task measure of explicit knowledge, assessing confidence ratings during the Word-Picture Matching task. Second, as in Experiment 1, we examined the role of retrospective, verbally reported explicit knowledge using the Explicit Knowledge Questionnaire.

Word-Picture Matching: Explicit Knowledge. To assess explicit knowledge of the underlying grammatical mappings two versions of the Word-Picture Matching task were administered: one in which participants provided confidence ratings for items they had been presented with during training (Word-Picture Matching Training task) and one in which they provided confidence ratings for new, previously unseen items (Word-Picture Matching Generalisation task).

For the version of this task with the trained items, participants were presented with the 32 trained items across 64 trials in total. On half of the trials, trained novel words were presented with pictures of objects with which they had been trained (matched trials, e.g., *tib zarfeem* = queen) and on the other half trained novel words were presented with trained pictures but of the opposite semantic category (mismatched trials, e.g., *tib zarfeem* = butcher). Mismatched trials were included to allow for a measure of accuracy, as explicit knowledge was assessed by examining whether participants showed greater confidence for correct responses (endorsing matched trials, rejecting mismatched trials) compared to incorrect responses (rejecting matched trials, endorsing mismatched trials).

In the generalisation version of this task, participants were presented with 24 new items, half of which featured word-picture pairings that conformed to the grammatical mappings presented during training (consistent trials), while the other half featured pairings that did not conform to the trained mappings (inconsistent trials). For example, at training, pictures of female characters were always paired with the determiner *tib* and suffix *eem* and pictures of male characters were paired with the determiner *ked* and suffix *ool.* Therefore, on consistent trials, new pictures of female characters were also paired with the determiner *tib* and the suffix *eem*; however, on inconsistent trials, new pictures of female characters were paired with mappings of the opposite semantic category (e.g., determiner *ked* and suffix *ool*).

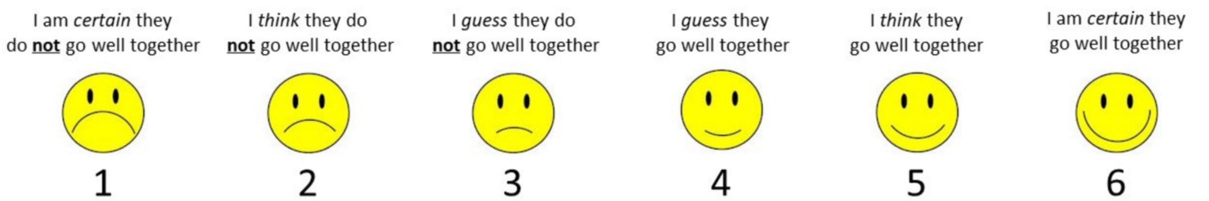

Figure 3.2. Response scale for the Word-Picture Matching training and generalisation tasks.

Before beginning the task, participants were presented with an instruction screen that displayed an image of the response scale (Figure 3.2) and outlined what they would be required to do. In Word-Picture Matching Training task, participants were informed that they would see one of the pictures and one of the words that had been presented to them previously (e.g., during training). For the Word-Picture Matching Generalisation task, they were told that they would see a new picture and hear a new made-up word. In both versions of the task, participants were told to select the smiley face in the response scale that best describes how well the word and the picture go together. Specifically, they were instructed to choose option 6 if they were certain the word and picture went well together, option 5 if they thought they went well together, and option 4 if they were guessing that they went well together. Conversely, they chose option 1 if they were certain the word and picture did not go well together, option 2 if they thought they did not go well together, and option 3 if they were guessing they did not go well together. Thus, participants indicated their word-picture matching decision and their confidence judgement simultaneously by selecting a single response on the scale.

In both versions of this task, each trial started with the presentation of a fixation cross for 500 ms, followed by the presentation of one of the novel words and a picture. Below the picture, the six-point response scale was displayed (see Figure 3.2). The next trial commenced once the participant had made their choice or until the trial timed out after 6000 ms. Since the time taken to reflect on one's knowledge may prompt the emergence of explicit awareness (Dienes, 2007), we limited response times in an attempt minimise the possibility that engaging with the task itself would enhance explicit knowledge. To further address this issue, we also counterbalanced the order of the two Word-Picture Matching tasks such that participants either completed the training block first or the generalisation block first. Two reasons motivated this decision. Firstly, and as noted above, one possibility of administering the training block first was that the task would encourage reflection and in doing so, enhance explicit knowledge of the underlying mappings presented during training. This opportunity for further reflection may then aid performance on the subsequent generalisation task, where participants had to apply their knowledge of the mappings to new items. Secondly, there was also a concern that if the generalisation block was administered first, asking participants to make judgements about new word-picture pairings in an unfamiliar and new task may also negatively impact performance. For both the training and generalisation blocks, each participant had the items presented to them in a random order and the first block of this task always began with two practise trials. Each block took approximately 5 minutes to complete.

Explicit Knowledge Questionnaire

As in Experiment 1, at the end of the experiment, participants were asked to complete the same short questionnaire that assessed their explicit knowledge of the underlying grammatical mappings. The same scoring scheme of the responses as in Experiment 1 was used for data analysis.

3.2.1.5 Data Analyses

All analyses reported below were conducted in R (R Core Team, 2020). The presented analysis follows the pre-registered analysis plan in OSF (pre-registration: https://osf.io/5an7j/).

2AFC

This task was used to assess word learning. To analyse the data, we conducted a mixed-effects binomial regression on accuracy data (0,1). Mapping type was included as a fixed effect and coded using sum contrast coding (prior knowledge dependent = 1, prior knowledge independent = -1). For the random-effects structure, participants and items were included as random intercepts, with a random slope for mapping type within the by-participant random effects. The package *buildmer* (Voeten, 2021) was used to establish the random-effect's structure, starting from the maximal model described above. The final random effects structure with the model that converged for each model is provided in the Results section below.

Picture Naming

This task was used to assess word and grammar learning. A mixed-effects binomial regression was conducted using the *glmer* function in the lme4 package in R (Bates et al., 2015) to analyse the data and assess differences in memory recall for the prior knowledge dependent and independent grammatical mappings (determiners + suffixes) and the words (stems) immediately after training. This analysis was performed for all three binary outcome measures: determiners, stems, and suffixes. Mapping type was included as a fixed effect using sum contrast coding (prior knowledge dependent = 1, prior knowledge independent = -1). The same approach as described above was used to specify the random effects structure.

Word-Picture Matching: a' generalisation

This task was used to assess whether participants had successfully generalised their knowledge of the grammatical mappings to new, untrained items. Specifically, this task was introduced to assess whether greater generalisation for the prior knowledge dependent mappings would emerge under conditions allowing explicit retrieval.

To assess generalisation of the underlying grammatical mappings, endorsement rates for consistent versus inconsistent items were used to compute an A' metric (Pallier, 2002). In this context, endorsements (responses 4, 5, and 6 in Figure 3.2) for consistent items were classified as hits, while endorsements for inconsistent items were classified as false alarms. A' scores ranged from 0 to 1. A score of 0.5 would suggest that a participant had endorsed both consistent trials and inconsistent trials equally. A score below 0.5 would indicate that a participant had endorsed more inconsistent trials than consistent trials. Conversely, A' scores above 0.5 would reflect higher endorsement rates for consistent trials over inconsistent ones and are therefore taken as evidence of successful generalisation of the determiner and suffix mappings to new items.

To evaluate the levels of generalisation for both prior knowledge dependent and prior knowledge independent mappings, we used one-tailed t-tests to compare a' scores against 0.5. These tests assessed whether participants' endorsement rates for consistent items were significantly greater than chance (0.5), indicating successful generalisation. In addition, to assess differences in generalisation performance for the two mapping types, we ran a linear mixed regression. For this, the package *buildmer* (Voeten, 2021) was used to establish the random-effects structure, starting from the maximal model, which included random intercepts for participants and a by-participant slope for mapping type. The final random effects structure with the model that converged is provided in the Results section.

Word-Picture Matching: Explicit knowledge

This task was used to assess participants' explicit knowledge of the underlying grammatical mappings, specifically examining whether explicit awareness was greater for mappings related to existing knowledge compared to those unrelated to existing knowledge.

To analyse the confidence ratings in both the Word-Picture Matching Training and Generalisation tasks, we converted the responses into scores of 0, 1, or 2, following a method described by Channon et al. (2002). Responses of 'I am certain', whether on the reject side of the scale or the endorse side of the scale (Figure 3.2), were classified as high confidence and assigned a score of 2. Responses of 'I think' on either side of the scale were classified as low confidence and assigned a score of 1. Neutral responses ('I guess goes well with' / 'I guess doesn't go well with') were assigned a score of 0.

To assess the possible presence of explicit knowledge, these confidence ratings were then compared against accuracy using the 'zero-correlation' criterion (Dienes et al., 1995). According to this, if explicit knowledge is present, participants will demonstrate greater confidence for correct responses compared to incorrect responses (Dienes, 2007; Rebuschat, 2013). In other words, a relationship between confidence and accuracy would indicate explicit knowledge. If participants lack explicit knowledge of what they have learned, there will be no relationship between their confidence and accuracy.

To derive a measure of accuracy, endorse responses (4, 5 and 6 in Figure 3.2) and reject responses (1, 2 and 3 in Figure 3.2) were used. Endorsing a 'matched' item and rejecting a 'mismatched' item in the Word-Picture Matching task with the trained items would be considered a correct response. Similarly, endorsing a 'consistent' item and rejecting an 'inconsistent' item in the Word-Picture Matching generalisation task would be considered a correct response. On the other hand, rejecting a 'matched/consistent' item and endorsing a 'mismatched/inconsistent' item would be considered incorrect responses. If confidence ratings are higher for correct responses than for incorrect responses, this would indicate a relationship between confidence and accuracy, suggesting the presence of explicit knowledge. However, if there is no significant difference in confidence ratings between correct and incorrect responses, this would indicate no relationship between confidence and accuracy, suggesting an absence of explicit knowledge.

To analyse the Word-Picture Matching data, we conducted a linear mixed regression, with confidence ratings as the outcome measure and accuracy (correct = 1, incorrect = -1) and mapping type (prior knowledge dependent = 1, prior knowledge independent = -1) coded as fixed effects, along with their interaction. Both fixed effects were coded using sum contrast coding. Random effects were included for both participants and items, with random slopes for mapping type within the by-participant random effects. The package *buildmer* (Voeten, 2021) was used to establish the random-effect's structure, starting from the maximal model described above. The final random effects structure with the model that converged for each model is provided in the Results section below.

Explicit Knowledge Questionnaire

This questionnaire was used to assess explicit awareness of the underlying grammatical mappings and was administered at the end of the session. The same coding scheme that was used in Experiment 1 was used, with scores ranging from 0 to 2 assigned separately to determiner mappings (both prior knowledge dependent and independent) and suffix mappings (both prior knowledge dependent and independent). To explore the influence of prior knowledge on participants' explicit awareness of the underlying mappings, we ran two linear mixed models using the lmer function in the lme4 package in R: one with awareness scores for the prior knowledge dependent and independent determiner semantic-phonological mappings and the second with awareness scores for the prior knowledge dependent and independent suffix semantic-phonological mappings. For both models, mapping type (prior knowledge dependent = 1, prior knowledge independent = -1) was entered as a fixed effect with random intercepts for participants.

3.2.3. Results

3.2.3.1 Word Learning

2AFC

Performance on the 2AFC task administered as the very last language task was used to assess how well participants had learned the meaning of the words in the prior knowledge dependent and prior knowledge independent conditions. We also tested whether word learning differed between the two mapping conditions.

Participants' overall performance replicated Experiment 1 in that word knowledge was high across both conditions (Prior Knowledge Dependent: M = 0.94; Prior Knowledge Independent: M = 0.92). The findings here replicate those of Experiment 1 in that there was no significant difference between the mapping conditions (Table 3.1).

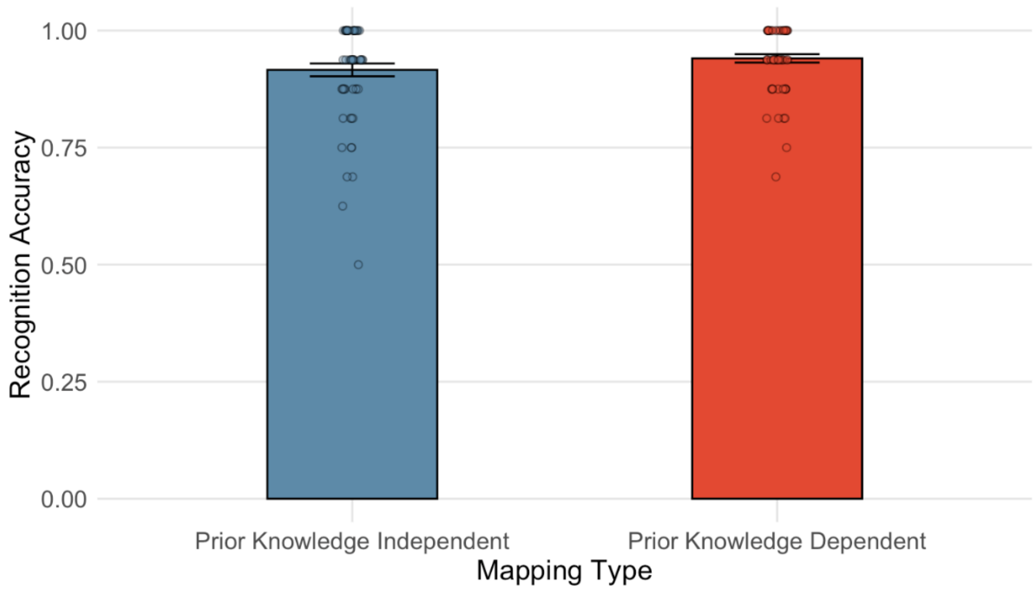

Figure 3.3. Accuracy (proportion correct) on the 2AFC recognition task measuring word learning. Error bars represent the standard error.

Table 3.1. Regression coefficients for the fixed factor using recognition accuracy as a binary outcome measure in the 2AFC word learning task.

	Estimate	Std. Error	z	p
Mapping	0.203	0.145	1.405	.160

Note. Random effects included in the final model that converged: by-participant and by-item intercepts.

Picture Naming: Stems

Word learning was also assessed using the Picture Naming task, and we examined whether performance differed based on mapping type. Overall, recall accuracy of the stems was slightly lower than in Experiment 1. No significant difference between the mapping conditions was observed (Table 3.2).

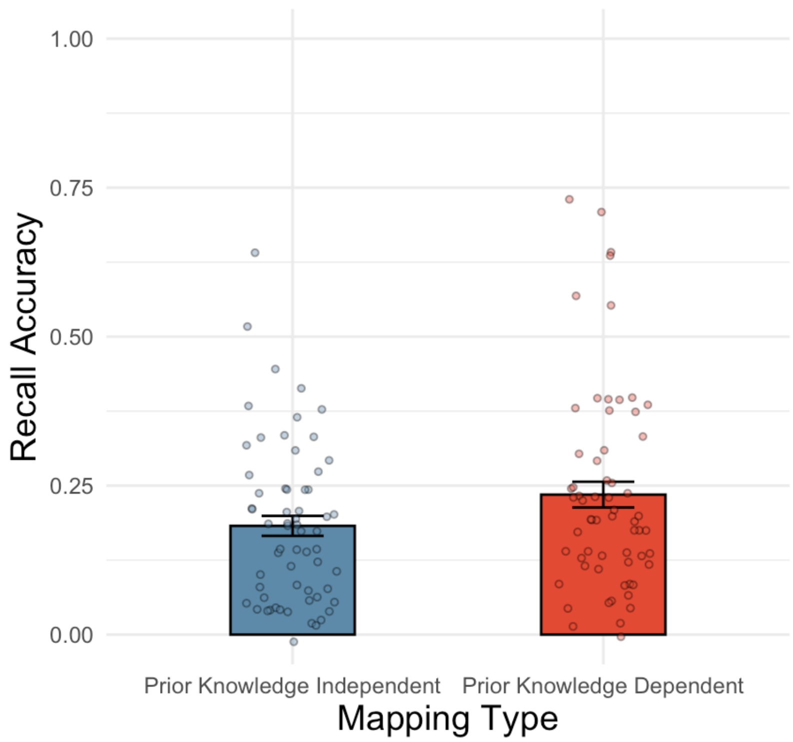

Figure 3.4. Stem recall accuracy (proportion correct) on the Picture Naming task. Error bars represent standard error.

Table 3.2. Regression coefficients for the fixed factors using stem recall accuracy as a binary outcome measure in the Picture Naming task.

	Estimate	Std. Error	z	p
Mapping	0.196	0.130	1.510	.131

Note. Random effects included in the final model that converged: by-participant and by-item intercepts.

3.2.3.2 Grammar Learning

The first key aim of the current experiment was to replicate the prior knowledge benefit observed in Experiment 1 for the grammatical mappings (determiners and suffixes) at initial learning, using the Picture Naming task.

Picture Naming: Determiners

As illustrated in Figure 3.5 and confirmed by the mixed effects analysis (Table 3.3), recall accuracy was greater for determiners in the prior knowledge dependent condition compared to those in the prior knowledge independent condition. This finding replicates the prior knowledge benefit found in Experiment 1.

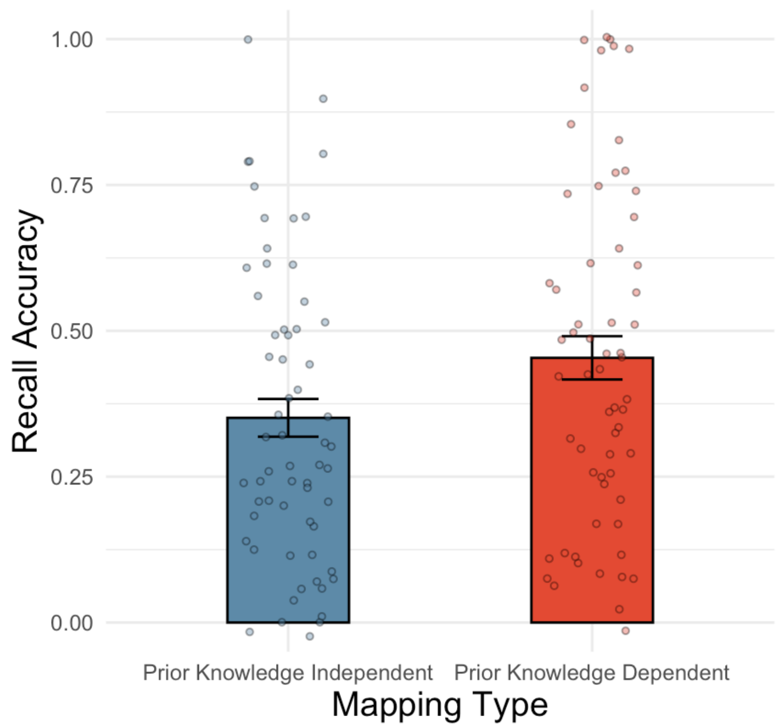

Figure 3.5. Determiner recall accuracy (proportion correct) on the Picture Naming task. Error bars represent standard error.

Table 3.3. Regression coefficients for the fixed factors using determiner recall accuracy as a binary outcome measure in the Picture Naming task.

	Estimate	Std. Error	z	p
Mapping	0.286	0.100	2.87	.004

Note. Random effects included in the final model that converged: by-participant and by-item intercepts, and by-participants random slopes for mapping.

Picture Naming: Suffixes

Similar to Experiment 1, participants showed greater recall accuracy for the suffixes compared to the determiners (prior knowledge dependent mappings: $M_{determiner} = 0.45$, $M_{suffix} = 0.60$; prior knowledge independent mappings: $M_{determiner} = 0.35$, $M_{suffix} = 0.47$). The overall accuracy for both affixes was slightly lower than in Experiment 1.

As illustrated in Figure 3.6, the suffix mappings followed the same pattern as the determiners, in that we see greater recall accuracy for the suffixes in the prior knowledge dependent condition compared to those in the prior knowledge independent condition, as shown by a significant main effect of mapping (see Table 3.4). This finding replicates the prior knowledge benefit observed in Experiment 1.

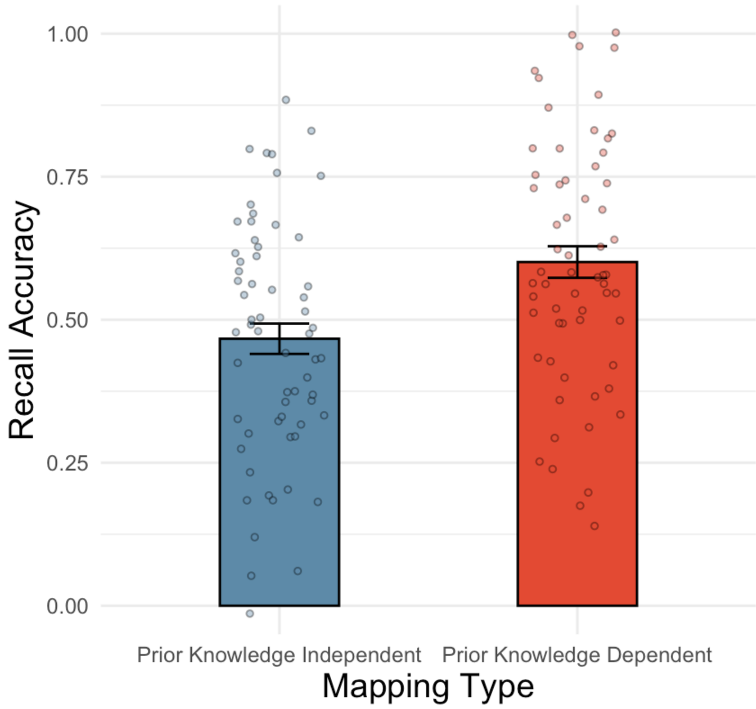

Figure 3.6. Suffix recall accuracy (proportion correct) on the Picture Naming task. Error bars represent standard error.

Table 3.4. Regression coefficients for the fixed factors using suffix recall accuracy as a binary outcome measure in the Picture Naming task.

	Estimate	Std. Error	z	p
Mapping	0.328	0.094	3.48	<.001

Note. Random effects included in the final model that converged: by-participant and by-item intercepts.

Overall, these findings replicate the findings observed in Experiment 1 in that participants show better memory for the prior knowledge dependent determiner and suffix mappings relative to the prior knowledge independent mappings, immediately after learning.

Word-Picture Matching: a' generalisation

In the current study, we also assessed participants' ability to generalise the grammatical mappings in a non-speeded task. Of key interest was whether participants would show greater generalisation for the prior knowledge dependent mappings under conditions that allow for explicit retrieval.

To assess the extent to which participants were able to generalise the prior knowledge dependent and independent mappings to new, previously unseen items, the a' metric was used, with a' scores above 0.5 interpreted as evidence of the ability to generalise. First, we examined whether a' scores were greater than 0.5 in both the prior knowledge dependent and prior knowledge independent conditions. Second, we compared differences in performance between the two mapping types.

One tailed t-tests showed that participants were able to generalise the mappings in the prior knowledge dependent condition (Figure 3.7; t (61) = 2.95, p = .002) and in the prior knowledge independent condition (Figure 3.7; t (61) = 2.62, p = .006). However, there was no evidence that prior knowledge influenced generalisation as indicated by the non-significant main effect of mapping (Table 3.5).

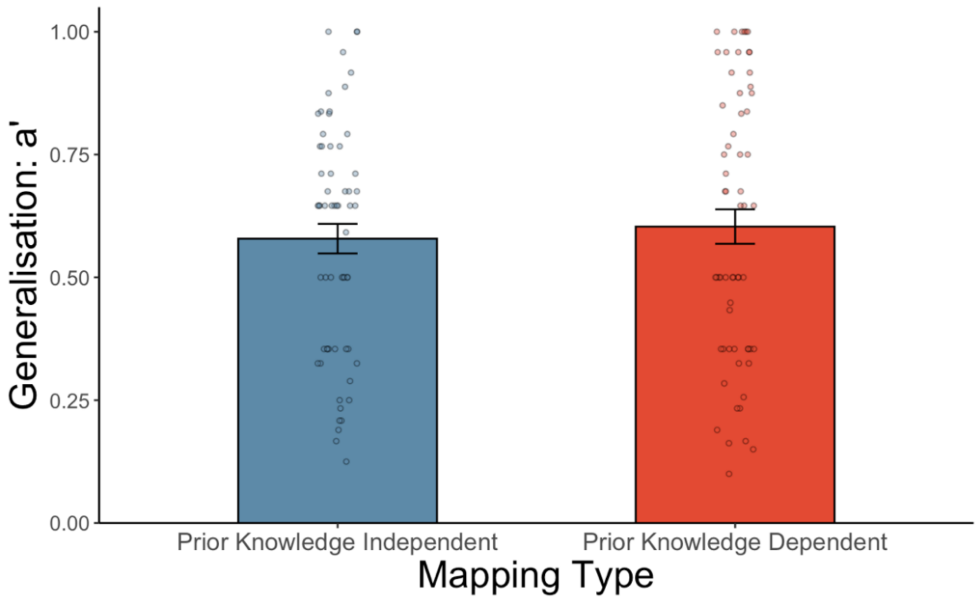

Figure 3.7. Generalisation performance for the determiner and suffix prior knowledge dependent and independent mappings. Error bars represent standard error.

Table 3.5 Regression coefficients for the fixed factors using a' scores as the outcome measure in the Word-Picture Matching task.

	Estimate	Std. Error	t	p
Mapping	0.012	0.021	0.572	.567

Note. Random effects included in the final model that converged: intercepts by participant.

These findings confirm that participants were able to generalise the underlying grammatical mappings in both conditions, albeit to a relatively weak degree ($M_{\text{prior knowledge dependent}}$ = 0.60; $M_{\text{prior knowledge independent}}$ = 0.58). Unlike the Picture Naming task with the trained items where we see evidence of a prior knowledge benefit in recall, participants were not better at generalising the mappings in the prior knowledge dependent condition compared to the mappings in the prior knowledge independent condition.

3.2.3.3. Explicit Knowledge

A second aim of the current study was to explore the possibility of an explicit knowledge contribution to the prior knowledge benefit observed in Experiment 1. In the following section, we first provide the results from the Word-Picture Matching task which provided a within task measure of explicit knowledge of the mappings in their trained item contexts and in new, untrained contexts. Second, as in our previous experiments, we assessed retrospective explicit knowledge using the 'Explicit Knowledge Questionnaire', which was administered at the end of the experimental session.

Word-Picture Matching: Explicit knowledge

To assess the use of explicit knowledge within the Word-Picture Matching training and generalisation tasks, confidence ratings from the six-point response scale were used in a zero-correlation analysis.

Word-Picture Matching: trained items

A linear mixed regression with confidence ratings for the trained items revealed a main effect of accuracy (Table 3.6), with significantly higher confidence ratings for correct compared to incorrect responses. A significant main effect of mapping was also found (Table 3.6, Figure 3.8), with higher confidence ratings for the mappings in the prior knowledge dependent condition compared to the mappings in the prior knowledge independent condition. However, the interaction between accuracy and mapping was not significant, indicating that the relationship between confidence and accuracy did not differ by mapping type. These results suggest that, while participants demonstrated explicit awareness for the grammatical mappings in their trained item contexts, this awareness was not greater for the mappings in the prior knowledge dependent condition.

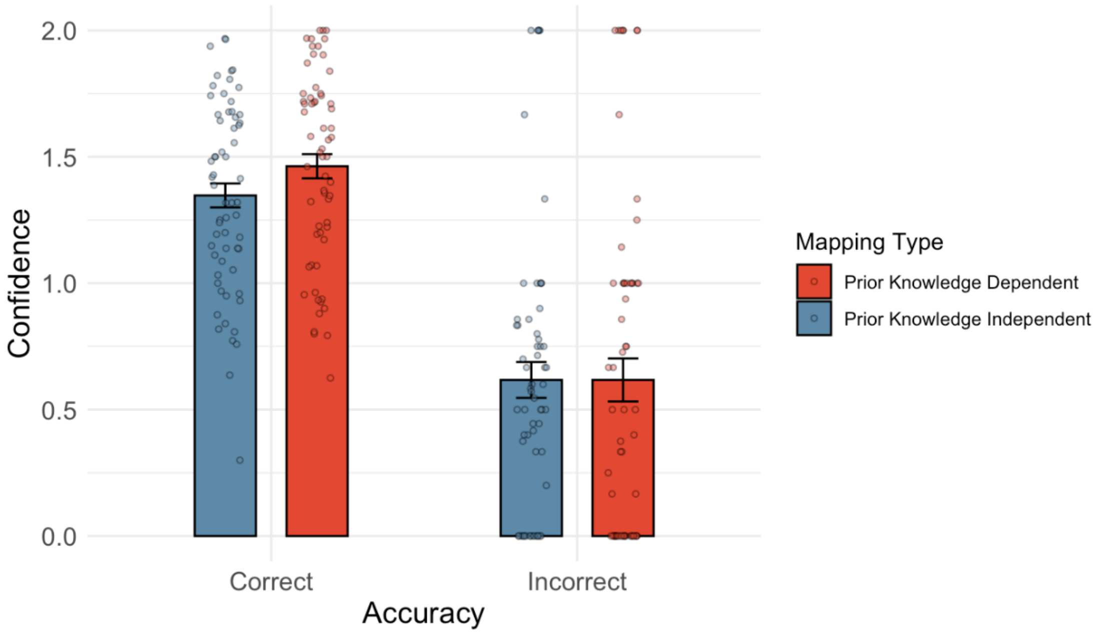

Figure 3.8. Explicit Knowledge: Mean confidence ratings for correct and incorrect responses in the Word-Picture Matching training task. Error bars represent standard error.

Table 3.6. Regression coefficients for the fixed factors using confidence ratings in the Word-Picture Matching training task as the outcome.

	Estimate	Std. Error	t	p
Accuracy	**0.267**	**0.017**	**15.46**	**<.001**
Mapping	**0.060**	**0.023**	**2.55**	**.011**
Accuracy × Mapping	-0.003	0.017	-0.15	.877

Note. Random effects included in the final model that converged: by-participant and by-item intercepts. For participants, random slopes were included for mapping. Significant results are highlighted in bold.

Word-Picture Matching: generalisation

For the generalisation items, confidence ratings did not differ significantly between correct and incorrect responses as no main effect of accuracy was observed (Table 3.7, Figure 3.9). There was, however, a main effect of mapping (Table 3.7), with overall higher confidence ratings for mappings in the prior knowledge dependent than the prior knowledge independent condition, but no significant interaction between mapping and accuracy. This suggests that, while participants were more confident in their responses in the prior knowledge dependent condition, their confidence was not related to accuracy in this task.

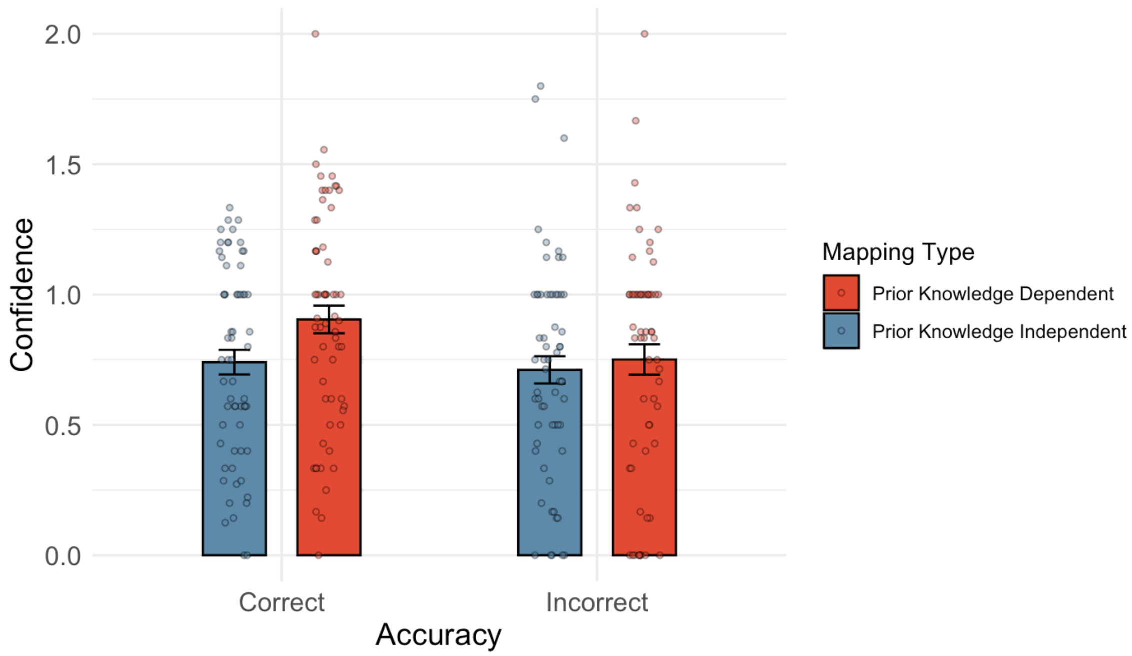

Figure 3.9. Explicit Knowledge: Mean confidence ratings for correct and incorrect responses in the Word-Picture Matching generalisation task. Error bars represent standard error.

Table 3.7. Regression coefficients for the fixed factors using confidence ratings in the Word-Picture Matching generalisation task as the outcome.

	Estimate	Std. Error	t	p
Accuracy	0.019	0.019	1.00	.318
Mapping	**0.067**	**0.022**	**3.08**	**.002**
Accuracy × Mapping	0.026	0.018	1.44	.151

Note. Random effects included in the final model that converged: by-participant and by-item intercepts. For participants, random slopes were included for mapping. Significant results are highlighted in bold.

Taken together, these findings suggest that while participants demonstrated explicit awareness of the mappings in the trained context, this awareness did not extend to novel contexts, as confidence ratings were not linked to accuracy in the generalisation task. Additionally, while confidence ratings were overall greater for the prior knowledge dependent than prior knowledge independent mappings across both trained and new items, there was no evidence that this was due to greater explicit knowledge in this condition.

Explicit Knowledge Questionnaire

This questionnaire assessed retrospective explicit knowledge of the underlying grammatical mappings and was administered at the end of the experimental session. Responses were scored separately for the determiner (both prior knowledge dependent and independent) and suffix mappings (both prior knowledge dependent and independent).

Figure 3.10 illustrates the level of explicit knowledge of the determiner and suffix mappings in the two conditions. The explicit knowledge reported at the end of the experiment was low overall but numerically higher for determiners than for suffixes.

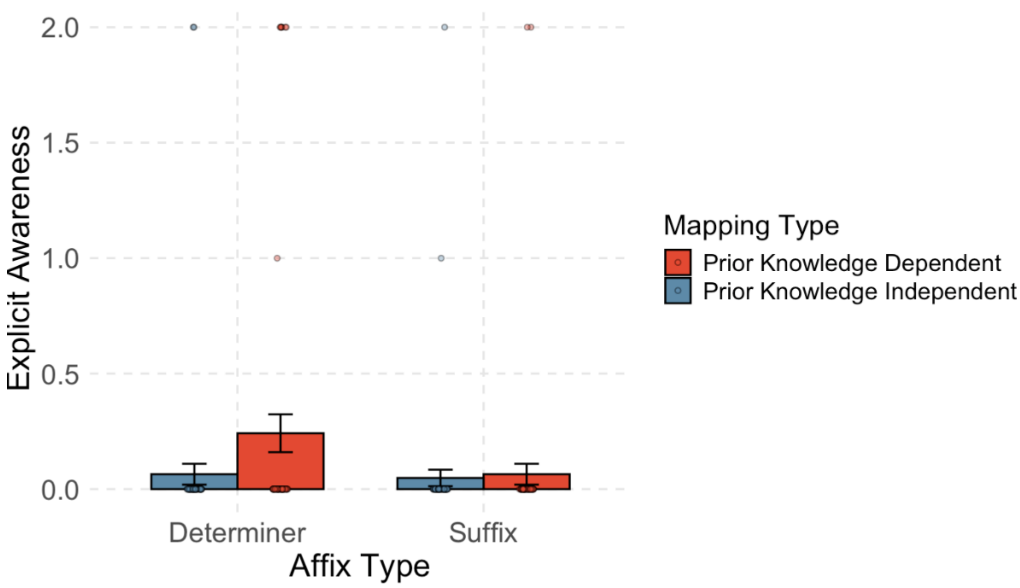

Figure 3.10. Levels of explicit awareness of the determiners and suffixes in the Explicit Knowledge Questionnaire. Error bars represent standard error.

A mixed-effect analyses revealed a significant main effect of mapping for the determiners, with greater levels of explicit knowledge for the prior knowledge dependent mappings (Table 3.8). For the suffixes, explicit knowledge was very low overall, and the main effect of mapping was not significant (Table 3.9).

Table 3.8. Regression coefficients for the fixed factor using determiner responses in the Explicit Knowledge Questionnaire as the outcome.

	Estimate	Std. Error	t	p
Mapping	**0.089**	**0.035**	**2.500**	**.012**

Note. Random effects included in the final model that converged: by-participant intercepts.

Table 3.9. Regression coefficients for the fixed factor using suffix responses in the Explicit Knowledge Questionnaire as the outcome.

	Estimate	Std. Error	t	p
Mapping	0.008	0.008	1.000	.317

Note. Random effects included in the final model that converged: by-participant intercepts.

3.2.4 Discussion

In Experiment 2, we further explored the role of prior knowledge in learning novel grammatical mappings. Our first aim was to replicate Experiment 1, which found greater recall, immediately after training, for mappings that were related to participants' existing knowledge compared to those that were unrelated. In addition to this, Experiment 2 also examined whether participants could generalise their learning of the mappings to new, untrained items in a non-speeded task, and whether generalisation would be greater for the mappings in the prior knowledge dependent condition. The final aim was to explore the possibility of an explicit knowledge contribution to the observed prior knowledge benefits. Specifically, we tested participants' explicit knowledge of the underlying grammatical mappings using a Word-Picture Matching task with an incorporated confidence rating measure in addition to a retrospective verbal report measure.

Overall, the results of Experiment 2 support and extend those of Experiment 1. We replicated the prior knowledge benefit in the Picture Naming cued recall task, with participants demonstrating greater recall for the determiner and suffix mappings in the prior knowledge dependent condition compared to the independent condition. This finding offers support for the theoretical predictions outlined by the CLS (McClelland, 2013) and SLIMM (Van Kesteren et al., 2012) frameworks, which posit that the time course for the learning and consolidation of new knowledge is influenced by its relation to existing knowledge. Specifically, these accounts suggest that prior knowledge related information is likely to benefit from connections to existing knowledge during learning, resulting in faster consolidation into long-term memory networks than when the new information is unrelated. Thus, the present finding provides further evidence that prior knowledge can support the initial learning of novel grammatical mappings.

In addition, we tested generalisation of the mappings, with a key question being whether participants would show greater generalisation for the prior knowledge dependent mappings. In Experiment 1, we found no evidence of a prior knowledge benefit in generalisation when using a speeded task. It was hypothesised that retrieval conditions in different tasks might play a role in whether generalisation may be influenced by prior knowledge. Specifically, a non-speeded task affords participants the opportunity to recall connections between new mappings and their related knowledge, whereas a speeded task does not. If participants relied on explicit knowledge to support their learning of prior knowledge dependent mappings, such benefits might be more likely to emerge under conditions that allow for conscious retrieval, as in word form learning studies (e.g., James et al., 2019). To test this, Experiment 2 introduced a non-speeded generalisation task to assess whether participants could apply their learning in these conditions. The results of the non-speeded generalisation task revealed that although participants showed some evidence of generalisation of the determiner and suffix mappings to new, previously unseen items, we found no evidence of a generalisation advantage for the prior knowledge dependent mappings. Thus, these findings are consistent with those in Experiment 1 in that, while we found evidence that prior knowledge supported the initial *learning* of grammatical mappings, this benefit did not extend to generalising this knowledge to new contexts.

We also directly assessed participants' explicit knowledge of the grammatical mappings. Importantly, this knowledge could have only emerged during training as the grammatical mappings were never mentioned to participants before or during the experiment. Explicit knowledge was assessed using two measures: a retrospective verbal report, which was also used in Experiment 1 and required participants to self-report their awareness of the mappings at the end of the experimental session, and a Word-Picture Matching task with an integrated confidence rating measure administered immediately after training. As previously mentioned, this task was included as an alternative, more sensitive measure to overcome the limitations of the verbal report. For example, the open-ended nature of verbal reports allows participants to subjectively determine which aspects of their knowledge to report. Consequently, they may fail to mention explicit grammatical knowledge they possess if they deem it irrelevant to the task (Rebuschat, 2013). Thus, Word-Picture Matching served as a more sensitive, within-task measure of explicit knowledge.

Consistent with Experiment 1, self-reported explicit knowledge in the retrospective verbal report was low overall, but participants demonstrated greater knowledge for the prior knowledge determiner mappings. No differences in awareness between the two mapping types were observed for the suffixes, and knowledge of suffixes was lower overall compared to determiners. Notably, this same pattern of results was found across two different time points: in Experiment 2, the retrospective verbal report was administered immediately after training, whereas in Experiment 1 it was administered at the end of the final experimental session a week after training.

The Word-Picture Matching task also showed evidence for a role of explicit knowledge when the mappings were presented in their trained item contexts, with participants demonstrating higher overall confidence ratings for correct compared to incorrect responses. However, contrary to our hypothesis, there was no difference in confidence ratings between the prior knowledge dependent and independent mappings, suggesting that explicit knowledge was not greater for mappings supported by prior knowledge in this task.

We found no evidence for a role of explicit knowledge in generalisation, as confidence ratings were unrelated to accuracy in the Word-Picture Matching generalisation task. This is consistent with the findings using a’ scores where despite some evidence of successful generalisation, it did not differ between the prior knowledge dependent and independent mappings.

Overall, the evidence of explicit knowledge in the Word-Picture Matching task with the trained but not the generalisation items suggests that explicit knowledge in this task may have been driven by item-specific familiarity rather than knowledge of the underlying, potentially more abstract grammatical mappings. For example, participants may have relied on explicit memory for previously encountered items in the training task, whereas explicit knowledge in the generalisation task, which used novel items, would more likely reflect awareness of the grammatical mappings themselves rather than familiarity with individual items.

Together, the findings from Experiment 2 are broadly in line with those of Experiment 1 in that, immediately after learning, participants showed better recall for the prior knowledge dependent mappings in the Picture Naming task. This benefit was only found for grammatical mappings, and not in the arbitrary stem-meaning mappings. As in Experiment 1, the prior knowledge benefit did not extend to generalising the grammatical mappings to new contexts, even when retrieval conditions were designed to support explicit access to the learned mappings. This, combined with the lack of strong evidence for a difference in explicit knowledge between the two prior knowledge conditions in the current experiment, challenges our earlier interpretation that the absence of a prior knowledge benefit in the speeded generalisation task in Experiment 1 might have been due to differences in retrieval demands across tasks. Nonetheless, in the current experiment, generalisation was weak for both mapping types, which may have reduced our ability to detect any differences between the two conditions. This issue was addressed in Experiment 3.

3.3 Experiment 3

3.3.1 Introduction

Experiment 3 aimed to enhance participants' overall generalisation of the new mappings and test whether the prior knowledge benefit that supports learning of the mappings can also facilitate greater generalisation.

One factor that has been shown to influence the generalisation of grammatical mappings is contextual diversity. For example, Tamminen et al. (2015) showed that, when learning a novel morphology, increasing the number of word contexts in which a novel morpheme appears during training can support generalisation. In this study, participants were taught novel words made up of an existing stem and a novel affix (e.g., *sleepafe*), with each affix consistently modifying the meaning of its stem in a semantically transparent way (e.g., "sleepafe is a participant in a study about the effects of sleep"). Generalisation of the affix meanings to new, untrained stems was then tested using a sentence congruency task. In this task, participants were presented with sentence frames followed by a semantically congruent or incongruent final word and were asked to read the sentence-final word aloud as quickly as possible. For example, if participants had learned that the affix *-nule* referred to a person, then the sentence "It was an honour to be visited by the sandnule" would be semantically congruent, whereas "The company had just relocated to a peachnule" would be incongruent. Generalisation on this task was assessed in two ways. First, in a speeded reading-aloud component, the time taken to read aloud the sentence-final target word was measured when it appeared in a semantically congruent versus a semantically incongruent sentence. Tamminen and colleagues suggested that if participants had consolidated the novel affix meaning, then their processing of the sentence-final target word would be faster in a semantically congruent context relative to an incongruent one (i.e., a congruency effect would be observed). In addition to this, participants were asked to make a non-speeded, explicit judgment about the semantic congruency of the sentence.

In one of a series of experiments using the same items, Tamminen et al. (2015) manipulated the number of stem contexts with which the novel affixes appeared and tested generalisation. For example, during training, half of the affixes appeared with eight different stems (high contextual diversity) and the remaining half appeared with only two (low contextual diversity). The findings showed that although contextual diversity did not influence generalisation in the explicit judgement component of the task, it did influence the size of the congruency effect in the reading aloud component. Specifically, a significant congruency effect (faster reading times for semantically congruent than incongruent sentence-final words) was observed only for affixes that had been trained in high-diversity contexts. No such effect was found for affixes trained with fewer stem contexts, suggesting that greater contextual diversity supported the formation of more generalisable affix representations.

Along similar lines, other research has also shown that increasing the variability in exemplars during training can enhance the generalisation of phonological distributional regularities. For instance, Gómez (2002) trained adults and 18-month-old infants on an artificial language structured with non-adjacent dependencies between the first and third elements in a sequence (an aXb structure). In this structure, the 'a' and 'b' elements formed non-adjacent dependencies signalling “grammatical” categories, separated by a variable middle stem 'X' (e.g., aXb – *pel-kicey-rud*; cf. *is* run*ning, is* walk*ing, is* talk*ing* in English). To assess whether increasing variability of the middle elements would enhance learning of the non-adjacent regularities, Gomez (2002) systematically manipulated the number of unique middle elements presented during training. The findings showed that as variability (number of stems) increased, participants demonstrated improved learning of the non-adjacent dependencies. Gómez (2002) suggested that greater variability in the middle elements enhances learning by reinforcing the consistency of non-adjacent dependencies, ultimately improving learners' ability to recognise grammatical patterns.

Based on the findings of Tamminen et al. (2015) and Gómez (2002), Experiment 3 aimed to enhance generalisation by increasing exemplar variability (i.e., the number of stems) during training. By doing so, we sought to determine whether this manipulation would lead to improved generalisation performance relative to Experiment 2 and, if so, allow us to observe any benefits for the prior knowledge dependent mappings relative to the prior knowledge independent mappings.

In addition to this, we considered the possibility that increasing exemplar variability during training might promote explicit awareness of the grammatical mappings. Therefore, explicit knowledge was again assessed using the Word-Picture Matching task with integrated confidence ratings, as well as a retrospective verbal report administered at the end of the experimental session.

Finally, we also administered a new two-alternative forced choice (2AFC) task to assess participants' recognition of the images in both the prior knowledge dependent and prior knowledge independent conditions for the trained items. The purpose of this was to ensure that the prior knowledge benefit observed for the trained items in Experiments 1 and 2 was not simply due to images within a given semantic category being more recognisable (e.g., pictures of male and female characters being easier to recognise than those of sea animals or insects).

To summarise, the key research question Experiment 3 aimed to address is as follows:

1) Will increasing the number of exemplars (i.e., stems) during training enhance grammar learning and generalisation, and if so, will the prior knowledge dependent

mappings show greater generalisation than those that are less dependent on prior knowledge?

3.3.2 Methods

3.3.2.1 Participants

A total of 58 participants (aged 18–40) were recruited from Prolific, an online participant recruitment platform. All participants were self-reported monolingual, native English speakers with no history of language, hearing, or visual impairments. To ensure data quality, only participants with a Prolific reputation score of at least 95% were eligible. Participants received £10 for their participation, with the opportunity to earn a £3.00 bonus for achieving at least 95% accuracy on the 2AFC word learning task (described in more detail below). Ethical approval for this study was granted by the York St John University Cross-School Ethics Committee.

A priori power analysis was conducted to determine sample size using effect size from the generalisation task in Experiment 3, as this was the primary focus of the current experiment. The power analysis was based on effect sizes derived from the a′ generalisation task (Cohen's d = 0.37 for prior knowledge dependent mappings; d = 0.33 for prior knowledge independent mappings). Using the smaller effect size (d = 0.33) as a conservative estimate, we determined with the pwr.t.test function in R that a sample size of 58 participants would provide 80% power at $\alpha = .05$.

3.3.2.2 Materials

Training set. In the current study, we increased the training set of the artificial language from 32 words in Experiments 1 and 2 to 48 words. The overall structure of the language remained the same as in the previous experiments, where *tib…eem* words were paired with images of female characters and insects, while *ked…ool* words were paired with images of male characters and sea animals. The primary difference in this study was the number of stems included. Specifically, in each prior knowledge condition, the number of exemplars was increased from 16 in Experiments 1 and 2 to 24 in Experiment 3. The additional stems and pictures were selected from the items used in the generalisation test in Experiment 1 and 2.

Generalisation set. We used the remainder of the generalisation items (8) as the generalisation set. The structure remained the same as in Experiment 2, with both consistent and inconsistent items, but instead of six generalisation items per semantic category, the current study included two items per category, one consistent and one inconsistent.

3.3.2.3 Procedure

All participants followed the same training and testing protocol as in Experiment 2, with three small differences. First, an additional 2AFC image recognition task was administered to assess recognition of the trained items. Second, unlike Experiments 1 and 2, in the word-learning 2AFC test task, participants received feedback on their performance, informing them whether they had scored above 95% and thus qualified for the £3.00 bonus payment. To ensure that this feedback did not negatively influence performance or motivation on the other tasks, this task was moved to the very end of the experiment (see Figure 3.11 for an overview of the procedure). Finally, unlike Experiment 2, participants in the current study were recruited via Prolific and completed the experiment remotely.

As in previous experiments, all training and test tasks were implemented in the Gorilla experiment builder platform (Anwyl-Irvine et al., 2020). The session lasted approximately 90 minutes, and participants were encouraged to take breaks between tasks. Since the study was conducted remotely, at the end of the experiment, participants were asked about their compliance with task instructions and whether they had used any strategies to aid performance (e.g., writing down words, skipping tasks).

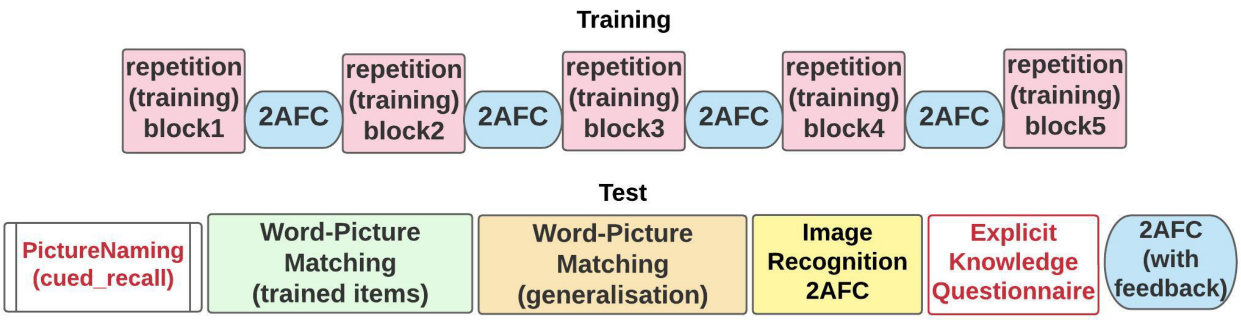

Figure 3.11 Experiment 3 procedure.

3.3.2.4 Tasks

Language Training: Word Repetition

This task was identical to the one used in Experiments 1 and 2. To account for the increased number of trained words in Experiment 3, we increased the number of exposures per word-picture pair from 8 (in Experiments 1 and 2) to 10 in the current experiment. These were distributed across five blocks, with each word repeated twice per block. The order of trials was randomised for each participant. There were two practice trials, and as in previous experiments, each repetition block was followed by a word recognition task (2AFC).

Language Tests: Word Learning

2AFC Recognition. As in our previous experiments, the 2AFC recognition task was used to assess participants' word learning performance. This tasks procedure was identical to the one used in Experiments 1 and 2 and recognition accuracy (0,1) was the outcome measure.

Picture Naming. This task followed the same procedure as in Experiments 1 and 2. Recall accuracy (0,1) was the outcome measure, and responses were scored offline by the experimenter using the same method described in Chapter 2.

Language Tests: Grammar Learning

Picture Naming (cued recall). As in Experiments 1 and 2, Picture Naming was also used to assess grammar learning by measuring participants' recall accuracy of the determiner and suffix mappings. As outlined above, recall accuracy (0,1) was the outcome measure and was scored offline using the same method described in Chapter 2.

Word-Picture Matching: a' generalisation. The same Word-Picture Matching task as in Experiment 2 was used to assess generalisation of the determiner and suffix mappings. A' was used at the outcome measure in this task.

Language Tests: Explicit Knowledge

Word-Picture Matching: Explicit Knowledge. The procedure for this task was the same as in Experiment 2. Participants were presented with the 48 trained items with 96 trials in total. As in Experiment 2, on half the trials, trained novel words were presented with pictures of objects with which they had been trained and on the other half trained novel words were presented with trained pictures but of the opposite semantic category.

In the generalisation version of this task, participants were presented with 8 new items, half of which featured word-picture pairings that conformed to the grammatical mappings presented during training (consistent trials), while the other half featured pairings that did not conform to the trained mappings (inconsistent trials).

Explicit Knowledge Questionnaire

This questionnaire followed the same procedure as Experiment 1 and 2 and the same scoring system was used.

Language Tests: Image Recognition

We used a 2AFC task to assess participants' recognition of the images in the prior knowledge dependent and prior knowledge independent conditions. This task was designed to ensure that the prior knowledge benefit observed in previous studies was not simply due to the images within a given semantic category being more recognisable (e.g., the pictures of male/female characters being easier to recognise than those of sea animals/insects).

Each trial began with a 500 ms fixation cross at the centre of the screen, followed by two images presented side by side. One image was the trained target, while the other was a foil (a visually altered version of the original image). The instructions outlined to participants that each trial contained one image they had seen during training and one foil, and they were asked to select the image they had previously seen. Participants selected a response by clicking on one of the images, and if no response was made within 4000 ms, the trial timed out. Before the task commenced, participants were presented with the instruction screen, followed by two practice trials.

Foils were created in Adobe Photoshop (Adobe Inc., 2023) by systematically altering the colours of specific features within each image. These modifications included changes to key features such as clothing and body markings (e.g., changing a character's clothing from green to blue or the colour patterns on a butterfly's wings).

The task consisted of 48 trials (one per trained item). The position of the target image was counterbalanced, appearing on the left in half of the trials and on the right in the other half. Trial order was randomised across participants, and no feedback was provided.

3.3.2.5 Data Analyses

All analyses reported below were conducted in R (R Core Team, 2020). The presented analysis follows the pre-registered analysis plan in OSF (pre-registration: https://osf.io/ucqkt/).

The analyses for all tasks were identical to those used in Experiment 2, except for the Image Recognition task, which is outlined below.

Image Recognition

To examine potential differences in image recognisability between the prior knowledge dependent and prior knowledge independent conditions, we conducted a mixed-effects binomial regression with accuracy (0,1) as the outcome variable. Mapping type was included as a fixed effect (prior knowledge dependent = 1, prior knowledge independent = -1). The random-effects structure included participants and items as random intercepts, with a random slope for mapping type within the by-participant effects. The package *buildmer* (Voeten, 2021) was used to establish the random-effect's structure, starting from the maximal model described above. The final random effects structure with the model that converged for each model is provided in the Results section below.

3.3.3 Results

3.3.3.1 Word Learning

2AFC

As in Experiment 2, word learning was assessed using the 2AFC task. Participants had similar levels of word learning as in Experiment 2 (Figure 3.12; Prior Knowledge Dependent: M = 0.92; Prior Knowledge Independent: M = 0.90). As in Experiment 2, the main effect of mapping was not significant (Table 3.10).

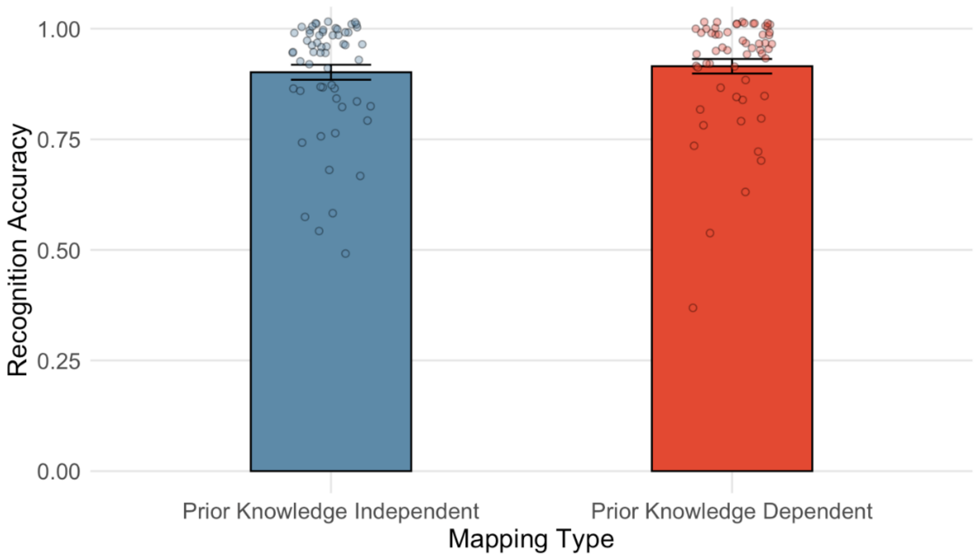

Figure 3.12. Accuracy (proportion correct) on the 2AFC task measuring word learning. Error bars represent standard error.

Table 3.10. Regression coefficients for the fixed factor using recognition accuracy as a binary outcome measure in the 2AFC word learning task.

	Estimate	Std. Error	z	p
Mapping	0.10	0.09	1.05	.294

Note. Random effects included in the final model that converged: by-participant and by-item intercepts.

Picture Naming: Stems

As illustrated in Figure 3.13, stem recall accuracy was low in both conditions but somewhat higher than in Experiment 2. In line with Experiment 2, there was no significant main effect of mapping on stem recall performance (see Table 3.11).

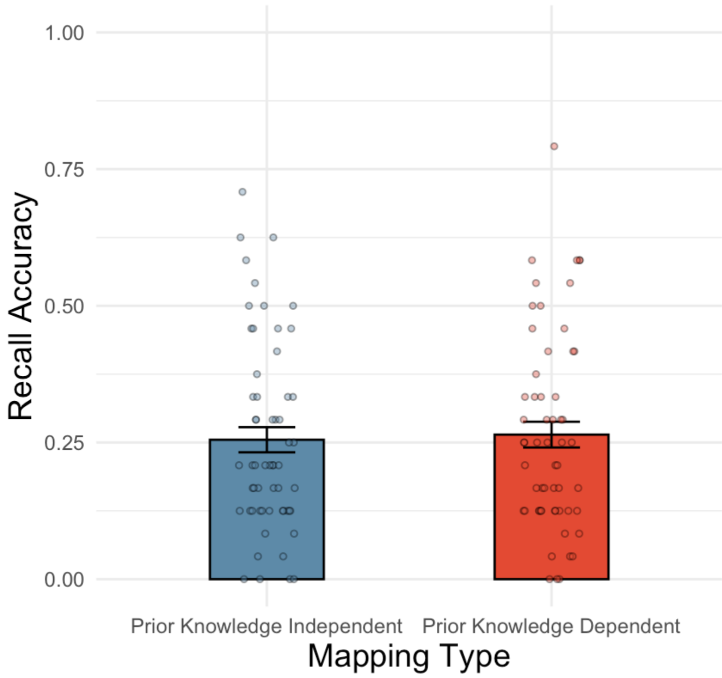

Figure 3.13. Stem recall accuracy (proportion correct) on the Picture Naming task. Error bars represent standard error.

Table 3.11. Regression coefficients for the fixed factors using stem recall accuracy as a binary outcome measure in the Picture Naming task.

	Estimate	Std. Error	z	p
Mapping	0.03	0.10	0.30	.766

Note. Random effects included in the final model that converged: by-participant and by-item intercepts.

Overall, the findings from both tasks replicate those of Experiment 2 in that the prior knowledge manipulation of the grammatical mappings did not influence word learning performance when using a greater number of exemplars (larger vocabulary).

3.3.3.2 Grammar Learning

Picture Naming: Determiners.

As illustrated in Figure 3.14, participants recalled the determiner mappings to a good level, with accuracy similar to that in Experiment 2 even with a larger number of exemplars. As in Experiment 2, there was a significant main effect of mapping type, with greater recall accuracy for the determiner mappings in the prior knowledge dependent condition (Table 3.12).

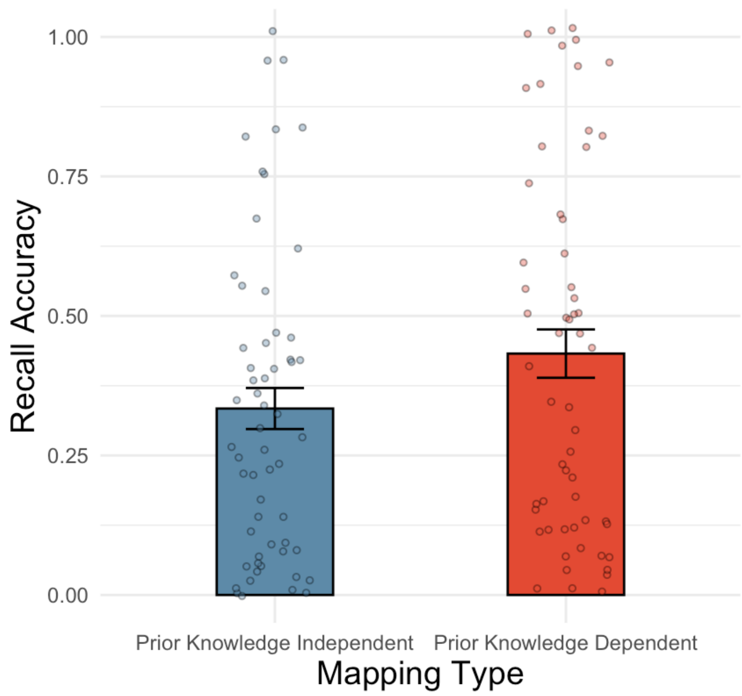

Figure 3.14. Determiner recall accuracy (proportion correct) on the Picture Naming task. Error bars represent standard error.

Table 3.12. Regression coefficients for the fixed factors using determiner recall accuracy as a binary outcome measure in the Picture Naming task.

	Estimate	Std. Error	z	p
Mapping	**0.313**	**0.094**	**3.32**	**.001**

Note. Random effects included in the final model that converged: by-participant and by-item intercepts. For participants, random slopes were included for the mapping contrast.

Picture Naming: Suffixes.

Suffix recall accuracy was also closely aligned with accuracy in Experiment 2 (Figure 3.15), and once again, participants demonstrated higher accuracy for the suffix mappings compared to the determiners (prior knowledge dependent mappings: $M_{determiner} = 0.43$, $M_{suffix} = 0.57$; prior knowledge independent mappings: $M_{determiner} = 0.33$, $M_{suffix} = 0.48$).

Furthermore, as in Experiments 1 and 2, a significant main effect of mapping type was observed, with higher accuracy for the suffix mappings in the prior knowledge dependent condition (see Figure 3.15, Table 3.13).

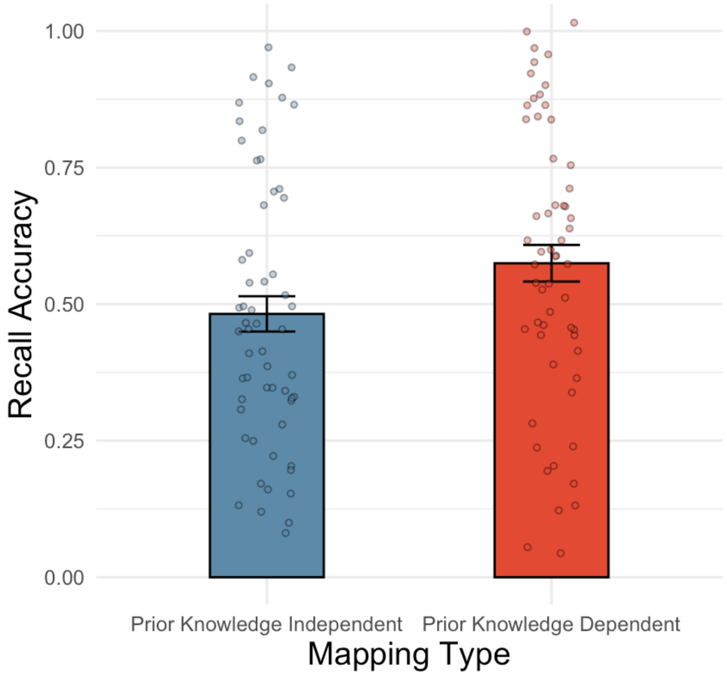

Figure 3.15. Suffix recall accuracy (proportion correct) on the Picture Naming task. Error bars represent standard error.

Table 3.13. Regression coefficients for the fixed factors using suffix recall accuracy as a binary outcome measure in the Picture Naming task.

	Estimate	Std. Error	z	p
Mapping	**0.254**	**0.084**	**3.01**	**.003**

Note. Random effects included in the final model that converged: by-participant and by-item intercepts. For participants, random slopes were included for the mapping contrast.

Overall, these findings are in line with those from Experiments 1 and 2, showing that participants recalled the prior knowledge dependent determiner and suffix mappings more accurately than the prior knowledge independent mappings immediately after learning even when using a larger number of exemplars (larger vocabulary).

Word-Picture Matching: a' generalisation

The primary aim of Experiment 3 was to investigate whether increasing the number of exemplars during training would enhance generalisation and, if so, whether the prior knowledge dependent mappings would exhibit greater generalisation than the prior knowledge independent mappings.

As shown in Figure 3.16, a' scores closely resembled those observed in Experiment 2. One-tailed t-tests revealed that participants successfully generalised the mappings in both the prior knowledge dependent condition, $t(57) = 2.68$, $p = .005$, and the prior knowledge independent condition, $t(57) = 3.79$, $p < .001$. As in Experiment 2, the main effect of mapping was not significant, suggesting that prior knowledge did not influence generalisation (Table 3.14).

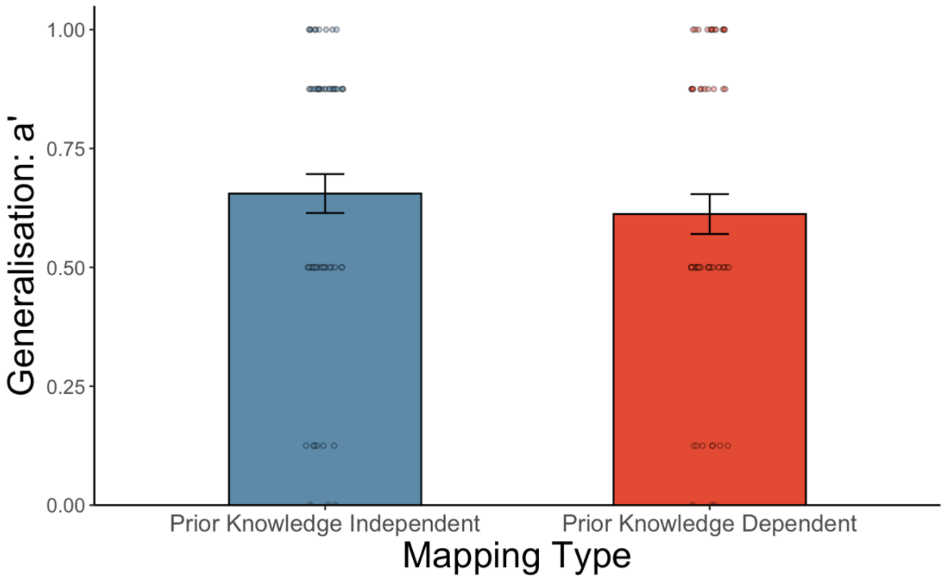

Figure 3.16. Generalisation performance for the determiner and suffix prior knowledge dependent and independent mappings. Error bars represent standard error.

Table 3.14 Regression coefficients for the fixed factors using a' scores as the outcome measure in the Word-Picture Matching task.

	Estimate	Std. Error	t	p
Mapping	-0.022	0.027	-0.796	.426

Note. Random effects included in the final model that converged: intercepts by participant.

In addition to this, we conducted an exploratory analysis to statistically compare a' scores assessing generalisation performance in the current study with that in Experiment 2. A mixed-effects analysis with experiment and mapping type as fixed effects and the interaction between them revealed no significant main effects and no significant interaction (Table 3.15).

Table 3.15. Regression coefficients for the fixed factors with a' as the outcome measure in the Word-Picture Matching Generalisation Task.

	Estimate	Std. Error	t	p
Study	0.021	0.020	1.074	.283
Mapping	-0.005	0.017	-0.271	.786
Study x Mapping	-0.017	0.017	-0.985	.325

Note. Random effects included in the final model that converged: by-participant intercept.

Overall, the results presented above indicate that increasing the number of stems in the current study did not improve generalisation performance. For both mapping types, performance was similar to Experiment 2, and there was no evidence that generalisation was significantly better for the prior knowledge dependent mappings compared to the prior knowledge independent mappings.

3.3.3.3 Explicit Knowledge

We also assessed participants' explicit knowledge of the underlying mappings using the same Word-Picture Matching tasks as in Experiment 2. Additionally, as in our previous experiments, explicit knowledge was assessed using a retrospective verbal report which was administered at the end of the session.

Word-Picture Matching: Explicit Knowledge

As in Experiment 2, confidence ratings from the six-point response scale were used in a 'zero-correlation' analysis separately for the trained and generalisation items

Word-Picture Matching: training

As in Experiment 2, a linear mixed regression analysis revealed a significant main effect of accuracy (see Table 3.16), with higher confidence ratings for correct responses compared to incorrect ones. Additionally, a significant main effect of mapping was observed (Table 3.16, Figure 3.17), with greater confidence ratings for mappings in the prior knowledge dependent condition than in the prior knowledge independent condition. However, the interaction between accuracy and mapping was not significant, indicating that the relationship between confidence and accuracy did not differ by mapping type.

These findings align with those of Experiment 2, suggesting that while participants demonstrated explicit knowledge of the grammatical mappings in the trained items, this knowledge was not greater for mappings in the prior knowledge dependent condition.

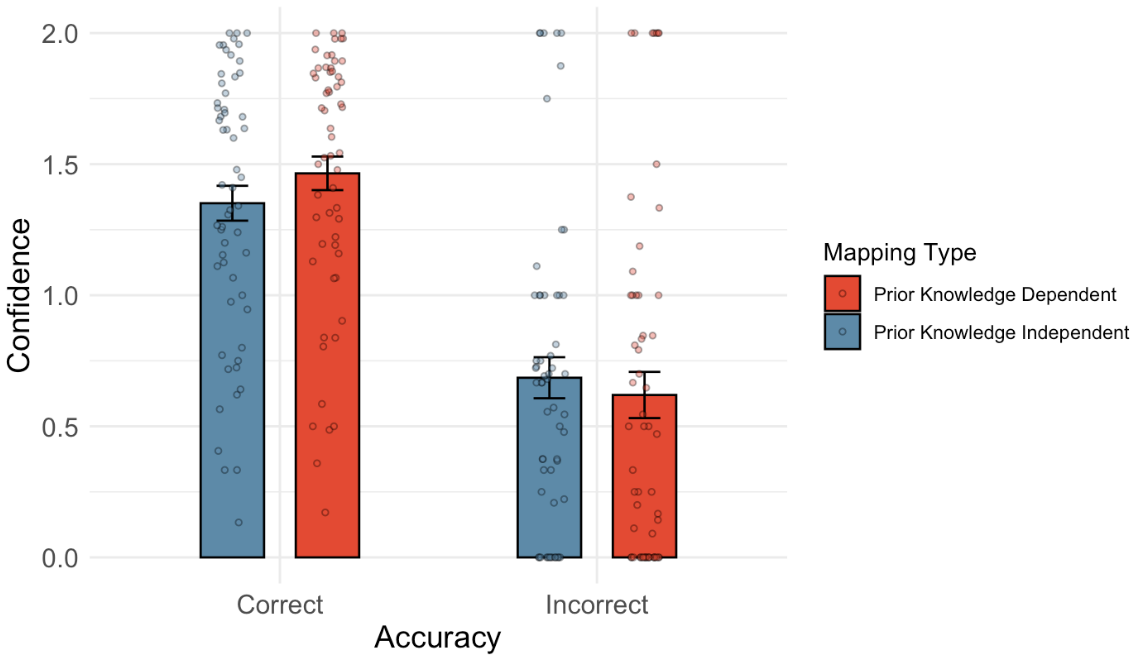

Figure 3.17. Explicit Knowledge: Mean confidence ratings for correct and incorrect responses in the Word-Picture Matching training task. Error bars represent standard error.

Table 3.16. Regression coefficients for the fixed factors using confidence ratings in the Word-Picture Matching training task as the outcome.

	Estimate	Std. Error	t	p
Accuracy	**0.218**	**0.012**	**17.84**	**<.001**
Mapping	**0.051**	**0.019**	**2.73**	**.006**
Accuracy × Mapping	0.008	0.012	0.71	.479

Note. Random effects included in the final model that converged: by-participant and by-item intercepts. For participants, random slopes were included for mapping. Significant results are highlighted in bold.

Word-Picture Matching: generalisation

For the generalisation items, a significant main effect of accuracy was found, with higher confidence ratings for correct responses compared to incorrect responses (Table 3.17; Figure 3.18). However, the confidence ratings for the correct responses were notably lower than with the trained items (Figure 3.17). Neither the main effect of mapping nor the interaction between accuracy and mapping were significant.

These results suggest that explicit awareness of the mappings influenced the performance in the Word-Picture matching task when they were presented in new, untrained contexts. This contrasts with Experiment 2, where there was no evidence for the role of explicit awareness in the Word-Picture Matching generalisation task. Thus, increasing the number of exemplars (stems) did seem to result in a greater role of explicit knowledge in this task.

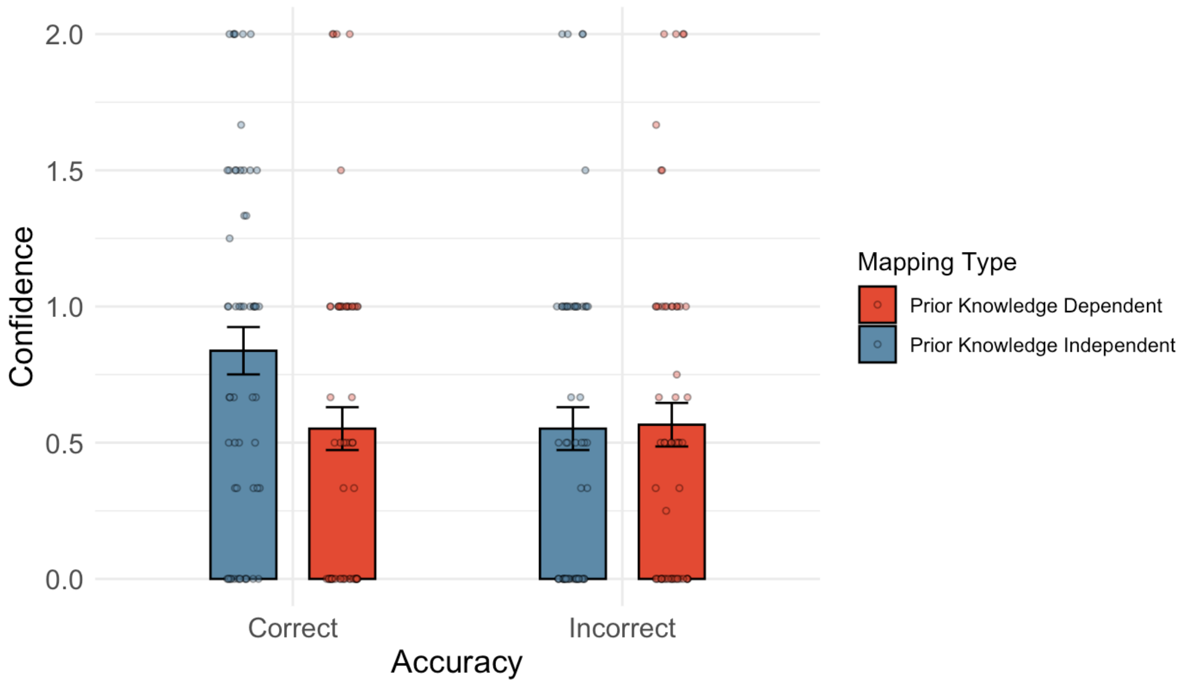

Figure 3.18. Explicit Knowledge: Mean confidence ratings for correct and incorrect responses in the Word-Picture Matching generalisation task. Error bars represent standard error.

Table 3.17. Regression coefficients for the fixed factors using confidence ratings in the Word-Picture Matching generalisation task as the outcome.

	Estimate	Std. Error	t	p
Accuracy	**0.073**	**0.031**	**2.33**	**.020**
Mapping	0.04	0.044	0.91	.361
Accuracy x Mapping	-0.012	0.031	-0.40	.693

Note. Random effects included in the final model that converged: by-participant and by-item intercepts. For participants, random slopes were included for the mapping contrast. Significant results are highlighted in bold.

Explicit Knowledge Questionnaire

Explicit knowledge of the mappings was also assessed using a retrospective verbal report. As in Experiment 2, a mixed-effect analyses revealed a significant main effect of mapping for the determiners, with greater levels of explicit knowledge for the prior knowledge dependent mappings (Table 3.18). For suffixes, the main effect of mapping was not significant (Table 3.19).

Similar to Experiment 2, levels of explicit knowledge were low overall but greater for the determiners than for suffixes (Figure 3.19), and somewhat greater for determiners than in Experiment 2 (Figure 3.10).

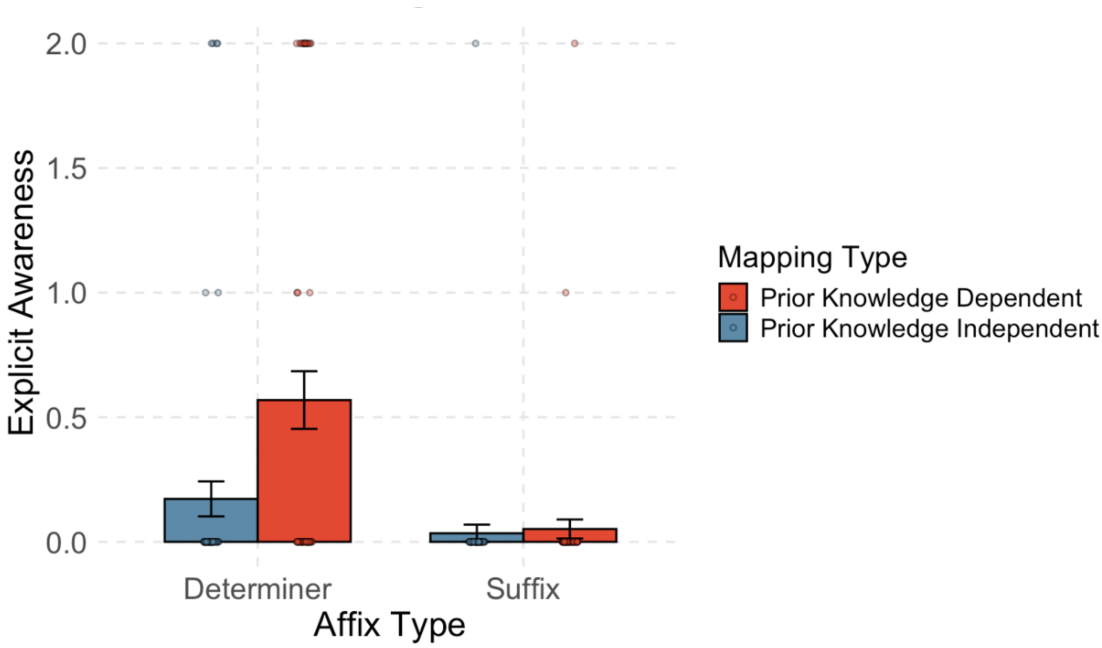

Figure 3.19. Levels of explicit awareness of the determiners and suffixes in the Explicit Knowledge Questionnaire. Error bars represent standard error.

Table 3.18. Regression coefficients for the fixed factor using determiner responses in the Explicit Knowledge Questionnaire as the outcome.

	Estimate	Std. Error	t	p
Mapping	**0.20**	**0.05**	**3.81**	**< .001**

Note. Random effects included in the final model that converged: by-participant intercepts.

Table 3.19. Regression coefficients for the fixed factor using suffix responses in the Explicit Knowledge Questionnaire as the outcome.

	Estimate	Std. Error	t	p
Mapping	0.01	0.01	1.00	.317

Note. Random effects included in the final model that converged: by-participant intercepts.

Image Recognition

To examine whether image recognisability differed between the prior knowledge dependent and prior knowledge independent conditions, we conducted a mixed-effects regression on the 2AFC Image Recognition task. As illustrated in Figure 3.20, participants' recognition accuracy was slightly higher for the images in the prior knowledge independent condition (M = 0.94) than for those in the prior knowledge dependent condition (M = 0.91) as shown by a significant main effect of mapping (Table 3.20). This difference is in the opposite direction than what would be expected if the prior knowledge benefit observed at recall was driven by the images.

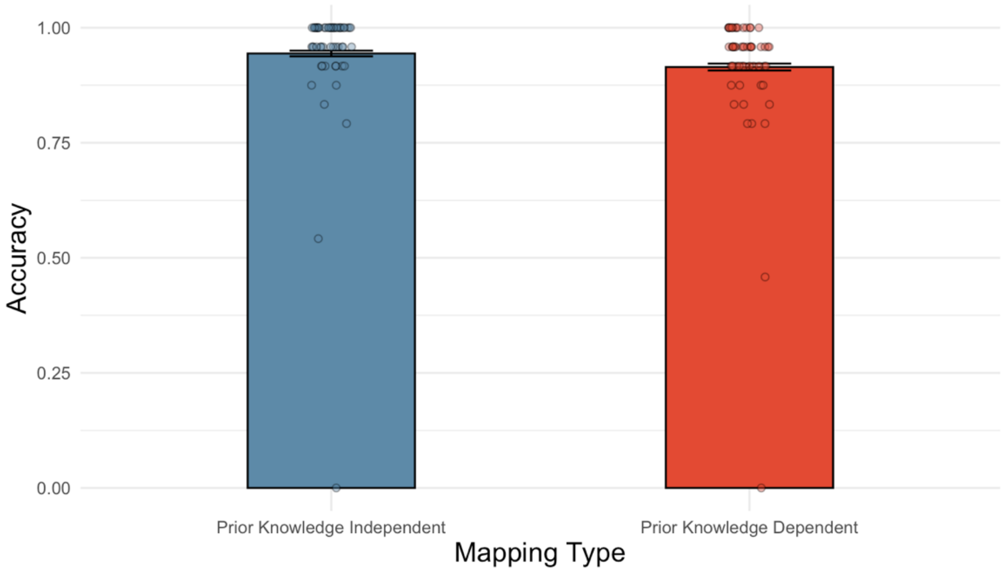

Figure 3.20. Accuracy for the Prior Knowledge Dependent and Independent conditions in the 2AFC Image Recognition task. Error bars represent standard error.

Table 3.20. Regression coefficients for the fixed factor using accuracy in the 2AFC Image Recognition task as the outcome.

	Estimate	Std. Error	z	p
Mapping	**-0.48**	**0.14**	**-3.34**	**<.001**

Note. Random effects included in the final model that converged: by-participant intercepts with random slopes included for the mapping contrast.

3.3.4 Discussion

The primary aim of Experiment 3 was to investigate whether increasing the number of stems during training would enhance generalisation and, if so, whether enhanced generalisation would allow us to detect any prior knowledge benefits. This manipulation was motivated by previous findings suggesting that greater variability in input can facilitate generalisation by enhancing learners' ability to detect shared regularities across items (e.g., Gómez et al., 2002; Tamminen et al., 2015). Hence the number of stems in each prior knowledge condition was increased from 16 to 24. Additionally, to ensure word learning remained comparable to Experiment 2, we increased the number of repetitions of individual words during training.

As in Experiment 2, word knowledge was assessed using the 2AFC recognition task and stem recall accuracy in the Picture Naming task, with both tasks showing similar levels of word learning across the two experiments. As in Experiments 1 and 2, there were no differences between the prior knowledge conditions in these tasks assessing word knowledge.

The Picture Naming task also assessed recall accuracy for the grammatical mappings immediately after training. Consistent with Experiments 1 and 2, participants demonstrated higher recall accuracy for prior knowledge dependent mappings relative to prior knowledge independent mappings, thus replicating the prior knowledge benefit for the grammatical mappings in the trained items. In contrast to our hypothesis and previous findings (Gómez et al., 2002; Tamminen et al., 2015; although see Brown et al., 2022 for contrasting findings), increasing exemplar variability did not improve generalisation performance. As in Experiment 2, participants' generalisation performance was relatively weak overall ($M_{\text{prior knowledge dependent}} = 0.61$; $M_{\text{prior knowledge independent}} = 0.65$), and no evidence of a generalisation advantage for the prior knowledge dependent mappings was observed. One possibility is that the increase in exemplar variability (from 16 to 24 stems in the current experiment) was too modest to yield an improvement in generalisation. Further increases in the number of stems were constrained by the fact that there were only a limited number of suitable items available within the semantic categories. This was particularly true for the semantic categories in the prior knowledge independent condition, where the pool of suitable items was more limited (e.g., there were only a finite number of identifiable sea animals and insects that could be used in the stimulus set). Nonetheless, these findings are consistent with Experiments 1 and 2, demonstrating that although prior knowledge supports the initial learning of grammatical mappings, this benefit does not extend to generalisation, at least under conditions where generalisation is relatively weak.

In Experiment 3, we also considered whether the images we used to train the participants on the experimental language contributed to the prior knowledge benefit observed for trained grammatical mappings in the Picture Naming task. For example, images of male and female characters in the prior knowledge dependent condition could have been more memorable or easier to recognise than images of sea creatures and insects in the independent condition. To examine this possibility, we assessed participants' recognition accuracy for the images presented during training. Recognition accuracy was high for the images in both conditions, but higher for the images in the prior knowledge independent condition, suggesting that differences in image recognisability are unlikely to account for the observed prior knowledge benefit at cued recall.

As in Experiment 2, participants' explicit knowledge of the underlying grammatical mappings was also examined using two measures: a retrospective verbal report and a Word-Picture Matching task with an integrated confidence rating measure. The retrospective verbal report was administered at the end of the single experimental session in Experiments 2 and 3, whereas in Experiment 1 it was administered at a session a week after training. In all cases, determiner and suffix mappings were assessed separately. As in Experiments 1 and 2, in

Experiment 3 self-reported explicit knowledge was low overall, though participants again demonstrated greater awareness for the prior knowledge dependent determiner mappings. In contrast, no differences between the two mapping types were observed in self-reported explicit knowledge for the suffix mappings, and knowledge of the suffixes was lower than for determiners. These findings are consistent with Experiments 1 and 2.

Consistent with Experiment 2, the Word-Picture Matching task provided evidence for the role of explicit knowledge when the mappings were presented in their trained stem contexts, with participants demonstrating higher confidence ratings for correct compared to incorrect responses. However, as observed in Experiment 2, there was no indication that explicit knowledge was greater for the prior knowledge dependent mappings, as confidence ratings did not differ between the two mapping types. Interestingly, and in contrast to Experiment 2, when the mappings were presented in novel, untrained contexts, there was evidence for the role of explicit knowledge in this task with the generalisation items, as indicated by higher confidence ratings for correct compared to incorrect responses. Confidence ratings for the correct responses were overall lower with the generalisation items relative to the trained items. One possibility is that the increased exemplar variability during training in Experiment 3 facilitated greater awareness of the underlying grammatical mappings that was also now found with the generalisation items. Nonetheless, explicit knowledge was still not greater for the prior knowledge dependent mappings in this task.

Overall, the findings from Experiment 3 closely mirrored those of Experiment 2. Once again, the Picture Naming task revealed higher recall accuracy immediately after training for the determiner and suffix mappings in the prior knowledge dependent condition. However, despite increasing the number of exemplars during training, generalisation performance remained similar to the levels observed in Experiment 2, and participants did not demonstrate greater generalisation for the prior knowledge dependent mappings compared to prior knowledge independent mappings.

3.4 General Discussion

There were three main aims of the experiments presented in Chapter 3. The first aim was concerned with replicating our finding from Experiment 1 of better initial learning of novel grammatical mappings that are more dependent on participants' existing knowledge compared to mappings that were less dependent. This prior knowledge benefit, assessed by the Picture Naming cued recall task, was replicated in both Experiments 2 and 3 for both determiners and suffixes, further reinforcing the conclusion that prior knowledge can support the initial learning of novel grammatical mappings. These results are consistent with the theoretical predictions outlined in the CLS (McClelland, 2013) and SLIMM (VanKesteren et al., 2012) models of memory consolidation which propose that the time course of the consolidation of new knowledge is influenced by its relationship to existing knowledge. Specifically, by these accounts, existing schematic knowledge should advance neocortical integration of related information at encoding, resulting in faster consolidation into long-term memory networks. However, for new information that is unrelated to existing knowledge, consolidation is predicted to be slower, requiring integration to happen offline (e.g., during post-learning sleep). The experiments outlined in this chapter assessed initial learning, and the greater immediate recall observed for prior knowledge dependent compared to prior knowledge independent mappings suggests they benefited from their connections to existing grammatical knowledge during encoding. These findings are also consistent with experimental evidence from the word learning literature which has shown that novel words that are phonologically similar to existing word forms are better recalled immediately after learning compared to novel words without such connections to existing knowledge (James et al., 2019). Overall, the findings from Experiment 2 and 3 further support those of Experiment 1 in showing that prior knowledge can support the initial learning of novel grammatical mappings.

As the second aim the experiments reported in Chapter 3 explored the possibility of an explicit knowledge contribution to the observed prior knowledge benefits. This was motivated primarily by previous findings showing that prior knowledge benefits in novel word learning occur specifically when encoding conditions enable learners to explicitly link new information to existing knowledge (James et al., 2019; 2021), and by our findings from Experiment 1, where a prior knowledge benefit in grammar learning was only observed in a task permitting explicit retrieval (the non-speeded Picture Naming task). Specifically in the speeded tasks in Experiment 1, no processing advantage was observed for the prior knowledge dependent mappings when they were presented in their trained item contexts or when they were presented in novel, untrained contexts. Thus, in light of these findings, we directly assessed participants' explicit knowledge of the grammatical mappings to determine whether explicit knowledge was greater for mappings related to participants' existing knowledge. To test this, two measures were employed: a retrospective verbal report that was also administered at the end of the final session in Experiment 1, and a Word-Picture Matching task with an integrated confidence rating measure. First, across both experiments and consistent with the findings from Experiment 1, participants' self-reported explicit knowledge, as measured by the retrospective verbal report, was greater for the prior knowledge dependent determiner mappings relative to the prior knowledge independent determiner mappings. However, explicit knowledge was low overall, and as in Experiment 1, no differences between the mapping types were observed for the suffixes in Experiments 2 and 3.

In Experiments 2 and 3, evidence of a role for explicit knowledge was also observed in the Word-Picture Matching task when the mappings were presented in their trained item contexts. This task assessed explicit knowledge in an integrated fashion while participants made recognition judgements about the match between the novel words and their meanings. However, unlike the retrospective verbal report measure, participants did not demonstrate greater explicit awareness for the prior knowledge dependent mappings in this task. Similarly, participants did not demonstrate greater explicit awareness for these mappings when they were presented in novel, untrained contexts, in either Experiment 2 or 3. Taken together, these findings challenge the earlier interpretation from Experiment 1, where it was speculated that the presence of a prior knowledge benefit in the non-speeded Picture Naming task, but not in the speeded tasks, was due to the non-speeded task allowing participants the opportunity to explicitly retrieve connections between their existing grammatical knowledge and the newly learned, related mappings.

The final aim of the current experiments was to examine whether the observed prior knowledge benefit extends beyond item-specific recall to support the generalisation of grammatical knowledge to novel, untrained items. In Experiment 1 no prior knowledge benefit was observed in generalisation when tested under speeded conditions. However, as outlined above, it was reasoned that if participants were using explicit knowledge to support their learning of the prior knowledge dependent mappings, then a prior knowledge benefit at generalisation might be more likely to emerge in a non-speeded task, where participants would have the opportunity to consciously retrieve and apply that knowledge. To test this, the current experiments employed the same non-speeded Word-Picture Matching task previously described for measuring explicit knowledge of the mappings. In Experiment 2, we found evidence that participants were able to generalise the mappings in this task; however, generalisation was weak overall and, crucially, participants were no better at generalising the prior knowledge dependent mappings than the independent ones.

In light of these findings, Experiment 3 aimed to determine whether increasing exemplar variability would enhance generalisation, and if so, whether we would observe greater generalisation for the prior knowledge dependent mappings. To this end, we increased the number of stems presented during training. However, despite increasing the number of exemplars during training, generalisation performance remained similar to that observed in Experiment 2: participants exhibited weak generalisation overall, and there was no evidence of a generalisation advantage for the prior knowledge dependent mappings.

The absence of a generalisation benefit following increased exemplar variability is surprising given prior evidence demonstrating that greater variability in exemplars can enhance the learning of grammatical regularities (e.g., Gómez et al., 2002; Gómez & Maye, 2009; Tamminen et al., 2015). However, differences in language complexity may account for this discrepancy. Specifically, the current study employed a complex artificial language, incorporating both distributional and semantic regularities. In contrast, the studies by Gómez (2002) and Gómez and Maye (2005) used simpler artificial languages that included only distributional regularities without semantic cues. Thus, increased exemplar variability alone might be insufficient to enhance grammatical generalisation in more complex learning contexts. Supporting this, Brown et al. (2022), also found no improvement in generalisation performance when exemplar variability was increased at training. Similar to the current study, Brown et al. (2022) used a language containing both distributional and semantic regularities, reinforcing the possibility that variability may be less beneficial for generalisation in more complex paradigms, or that an even greater number of exemplars is needed.

It is also possible that the type of generalisation assessed in Experiments 2 and 3 is less likely to benefit from greater exemplar variability. As outlined in Chapter 2, Tamminen et al. (2012; 2015) distinguish between two types of generalisation: one that arises from distinct episodic representations in the hippocampus, and another supported by overlapping representations in the neocortex. Specifically, generalisation arising from distinct hippocampal representations requires learners to retrieve and average across multiple context-dependent memory traces in order to form an abstract representation sufficient for generalisation. Thus, generalisation in this form requires additional processing time and so is more likely to support performance in non-speeded tasks. Conversely, generalisation that arises on the basis of overlapping cortical representations does not require learners to retrieve and combine multiple episodic memory traces. This is because the neocortex represents knowledge of related experiences in an overlapping and distributed manner, emphasising their shared, context-independent properties. Hence, generalisation in this form can be achieved more rapidly, supporting performance in speeded tasks. When generalising novel morphological forms and their meanings in speeded tasks, Tamminen et al. (2012; 2015) showed that offline consolidation is required. This is because hippocampal-neocortical integration during offline processes such as sleep provides one mechanism through which overlapping neocortical representations can develop and thus support learning in speeded tasks. By contrast, since generalisation in non-speeded tasks is supported by distinct episodic hippocampal representations, it does not depend on the formation of overlapping cortical representations and therefore does not require offline consolidation processes to emerge.

In addition to offline consolidation, Tamminen et al. (2015) also showed that generalisation in speeded contexts is influenced by contextual diversity (i.e., exemplar variability), whereas generalisation in non-speeded contexts is not. Specifically, participants only showed evidence of generalisation for affixes trained across many stem contexts in a speeded reading-aloud task, with no generalisation observed for affixes trained across fewer stem contexts. In contrast, contextual diversity did not influence generalisation performance on an explicit, non-speeded judgment task. As with offline consolidation, Tamminen et al. (2015) reasoned that this task dissociation in the influence of contextual diversity speaks to the nature of the representations that support generalisation in speeded and non-speeded contexts. For instance, overlapping, context-independent cortical representations, which support generalisation in speeded contexts, should particularly benefit from increased contextual diversity. This is because when a novel affix occurs over a large number of stem contexts, the neocortical system is better able to form a representation that is independent of the stem contexts in which it was trained, making it more readily generalisable. By contrast, since performance in non-speeded tasks is likely supported by non-overlapping, context-dependent hippocampal representations, factors enhancing context-independent features of related representations/ items (e.g., greater contextual diversity) are unlikely to influence performance here. Taking these findings into consideration, one possible reason that increasing exemplar variability in Experiment 3 did not lead to improvements in generalisation is because we used a non-speeded task, which, due to its reliance on non-overlapping hippocampal representations, may be less sensitive to increases in input variability.

Nonetheless, the findings from Experiments 2 and 3 are consistent with those of Experiment 1 in that, although prior knowledge supported the initial learning of related grammatical mappings, this benefit did not extend to generalisation. This consistency is particularly noteworthy given that generalisation was assessed using both speeded and non-speeded tasks. For example, in Experiment 1, it was speculated that the prior knowledge benefit in the Picture Naming cued recall task and its absence in a speeded generalisation task reflected differences in the retrieval conditions of the two tasks, with the non-speeded task affording participants the opportunity to recall connections between their existing knowledge and new, related knowledge. However, the absence of a prior knowledge advantage in the non-speeded generalisation tasks in Experiments 2 and 3 suggests that differences in retrieval conditions across tasks are unlikely to account for the absence of a prior knowledge benefit in generalisation in the current paradigm.

Instead, in light of the present findings, one possible explanation is that the representations of the grammatical mappings formed during training were not sufficiently abstract to support generalisation. For instance, according to abstractionist theories of category learning, learners generalise on the basis of a prototype representation, which serves as an integrated summary or average of previously encountered exemplars (Posner & Keele, 1968). This model of generalisation is consistent with the type of generalisation that likely supported performance in speeded contexts in Tamminen et al. (2012; 2015) in that its emergence depended on the discovery of shared structure across a number of different exemplars of a given category. As outlined above, in Tamminen et al. (2015), contextual diversity was one factor that aided the abstraction and generalisation of morphemic representations; however, the authors also found that semantic consistency played an equally important role. That is, when participants were trained on two types of novel affixes, those that referred to a single semantic category across all instances (semantically consistent affixes), and those that referred to one semantic category (e.g., tools) in half of the novel words and a different category (e.g., people) in the other half (semantically inconsistent affixes), only the affixes in the semantically consistent condition supported generalisation under speeded conditions (i.e., where generalisation likely relies on overlapping and integrated cortical representations). These findings are broadly in line with evidence from visual categorisation tasks which have shown that when individual category members have high perceptual coherence at training, it is easier for learners to form a prototype representation of the category that can be used to guide generalisation judgements to novel items (Bowman & Zeithamova, 2020).

Taking the above into consideration, one factor that may have contributed to the absence of a prior knowledge benefit in the speeded generalisation task in Experiment 1 is the fact that the training language comprised grammatical mappings that were semantically inconsistent or less coherent. For example, each determiner and suffix predicted two distinct semantic categories, with *tib* and *eem* denoting both female characters and insects, and *ked* and *ool* referring to both male characters and sea animals. This inconsistency may have made it difficult for participants to form a sufficiently abstract or prototype representation of the mappings in either condition, which could have supported performance under the time constraints of the speeded task. Furthermore, semantic inconsistency may also account for the weak generalisation and lack of a prior knowledge benefit in the non-speeded generalisation tasks in Experiments 2 and 3. For instance, Tamminen et al. (2015) showed that although semantic inconsistency specifically disrupted generalisation in speeded contexts, it also reduced generalisation performance in a non-speeded context. Specifically, in a non-speeded explicit congruency judgement task, participants were less accurate when making congruency judgements about affixes that were semantically inconsistent at training compared to affixes that were semantically consistent. Thus, one possibility is that the semantic inconsistency of the determiner and suffix mappings during training reduced participants' accuracy when making word-picture matching judgments for both the prior knowledge dependent and independent mappings in the non-speeded generalisation task.

More broadly, the findings from Experiment 2 and 3 should also be interpreted in light of several limitations of the current paradigm. First, one possible limitation of the current paradigm concerns the potential weakness of the prior knowledge manipulation of the systematic mappings. For instance, although English encodes natural gender in the pronouns, he/him and she/her, it does not mark grammatical gender systematically across determiners or noun endings. In addition, it should also be considered that the semantic categories associated with the prior knowledge dependent mappings (male and female characters) were easier to categorise than those associated with the prior knowledge independent condition (insects and sea animals), given that the distinction based on biological sex may be more salient to participants than semantic distinctions within animal species. If so, the recall advantage observed for the prior knowledge dependent mappings could reflect the greater salience of their associated semantic categories, rather than consistency with existing grammatical knowledge. However, some of the findings observed in the current experiments suggest that greater salience is unlikely to provide a full explanation for the observed prior knowledge benefit. For instance, in both Experiments 2 and 3, participants' word recall and recognition were no greater for the words in the prior knowledge dependent condition than for those in the prior knowledge independent condition. Furthermore, in Experiment 3, participants' image recognition accuracy was greater for the images in the prior knowledge independent category than for those in the prior knowledge dependent category. This suggests that category salience alone is unlikely to provide a full explanation for the observed prior knowledge benefit in the recall of the systematic mappings.

One final limitation to be considered concerns the fact that Experiment 3 was conducted online. Although levels of word learning in Experiment 3 were similar to those of Experiment 2, the potential limitations of running experiments online ought to be acknowledged. For example, testing online meant that there was less control over the environment in which participants completed the study. As a result, there is less control over distractions and background noise, both of which may have affected engagement with the tasks and consequently, reduced data quality. Moreover, there is a risk with remote testing that participants may have written the novel words down during the training session. However, several steps were taken to mitigate these risks. For instance, participants were instructed to complete the study in a quiet environment and to follow all task instructions carefully. They were also informed that their verbal responses to the tasks would be checked for intelligibility and compliance with task instructions (e.g., repeating back the words in the training task). In addition, Experiment 3 also included a word recognition task with feedback at the end of the experimental session. This was intended to encourage engagement with the training, as participants were offered a bonus payment if they achieved above 95% accuracy. Nevertheless, despite the checks outlined above, some aspects of remote testing were more difficult to mitigate than others. For instance, while the questionnaire administered at the end of the session also included checks relating to cheating (i.e., asking participants whether they had written any of the novel words down during training or whether they had avoided completing any of the tasks), these checks relied on participants' honesty.

In summary, the two experiments outlined in this chapter replicate the finding from Experiment 1 in showing that prior linguistic knowledge facilitates the initial learning of novel grammatical mappings. However, contrary to what was hypothesised, explicit knowledge does not appear to play a significant role in the learning of the prior knowledge related mappings. Moreover, prior knowledge does not appear to support greater generalisation of related mappings even when retrieval conditions allow learners to explicitly draw on connections between their new and existing linguistic knowledge (i.e., in non-speeded tasks). While these findings are broadly consistent with the CLS (e.g., McClelland, 2013) and SLIMM (Van Kesteren, 2012) frameworks, they require further understanding of the processes that underpin the abstraction of grammatical regularities and how they may be supported by the neocortical and hippocampal systems.

Chapter 4. The Contribution of Implicit and Explicit Processes to Grammar Learning

4.1 Introduction

The current chapter outlines an exploratory study that set out to interrogate the contribution of explicit and implicit processes to grammar learning in conditions that are a step closer to naturalistic language learning. In particular, we used a larger training language with regularities that matched natural languages more closely, and we trained and tested participants over four consecutive days.

In Experiments 2 and 3 (Chapter 3 of this thesis), the focus was on examining the contribution of explicit knowledge to the learning and generalisation of novel grammatical mappings that were either related or unrelated to participants' existing knowledge. To address this, Experiments 2 and 3 employed direct measures (i.e., measures that are more reliant on deliberate, conscious retrieval) to assess the learning and generalisation of the mappings. However, direct measures alone are unlikely to provide a comprehensive overview of the processes underpinning the learning and consolidation of more systematic types of knowledge such as regularities in grammar which are typically thought to rely more on implicit, automatic processes (e.g., Aslin & Newport, 2012). In the context of CLS (e.g., Gaskell, 2024), systematic properties of novel words are more suited to cortical learning and a key property of this system is that retrieval (of cortical knowledge) is strongly automatic.

Eye-tracking is one method that has been used extensively to assess automatic implicit mechanisms in online language processing (e.g., Tanenhaus & Brown-Schmidt, 2009). It has been particularly informative as a measure of automatic use of predictive linguistic cues. For example, Lew-Williams and Fernald (2010), employed an eye-tracking procedure known as the looking while listening paradigm, to investigate whether adults who learned Spanish as a first language (L1) and adults who learned Spanish as a second language (L2) would take advantage of gender-marked determiners to correctly predict an upcoming referent before it is named. In their study, participants were presented with a two-picture visual display in which the two objects either belonged to the nouns of the same grammatical gender (same gender trials) or different grammatical gender (different gender trials). For instance, on same gender trials, participants may have seen the picture of a ball (*la pelota* - 'the-FEM ball-FEM') displayed alongside the picture of a cookie (*la galleta* 'the-FEM cookie-FEM') whereas on different gender trials, the ball may have been displayed alongside the picture of a shoe (*el zapato* - 'the-MASC shoe-MASC'). On both trials, participants' eye gaze was recorded as they listened to a sentence asking them to identify one of the objects (e.g., Encuentra la pelota. ¿La ves? – 'Find the ball. Do you see it?). The key outcome measure was the speed at which they moved their eyes to the target object. Across three experiments which involved seeing familiar and novel objects, L1 adults were faster to orient their gaze to the target objects on different gender trials than on same gender trials. This was taken as evidence of the automatic use of the predictive gender cues: on different gender trials, the determiner was informative about the gender of the noun and that allowed the participants to use them as a predictive cue for the upcoming noun. This facilitated the planning of the eye movements and faster looks to the target object. However, non-native Spanish speakers who were only moderately proficient were unable to use the gender-marked determiner predictively and instead had to wait until hearing the noun before initiating their gaze shift toward the target. In follow up experiments frequency of exposure was manipulated by training and testing both L1 and L2 speakers using a small set of novel nouns and novel objects, rather than real Spanish noun and objects. Both L1 and L2 speakers were able to use the determiners predictively when they were presented with the same articles as at training (*el/la*), but only L1 speakers were able to generalise when the novel nouns were presented with different determiners at test (e.g., *un/una* at training, *el/la* at test).

4.1.1 The Current Experiment

The experiment outlined in this chapter presents an exploratory study designed to examine the contribution of implicit and explicit processes to the learning of novel grammatical mappings. In the current experiment, the same artificial language as in Mirković et al. (2011) was used that incorporated both prior knowledge dependent and prior knowledge independent mappings as in Experiments 1-3. However, to more closely simulate the complexity of natural languages, some key changes were implemented. First, participants in the current study were trained on a larger number of words (64 compared to 32 words in Experiments 1 and 2, and 48 words in Experiment 3). As in previous experiments, half of the words were assigned to the prior knowledge dependent mapping condition and half assigned to the prior knowledge independent mapping condition. For the prior knowledge dependent mappings, as before, all *tib…eem* words were paired with pictures of female characters/occupations and all *ked…ool* words were paired with pictures of male characters/occupations. However, because the semantic categories previously used for the prior knowledge independent mappings (insects and sea animals) had a limited number of suitable exemplars, these categories were changed to accommodate the increased number of training items. For instance, while in previous experiments *tib…eem* words were paired exclusively with pictures of insects and *ked…ool* words exclusively with sea animals, in the current study, *tib…eem* words were paired with domesticated animals and insects, and *ked…ool* words were paired with wild animals and reptiles/amphibians. In addition to increasing the number of training words, the current study also increased the number of suffixes associated with each determiner to more closely replicate the complexity of regularities in natural languages (Mirković et al., 2005). Specifically, while in previous experiments each determiner combined with only one suffix, in the current experiment each determiner combined with two suffixes, such that all tib words ended in either *-eem* or *-esh*, and all ked words ended in either *-aff* or *-ool*.

Participants were trained and tested on the artificial language across four consecutive days. To track changes in the learning and consolidation of the novel words and grammatical mappings and to assess the contribution of explicit and implicit processes, we employed the same tasks as those used in Experiments 2 and 3 (Chapter 3 of this thesis) and a new implicit measure designed to test automatic processing of the grammatical mappings. First, to examine the use of the determiner mappings as predictive cues, the same eye-tracking paradigm as that used in the Lew-Williams and Fernald (2010) study outlined above was administered on each of the four days. Following their procedure, we recorded participants' eye movements while they viewed two pictures from the stimulus set and heard a novel word referring to one of the pictures. Following Lew-Williams and Fernald (2010) this task consisted of two sets of trials: same gender and different gender trials. On the same gender trials, the novel words for the two pictures in the visual display belonged to the same grammatical/gendered category (e.g., both *tib* or both *ked* words). On the different gender trials, the novel words for the two pictures belonged to different gendered categories (e.g., one picture was a *tib* word and one picture a *ked* word). Of key interest was whether participants would orient their gaze more quickly toward the target picture (the correct referent for the novel word) on different gender trials, where the gender-marked determiner provided a predictive cue, compared to same gender trials, where the determiner was not informative. Faster shifts on different gender trials would indicate that participants were able to exploit the gender-marked determiner to predict the upcoming novel words in real-time, thus providing evidence of implicit learning of the determiner mappings. Also of interest was when this behaviour would emerge over the course of training and how performance would compare across the two mapping types. For example, would participants orient their gaze more quickly toward the target referent on different gender trials when the novel word belonged to the prior knowledge dependent condition? On the final day of training, an additional block of the eye-tracking task was administered to assess generalisation. This allowed us to determine whether participants could use the determiner mappings acquired during training to make predictions about novel, previously unseen items.

Alongside implicit learning, explicit knowledge of the mappings was tested on each of the four training days. For this, we employed the same Word-Picture Matching task used in Experiments 2 and 3. In the current experiment, this task also provided participants with training on the language, replacing the Word Repetition task used to train participants in Experiments 1, 2 and 3. The main aim of this task was to assess whether explicit awareness of the underlying mappings emerged over the course of training, how this awareness developed across training days, and whether explicit knowledge differed for the two mapping types. On the final day of training, explicit knowledge was further assessed using the same retrospective verbal report as in all previous experiments. Additionally, an extra block of the Word-Picture Matching task was administered with new, previously unseen items to assess participants' awareness of the trained grammatical mappings when they were presented in novel contexts. As in Experiments 2 and 3, the Word–Picture Matching task with the new items was also used to assess generalisation. This provided an additional, more direct measure of generalisation (beyond the implicit generalisation test provided by the eye-tracking task described above), since the reflective nature of this task would allow learners the opportunity to consciously retrieve and apply their knowledge to support performance.

Finally, the same Picture Naming task used in previous experiments was administered at the end of each training day to assess participants' recall of the novel words and grammatical mappings. This task provided another direct measure of learning, with key interests being how recall would change across training days and whether the previously observed prior knowledge benefit for recalling the determiner and suffix mappings would replicate in a more complex language. Additionally, Picture Naming was used to assess offline changes in recall accuracy of the novel words and mappings across consolidation intervals (i.e., the intervals between consecutive training sessions), and to examine whether these changes differed between earlier and later intervals as the cumulative exposure to the language increased over the training sessions.

4.2 Method

4.2.1. Participants

38 adult participants aged between 18 and 40 were recruited from York St John University and participated in the study for monetary payment. Of the 38 recruited, the data from two participants were excluded from the analyses entirely due to not returning for the final session. All participants were monolingual, native speakers of English with no self-reported history of language, hearing or visual difficulties. Ethical approval for the study was obtained from the York St John University Cross-School Ethics Committee.

4.2.2. Materials

4.2.2.1 Artificial language

Training set. The training set consisted of 64, pronounceable, English pseudowords designed using the ARC database (Rastle et al., 2002). Each item featured a unique stem combined with one of four suffixes: *-eem, -esh, -aff*, or *-ool.* As in previous experiments, all words were paired with one of two determiners, *ked* or *tib,* with half of the words assigned to each. However, unlike the artificial language used in Experiments 1, 2 and 3, where each determiner corresponded to a single suffix, the current design assigned two suffixes to each determiner. Specifically, all words ending in *-eem* or *-esh* appeared with the *tib* determiner, while those ending in *-aff* or *-ool* were assigned to the *ked* determiner. As in the previous experiments, all novel words were paired with pictures (e.g., ballerina, policeman) selected from the commercially available ClipArt image database and were digitally recorded in a sound-proof booth by a native English speaker. There was no phonological overlap in the English word for the depicted character or animal and the novel word.

As in the previous experiments, the artificial language incorporated two types of grammatical mappings: prior knowledge dependent and prior knowledge independent. For the prior knowledge dependent mappings, there were 32 exemplars (16 *tib* and 16 *ked* words). The words were presented with pictures depicting stereotypical female or male characters, with *tib...eem/esh* words paired with pictures depicting female characters and *ked...ool/aff* words paired with pictures depicting male characters. For the prior knowledge independent mappings, there were 32 exemplars also (16 *tib* and 16 *ked* words). For this category, *tib...eem/esh* words were paired with pictures depicting domesticated animals and insects and *ked...ool/aff* words paired with pictures depicting wild animals and reptiles (see Table 4.1).

Generalisation set. An additional 16 items were administered on the final day of testing to assess participants' ability to generalise the mappings acquired in training to new, untrained items. As in previous experiments, the generalisation items differed in whether they were consistent or inconsistent with the trained regularities. Half of these (8 in total with 4 in each mapping condition) were consistent with the training set in that the mapping between the determiner, suffix and the semantic category represented by the picture (e.g., female/male characters, domesticated animals/insects and wild animals/reptiles) remained the same. For the other half of items (8 in total with 4 in each mapping condition), the mappings were inconsistent with the mappings presented during training.

Table 4.1 Example items from prior knowledge dependent and prior knowledge independent mappings and their assigned semantic category.

Mapping	Semantic Category	Determiner	Stem + Suffix	Picture
Prior Knowledge Dependent:	female character	*tib*	scoiff*esh*	*ballerina*
	male character	**ked**	lup**aff**	**priest**
Prior Knowledge Independent:	domesticated animals/insects	*tib*	doud*eem*	pig
	wild animals/reptiles/amphibians	**ked**	jor**ool**	**frog**

Note. Cues associated with *tib* words are marked in *italics*, *ked* in **bold**.

4.3. Procedure

4.3.1. Testing Schedule

Participants were trained and tested over four sessions spread across four consecutive days (Figure 4.1). All testing took place in a quiet room within the university's psychology labs. Upon arrival on Day 1, participants were presented with an information sheet and asked to provide written informed consent. As in Experiments 1, 2, and 3 no reference was made to the existence of the underlying grammatical mappings during the initial briefing, nor at any point during the study.

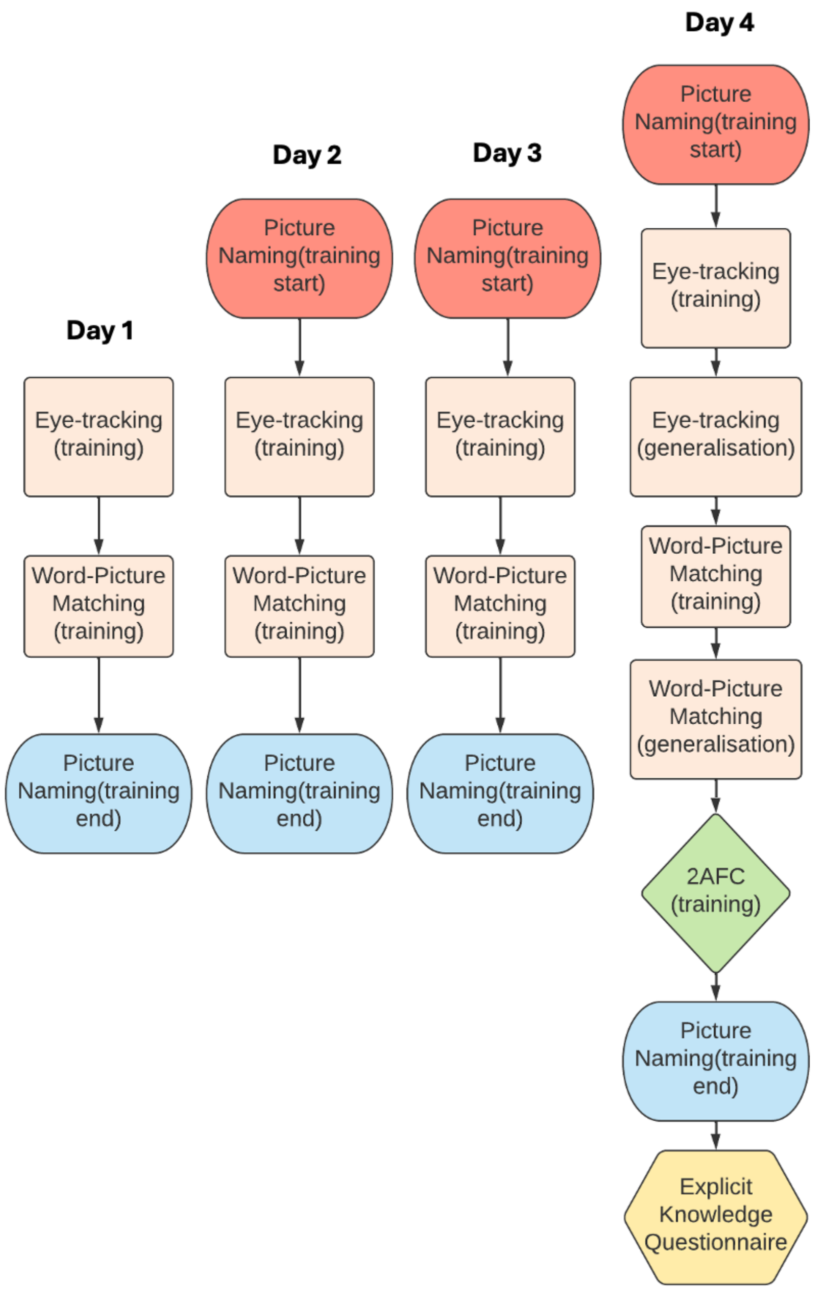

Figure 4.1 Experimental procedure: order of tasks per session. The stimulus set used in each task is provide in the parenthesis.

On Day 1, the session started with the eye-tracking task, followed by the Word-Picture Matching task, both of which provided training on the novel words. In addition to providing language training, the eye-tracking task was used to assess participants' ability to use the determiner cues predictively, while the Word–Picture Matching task provided a measure of explicit knowledge of the underlying mappings. At the end of Day 1, participants also completed the Picture Naming task to assess recall of the novel words and grammatical mappings following training (Figure 4.1).

Days 2 and 3 followed the same procedure as Day 1, with the addition of a Picture Naming task at the start of the session (Figure 4.1). This allowed us to compare recall accuracy of the words and mappings to that at the end of the previous day and therefore assess changes in accuracy between training sessions (i.e., over offline consolidation).

On the final day (Day 4), the same procedure was followed, including Picture Naming at both the start and end of the session. In addition to the tasks administered on previous days, participants also completed several new tasks (Figure 4.1). These included two extra blocks of the Eye-tracking and Word–Picture Matching tasks, to assess whether participants could generalise the grammatical mappings presented during training to novel, untrained items. Two additional tasks were also administered: a two-alternative forced-choice (2AFC) recognition task to assess learning of the novel words, and the Explicit Knowledge Questionnaire (see Figure 4.1).

Tasks per session were administered in the order shown in Figure 4.1. Picture Naming and Word-Picture Matching tasks were implemented in DMDX (Forster & Forster, 2003) and eye-tracking and 2AFC recognition in E-prime (Psychology software Inc.). Auditory stimuli were presented via a PC stereo speaker.

4.3.2. Language Training and Language Tests

4.3.2.1. Eye-tracking with training items

This task was used as part of training (exposure) to the novel words and also to assess whether participants could use the gender-marked determiners predictively during online language processing, thus providing a measure of automatic use of predictive cues of novel grammatical mappings.

On each trial, two pictures (a target and a distractor) were presented side by side for 2000 ms, followed by a centrally located fixation cross for 1000 ms. The fixation cross then disappeared, and an auditory presentation of the novel word began, with the two pictures remaining on screen. The target picture was the correct referent for the novel word. Of key interest was whether participants would orient to the correct referent (i.e., the target picture) more quickly on trials where the distractor differed in gender from the target (different gender trials) than on trials where the distractor was the same gender as the target (same gender trials). The instructions informed participants that two pictures would appear on the screen, followed shortly by a cross in the centre. They were instructed to look at the cross when it appeared, and shortly afterwards, they would hear a made-up word. Participants were told their task was simply to look at the pictures and listen carefully to the words.

There were four 5-minute blocks on days 1-3, each containing 32 pairs of items. Within each block, half of the trials were same gender trials, and half were different gender trials. Across the four blocks on each day, each item appeared twice as a target and twice as a distractor. Target-distractor pairs were not repeated across days, and the trial order was randomised such that the same target picture never occurred in consecutive trials.

The pictures were presented on a 22″ monitor, and the audio stimuli over a PC stereo speaker. Eye movements were recorded using an SMI RED 500 remote eye-tracker, sampling at 500 Hz with a spatial resolution of 0.03° and a gaze position accuracy of 0.4°. A 9 point calibration was conducted at the start of every block and was deemed successful if deviations reported by SMI were less than 2°.

Each eye-tracking task started with the calibration screen. After calibration was completed successfully, an instruction screen and two practice trials were presented at the start of the first block of each session.

4.3.2.2. Eye-tracking with generalisation items

On the final day of testing (Figure 4.1), an additional block of items was administered in the eye-tracking task, containing the 8 consistent items from the generalisation set described above. The procedure for this block was identical to previous blocks and there were 16 trials in total. In half of the trials, the grammatical gender of the distractor and target picture matched; however, in the other half, the distractor and target picture belonged to different gender categories. All trials included the items from the generalisation set.

4.3.2.3. Word-Picture Matching with training items

This task provided exposure to the novel words, while also serving as a measure of the possible emergence of explicit awareness of the grammatical mappings. The task procedure was similar to the Word–Picture Matching task used in Experiments 2 and 3 (Chapter 3 of this thesis), in that participants were asked to make a single integrated judgement, indicating how well they thought each word and picture went together, as well as their confidence in that judgement. However, unlike the Word–Picture Matching task in Experiments 2 and 3, where participants provided judgements on a 6-point Likert scale, the response scale in the current experiment used a 5-point Likert scale (see Figure 4.2).

Figure 4.2. Response scale for the Word-Picture Matching training and generalisation tasks.

On each day, participants were presented with 256 word-picture trials in total, consisting of 128 matched and 128 mismatched pairs. Across the 256 trials, they were exposed to each item from the training set 4 times. Specifically, in the matched trials, participants saw each correct word-picture pairing (e.g., *tib scoiffesh* + ballerina) twice. In the mismatched trials, each word was also presented twice but once with an incorrect picture from the *same* semantic category (e.g., *tib scoiffesh* + mermaid) and once again with an incorrect picture from the *opposite* semantic category (e.g., *tib scoiffesh* + male sailor). As in the Word-Picture Matching task used in Experiments 2 and 3, mismatched trials were included to allow for a measure of accuracy, as explicit knowledge was assessed by examining whether participants showed greater confidence for correct responses (endorsing matched trials, rejecting mismatched trials) compared to incorrect responses (rejecting matched trials, endorsing mismatched trials). Overall, across the 4 days, participants were presented with the correct word-picture pairing (e.g., *tib scoiffesh* + ballerina) on a total of 8 trials (twice on each day). Each incorrect word–picture pairing was presented only once across the four days.

Each trial in the Word-Picture Matching task began with a 500 ms centrally located fixation cross, followed by the presentation of a picture. The picture remained on screen for 500 ms before a novel word was played auditorily, while the picture remained on the screen. After the auditory presentation of the novel word, the five-point response scale (see Figure 4.2) appeared below the picture. The next trial began either when the participant made a response or after a 6000 ms timeout. The task took approximately 20 minutes to complete and was preceded by two practice trials. Participants were given an opportunity to take a break halfway through and were asked to resume the task once they felt that they were ready to continue. The presentation of items was randomised across participants. No feedback was provided.

4.3.2.4. Word-Picture Matching with generalisation items

On the final day of training, participants completed an additional 2-minute block of the Word-Picture Matching task, comprising the 16 novel items from the generalisation set. This task procedure was identical to the Word-Picture Matching generalisation task used in Experiments 2 and 3 in that half of the items in this task featured word-picture pairs that conformed to the grammatical mappings presented during training (consistent trials), while the other half presented pairings that did not conform to the trained mappings (inconsistent trials).

As in Experiments 2 and 3, this task served two purposes: to assess participants' explicit knowledge of the mappings when they were presented in new, untrained contexts and to provide a measure of generalisation. As with the trained items, inconsistent trials were included to allow for a measure of accuracy since explicit knowledge was assessed by examining whether participants would report greater confidence for correct responses (endorsing consistent trials, rejecting inconsistent trials) compared to incorrect responses (rejecting consistent trials, endorsing inconsistent trials). As in Experiments 2 and 3, to assess generalisation, endorsement rates for consistent versus inconsistent trials were used to derive an A' score (Pallier, 2002).

4.3.2.5 Picture Naming

As in previous experiments, this task was used to assess word and grammar learning. Accuracy of recall for the stem portion of each novel word (e.g., ***scoi*** in *scoiffesh*) was used to measure vocabulary learning, while recall of the determiner (e.g., *ked/tib*) and suffix (e.g., *ool* in *larshool*) components was used to assess learning of the grammatical mappings.

The procedure for this task was identical to that in previous experiments. Each trial began with the presentation of a fixation cross for 500 ms, followed by a picture. Both the fixation cross and image were presented at the centre of the screen. At the start of the task, participants were instructed to name each picture using words from the artificial language. If they could not recall the exact word, they were encouraged to name any part that they could remember. A keypress enabled participants to proceed to the next trial. All responses were audio-recorded and transcribed verbatim by the experimenter offline and scored according to the same criteria outlined in Chapter 2 (Experiment 1). The order of items was randomised for each participant and for each session. The task consisted of 64 trials in total, with one trial for each of the 64 words presented during training.

In the current experiment, the Picture Naming task was administered at the end of the training session on Day 1, and then both the start and end of each session on Days 2-4 (Figure 4.1), allowing us to examine changes in memory recall across the consecutive training sessions and offline consolidation periods. For example, to assess word learning at each session, accuracy (scored as a binary outcome measure) for the stem from the Picture Naming task administered at the end of each session was used. To evaluate changes in performance over offline consolidation, accuracy at the start of each session was compared to performance at the end of the previous session.

4.3.2.6 2AFC

On the final day of testing, a 2AFC task was administered on the eye-tracker. As in previous experiments, this task was used to assess participants' word knowledge. For this, the procedure was identical to the previous eye-tracking blocks; however, participants were also required to respond via a keypress (p or q) as to whether the word matched the picture on the left (q) or the picture on the right (p). The position of the target picture on the screen was counterbalanced across trials so that in half the trials, the target picture was presented on the left and for the other half it was presented on the right.

Participants completed 2 blocks, with each block containing 32 pairs of items. Thus, there were 64 trials in total, with one trial for each of the 64 words presented during training. As the aim of this task was to assess word learning, in contrast to the previous eye-tracking blocks, the two alternative pictures were always of the same grammatical category. This was to ensure that the participants would not be able to rely on the determiner and suffix cues to complete the task but instead, rely on their knowledge of the word itself (stem components).

4.3.2.7 Explicit Knowledge Questionnaire

As in Experiments 1, 2 and 3, at the end of the final session (Figure 4.1), participants were asked to complete a short questionnaire to assess their explicit knowledge of the underlying grammatical mappings.

4.4 Data Analyses

All analyses outlined below were conducted in R (R Core Team, 2020). This section describes the data analysis approach for the tasks assessing vocabulary learning first, followed by the tasks assessing learning of the grammatical mappings.

4.4.1 2AFC

This task was used to assess word learning and was administered on the final day of training (Figure 4.1). Due to a task programming error, the data for five participants were lost. Therefore, the analyses included the data for 31 participants. To analyse the data, a mixed-effects binomial regression with accuracy (0,1) as the outcome measure was conducted using the glmer function from the lme4 package in R (Bates et al., 2015). Mapping type was included as a fixed effect using sum contrast coding (prior knowledge-dependent = 1, prior knowledge-independent = -1). For the random-effects structure, participants and items were included as random intercepts, with a random slope for mapping type within the by-participant random effects. The package *buildmer* (Voeten, 2021) was used to establish the random-effect's structure, starting from the maximal model described above. The final random effects structure with the model that converged is provided in the results below.

4.4.2 Picture Naming

Due to a task programming error, the data for six participants were lost. Therefore, the analyses outlined below include the data for 30 participants.

4.4.2.1 Picture Naming: training

To assess learning across the training sessions, accuracy data (0,1) from the picture naming task administered at the end of each training session were analysed using three separate mixed-effects binomial regressions, one for each word component: stem, determiner, and suffix, using the glmer function from the lme4 package in R (Bates et al., 2015). Mapping type (prior knowledge dependent vs. independent) was included as a fixed effect, using sum contrast coding (prior knowledge dependent = 1, independent = –1). Day was also entered as a fixed effect with repeated contrasts applied using the contr.sdif() function from the MASS package in R (Venables & Ripley, 2002), allowing us to assess changes in performance relative to the previous training session. Random effects were specified for both participants and items, with a random slope for mapping type and day within the by-participant random effects and random slopes for day within the by-item random effects. The package *buildmer* (Voeten, 2021) was used to establish the random-effect's structure, starting from the maximal model described above. The final random effects structure with the model that converged is provided in the results below.

4.4.2.2 Picture Naming: offline consolidation

To assess changes in the memory of the novel words and grammatical mappings across offline consolidation periods (i.e., between training sessions), difference scores were calculated for each participant by subtracting their performance at the start of each session (red boxes in Figure 4.1) from their performance at the end of the previous session (blue boxes in Figure 4.1). This resulted in three consolidation periods: between Days 1 and 2 (calculated by subtracting the end of the Day 1 performance from the beginning of Day 2 performance; Figure 4.1), between Days 2 and 3 (Day 3(start) - Day 2(end), and between Days 3 and 4 (Day 4(start) - Day 3(end)). For each consolidation period, participants contributed one difference score per mapping condition (prior knowledge dependent vs.

independent) for each of the three outcome measures: determiners, stems, and suffixes. To analyse the data, three linear mixed-effects regressions were conducted, one for each word component (determiner, stem, and suffix), with difference scores as the outcome variable. Mapping type was included as a fixed effect, using sum contrast coding (prior knowledge dependent = 1, independent = –1). Consolidation period was also included as a fixed effect, with repeated contrasts applied using the contr.sdif() function from the MASS package in R (Venables & Ripley, 2002). This approach resulted in two contrast comparisons across the three consolidation periods, with the first contrast (consolidation contrast 1) comparing the difference scores between Days 1 and 2 to the difference scores between Days 2 and 3, and the second contrast (consolidation contrast 2) comparing the difference scores in recall between Days 2 and 3 with the difference scores between Days 3 and 4. Random effects were specified for both participants and items, with a random slope for consolidation period and mapping type within the by-participant random effects and random slopes for consolidation period within the by-item random effects. The package *buildmer* (Voeten, 2021) was used to establish the random-effect's structure, starting from the maximal model described above. The final random effects structure with the model that converged is provided in the results below.

4.4.3 Eye-tracking

4.4.3.1 Eye-tracking: training

The analyses of the eye movements followed the same approach as used by Lew-Williams and Fernald (2010). To assess whether participants could use the determiner cues predictively we compared the latency to initiate a gaze shift from the fixation cross toward the target picture from the determiner onset (RTs), on same gender trials versus different gender trials. If participants had learned the mappings between the determiner and semantic referent, it was expected that fixation to target shifts would begin more rapidly on different gender trials, where the determiner provides a cue as to the identity of the target referent. Additionally, we compared RTs between mapping types to examine whether gaze shifts would be more rapid for the targets in the prior knowledge dependent condition.

Following Lew-Williams and Fernald (2010), only RTs occurring between 300 and 1300 ms from determiner onset were included to account for the time it typically takes to plan the eye movements and the average duration of the novel words. Similarly, trials were only included if participants were fixating on the fixation cross at determiner onset.

Before proceeding with the analyses, we examined the proportion of trials in the two prior knowledge conditions that would be included in the analyses. For the prior knowledge dependent condition, the proportions of trials that met the above criteria were as follows: Day 1: 12% (different gender), 12% (same gender); Day 2: 17% (different gender), 17% (same gender); Day 3: 23% (different gender), 24% (same gender); Day 4: 29% (different gender), 25% (same gender). For the prior knowledge independent condition, the proportions of trials that met the criteria were: Day 1: 11% (different gender), 12% (same gender); Day 2: 18% (different gender), 17% (same gender); Day 3: 25% (different gender), 22% (same gender); and Day 4: 26% (different gender), 24% (same gender).

Based on these proportions, only Day 4 was selected for analysis, as this day had the highest proportion of trials meeting the inclusion criteria. To analyse the data, a linear mixed-effects regression was conducted on Day 4 RTs. Trial type (different gender = -1, same gender = 1) and mapping type (prior knowledge dependent = 1, prior knowledge independent = -1) were included as fixed effects using sum contrast coding. Random effects were specified for both participants and items with random slopes for trial type within the by-item random effects and random slopes for mapping type within the by-participant random effects. The package *buildmer* (Voeten, 2021) was used to establish the random-effect's structure, starting from the maximal model described above. The final random effects structure with the model that converged is provided in the results section.

4.4.3.2 Eye-tracking: generalisation

To determine whether participants could generalise the determiners by the final day of training (Day 4), we assessed RTs on items in the generalisation set using the same approach as above. For the prior knowledge dependent mapping condition, the proportions of trials that met the inclusion criteria were: 33% (different gender), 28% (same gender). For the prior knowledge independent mapping condition, the proportions of trials that met the inclusion criteria were: 39% (different gender), 29% (same gender).

As above, the data for the generalisation items were analysed using a linear mixed-effects regression, with RTs as the outcome. Trial type (different gender = -1, same gender = 1) and mapping type (prior knowledge dependent = 1, prior knowledge independent = -1) were included as fixed effects using sum contrast coding. Random effects were specified for both participants and items with random slopes for trial type within the by-item random effects and random slopes for mapping type within the by-participant random effects. The package *buildmer* (Voeten, 2021) was used to establish the random-effect's structure, starting from the maximal model described above. The final random effects structure with the model that converged is provided in the results section.

4.4.4 Word-Picture Matching

4.4.4.1 Word-Picture Matching: a' generalisation

This task was administered on the final day of training and provided an additional measure of generalisation.

As in Experiments 2 and 3, to evaluate generalisation of the underlying grammatical mappings, endorsement rates for consistent and inconsistent items were used to calculate an

A′ score (Pallier, 2002). Neutral responses (midpoint of the confidence scale: ‘I guess’) were excluded from the analyses, as they did not reflect either endorsement or rejection of a word-picture pairing. These responses were treated as missing (NA), which led to slight differences in the number of observations across mapping conditions.

In this analysis, endorsing a consistent trial was treated as a hit, while endorsing an inconsistent trial was considered a false alarm. A′ values ranged from 0 to 1, with a score of 0.5 indicating chance-level performance (i.e., equal endorsement of consistent and inconsistent trials). Scores below 0.5 reflect greater endorsement of inconsistent trials, whereas scores above 0.5 indicate higher endorsement rates for consistent trials over inconsistent ones and are therefore taken as evidence of successful generalisation of the mappings to new items.

To assess the extent of generalisation for both prior knowledge dependent and prior knowledge independent mappings, one-tailed t-tests were conducted to compare a’ scores against 0.5. These tests evaluated whether participants endorsed consistent items at rates significantly above chance, indicating successful generalisation. In addition, a linear mixed regression was conducted to assess differences in generalisation performance for the two mapping types. The package *buildmer* (Voeten, 2021) was used to establish the random-effects structure, starting from the maximal model, which included random intercepts for participants and a by-participant slope for mapping type. The final random effects structure with the model that converged is provided in the results section.

4.4.4.2 Word-Picture Matching: Explicit Knowledge (training)

This analysis was used to assess participants’ explicit knowledge of the underlying grammatical mappings when presented in their trained item contexts. Of key interest was whether participants would develop awareness of the underlying grammatical mappings during the four-day training period. In addition, we also assessed whether levels of explicit knowledge would be greater for mappings related to existing knowledge compared to those that were unrelated.

To analyse the accuracy of participants' word-picture matching judgements, endorse responses (the two right-most smileys in Figure 4.2) and reject responses (the two left-most smileys in Figure 4.2) were used. This scoring followed the same procedure as that used and outlined in Experiments 2 and 3 (Chapter 3 of this thesis). As outlined in Chapter 3, endorsing a 'matched/consistent' item and rejecting a 'mismatched/inconsistent' item in this task would be considered a correct response. Conversely, rejecting a 'matched/consistent' item and endorsing a 'mismatched/inconsistent' item would be considered incorrect responses. If confidence ratings are higher for correct responses than for incorrect responses, this would indicate a relationship between confidence and accuracy, suggesting the presence of explicit knowledge. However, if there is no significant difference in confidence ratings between correct and incorrect responses, this would indicate no relationship between confidence and accuracy and suggest an absence of explicit knowledge.

For the confidence ratings, scoring again followed the same procedure as that used in Experiments 2 and 3 with responses converted into scores indicating high or low confidence. Responses of 'I am certain', whether on the reject side of the scale or the endorse side of the scale (Figure 4.2), were classified as high confidence and assigned a score of 2. Responses of 'I think' on either side of the scale were classified as low confidence and assigned a score of 1. Neutral responses (the midpoint of the confidence scale: 'I guess') were not included in the analysis.

To analyse the Word-Picture Matching data with the trained items, we conducted a linear mixed regression, with confidence ratings as the outcome measure and accuracy, mapping type, and day coded as fixed effects, along with their interactions. Both accuracy (correct = 1, incorrect = −1) and mapping type (prior knowledge dependent = 1, prior knowledge independent = −1) were coded using sum contrast coding. For day, repeated contrasts were applied using the contr.sdif() function from the MASS package (Venables & Ripley, 2002) in R. This allowed us to compare confidence ratings for each day to those of the previous day and, in doing so, assess changes in performance across the four-day training period. Random effects were included for both participants and items, with random slopes for day and mapping type nested within the by-participant random intercepts, and random slopes for day within the by-item random intercepts. The package *buildmer* (Voeten, 2021) was used to establish the random effects structure, starting from the maximal model described above. The final random effects structure and model that converged for each analysis are reported in the results section.

4.4.4.3 Word-Picture Matching: Explicit Knowledge (generalisation)

This task was administered on the final day of testing and was used to assess participants' explicit knowledge of the underlying grammatical mappings when presented in new, untrained contexts. As with the trained items, we also examined whether explicit knowledge would be greater for those mappings related to existing knowledge compared to those that were unrelated.

Scoring of responses followed the same procedure as that for the Word-Picture Matching task with the trained items (outlined above).

To analyse the Word-Picture Matching data with the generalisation items, a linear mixed regression was used, with confidence ratings as the outcome measure and accuracy (correct = 1, incorrect = -1) and mapping type (prior knowledge dependent = 1, prior knowledge independent = -1) coded as fixed effects, along with their interaction. Both fixed effects were coded using sum contrast coding. Random effects were included for both participants and items, with random slopes for mapping type within the by-participant random effects. The package *buildmer* (Voeten, 2021) was used to establish the random-effect's structure, starting from the maximal model described above. The final random effects structure with the model that converged for each model is provided in the results below.

4.4.5 Explicit Knowledge Questionnaire

This questionnaire was administered at the end of the final session and was used to assess participants' explicit knowledge of the underlying grammatical mappings. Responses were scored using the same coding scheme as that used in Experiments 1, 2, and 3. Each participant received a score between 0 and 2 for both the determiner and suffix mappings, with separate scores calculated for the prior knowledge dependent and independent conditions. To investigate the role of prior knowledge in explicit awareness of the trained mappings, two linear mixed-effects models were conducted using the lmer function from the lme4 package in R. The first model analysed awareness scores for the determiner mappings, and the second for the suffix mappings. In both models, mapping type (prior knowledge dependent = 1, independent = –1) was included as a fixed effect, and random intercepts were included for participants. The package *buildmer* (Voeten, 2021) was used to establish the random-effect's structure. The final model that converged is presented in the results section.

4.5 Results

This section presents the results for each task, starting with vocabulary learning first, followed by grammar learning and then explicit knowledge.

4.5.1 Word Learning

4.5.1.1 2AFC

This task was administered on the final day of training to assess how well participants had learned the novel words in this larger and more complex language following the four-day training period. We also examined whether the prior knowledge manipulation of the grammatical mappings influenced word knowledge by comparing performance across the two mapping conditions.

As shown in Figure 4.3, by the end of training, participants' word knowledge was high in both conditions (Prior Knowledge Dependent: M = .90; Prior Knowledge Independent: M = .86) and the main effect of mapping did not reach statistical significance (p = .054, Table 4.2)

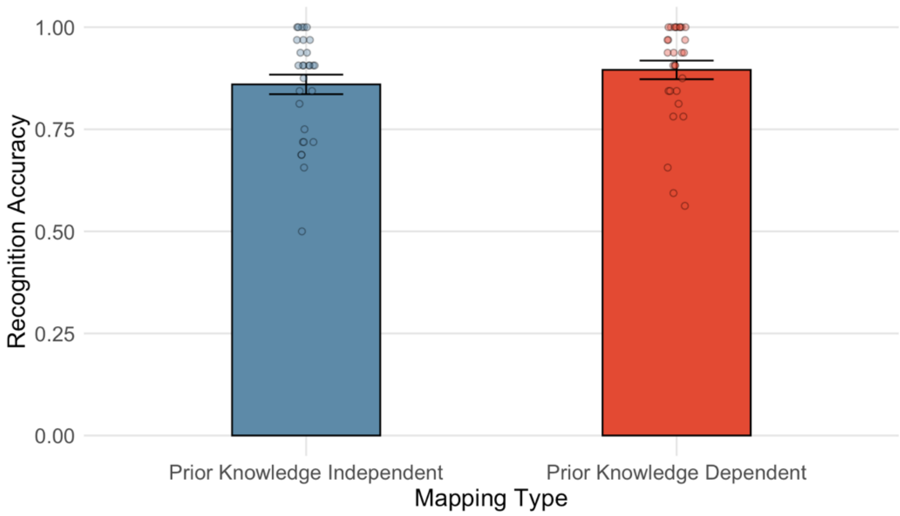

Figure 4.3. Accuracy (proportion correct) on the 2AFC task measuring word learning. Error bars represent the standard error.

Table 4.2. Regression coefficients for the fixed factor using recognition accuracy as a binary outcome measure in the 2AFC word learning task.

	Estimate	Std. Error	z	p
Mapping	0.196	0.102	1.925	.054

Note. Random effects included in the final model that converged: by-participant and by-item intercepts.

4.5.1.2 Picture Naming (training): Stems

The time course of word learning over the four sessions was assessed using the Picture Naming task administered at the end of each session. As shown in Figure 4.4, recall accuracy for the stems was low overall but similar to the previous experiments by the end of training on Day 4. As indicated by the significant main effect of the day contrasts (Table 4.3), participants showed steady improvements following each training session relative to the previous day. No significant main effect of mapping or interaction between mapping and day was observed (Table 4.3), suggesting that the prior knowledge manipulation of the grammatical mappings did not influence stem learning.

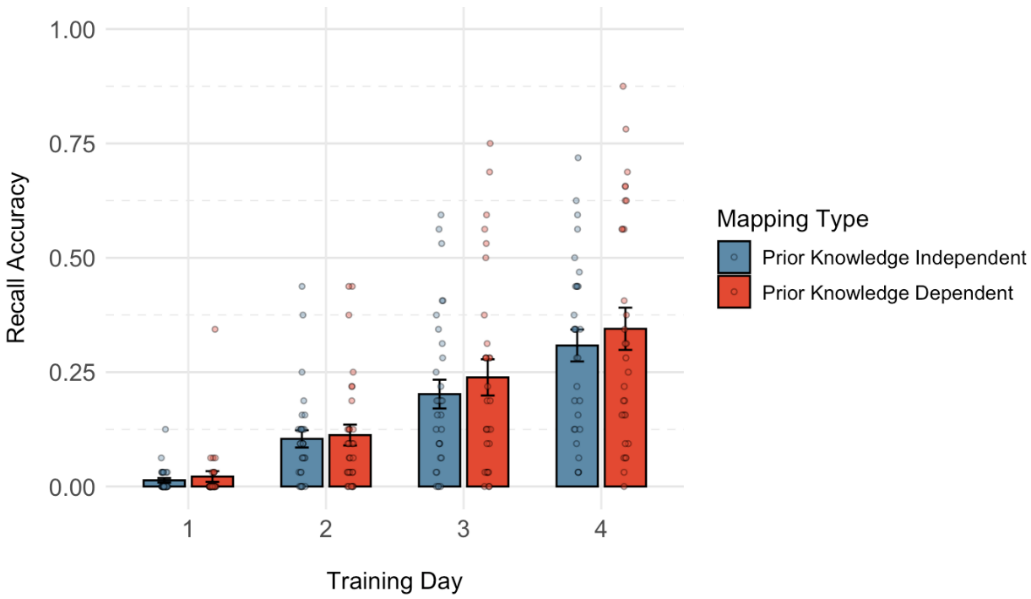

Figure 4.4. Stem recall accuracy (proportion correct) on the Picture Naming task administered at the end of each training session. Error bars represent standard error.

Table 4.3 Regression coefficients for the fixed factors using stem recall accuracy in the Picture Naming tasks administered at the end of each training session as the outcome.

	Estimate	Std. Error	z	p
Day2vs.Day1	**2.237**	**0.202**	**11.060**	**<.001**
Day3vs.Day2	**1.267**	**0.106**	**10.631**	**<.001**
Day4vs.Day3	**0.779**	**0.088**	**8.332**	**<.001**
Mapping	0.106	0.142	0.743	.457
Day2vs.Day1 x Mapping	-0.204	0.202	-1.009	.313
Day3vs.Day2 x Mapping	0.104	0.106	0.979	.328
Day4vs.Day3 x Mapping	-0.016	0.088	-0.172	.863

Note. Random effects included in the final model that converged: intercepts by participant and items. Significant results are highlighted in bold.

4.5.1.3 Picture Naming (offline consolidation): Stems

To evaluate changes in stem recall across offline consolidation periods, two successive difference contrasts were used to compare recall changes across the three consolidation periods (between Days 1 and 2, between Days 2 and 3, and between Days 3 and 4). Specifically, the contrast labelled 'consolidation contrast 1' compared the difference scores reflecting offline consolidation between the end of the Day 1 session and the start of the Day 2 session with the difference scores reflecting offline consolidation between the end of the Day 2 session and the start of the Day 3 session. Similarly, consolidation contrast 2 compared the difference scores reflecting offline consolidation between the end of the Day 2 session and the start of the Day 3 session with the difference scores reflecting offline consolidation between the end of the Day 3 session and the start of the Day 4 session.

As illustrated in Figure 4.5, there was little change in recall accuracy over the first two consolidation periods, and no significant effect for the consolidation contrast 1 or for the consolidation contrast 2. The main effect of mapping was also not significant (Table 4.4).

For the interactions, neither the Consolidation Contrast 1 × Mapping nor the Consolidation Contrast 2 × Mapping interaction reached significance. However, as shown in Figure 4.5, the change from Day 4 to Day 3 suggests some benefit for the stems in the prior knowledge independent condition, although this did not reach significance (Table 4.4).

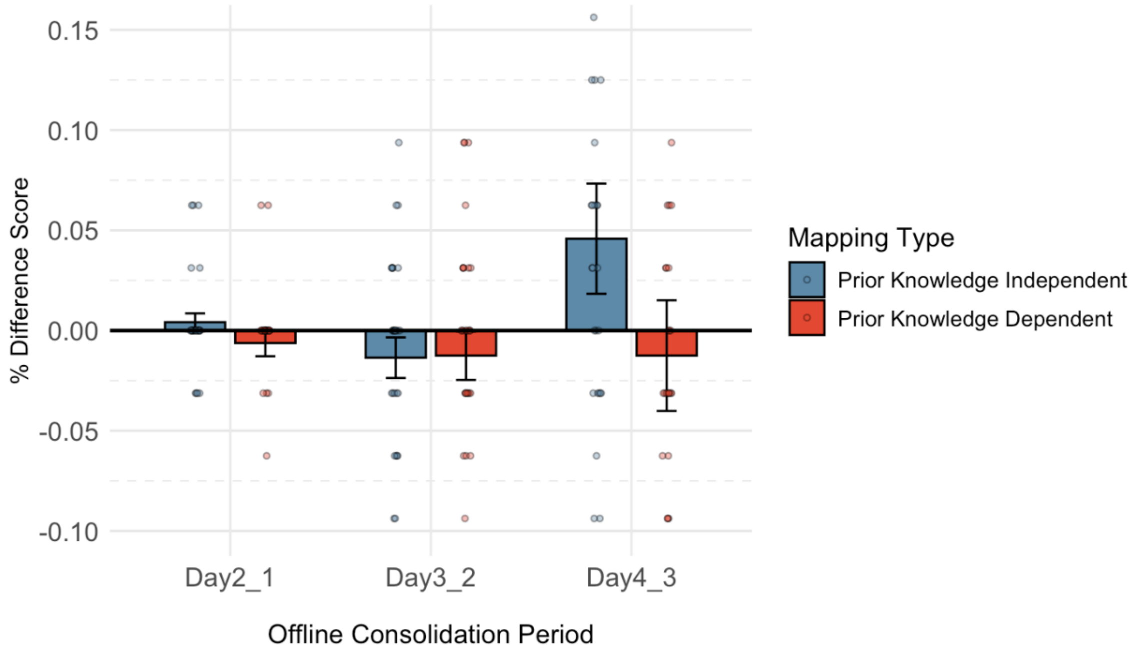

Figure 4.5. Difference scores comparing end of session vs. start of session on the following day on the Picture Naming task for stem accuracy (proportion correct). Day2_1 = difference between the end of Day 1 and the start of Day 2; Day3_2 = difference between the end of Day 2 and the start of Day 3; Day4_3 = difference between the end of Day 3 and the start of Day 4. Positive values indicate improvements in accuracy, while negative values indicate decreases in accuracy. Error bars represent standard error.

Table 4.4 Regression coefficients for the fixed factors using difference scores in stem recall accuracy in the Picture Naming task as the outcome.

	Estimate	Std. Error	t	p
Consolidation Contrast 1	-0.012	0.017	-0.716	.474
Consolidation Contrast 2	0.030	0.017	1.774	.076
Mapping	-0.011	0.007	-1.652	.099
Consolidation Contrast 1x Mapping	0.005	0.017	0.342	.732
Consolidation Contrast 2x Mapping	-0.030	0.017	-1.774	.076

Note. Random effects included in the final model that converged: intercepts by participant with random slopes for the consolidation contrasts.

4.5.2 Grammar Learning

Participants' recall accuracy for the determiner and suffix mappings was also assessed using the Picture Naming task. As with stem recall, memory recall was measured at the end of each training session, and offline changes in recall were evaluated by comparing accuracy at the start of each session to accuracy at the end of the previous session. Additionally, implicit learning and generalisation of the determiner mappings were examined using an eye-tracking task, which assessed participants' ability to exploit the gender-marked determiners during online processing of novel words. Finally, generalisation of the determiner and suffix mappings to novel, previously unseen items was evaluated with the Word–Picture Matching task, providing an additional and more direct measure of generalisation alongside the eye-tracking task.

4.5.2.1 Picture Naming: training

Determiners. As illustrated in Figure 4.6 and as indicated by the significant main effect of the day contrasts (Table 4.5), participants' recall accuracy for the determiner mappings improved with each training session. In addition, a significant main effect of mapping was observed, with greater recall accuracy for the determiner mappings in the prior knowledge dependent condition compared to the independent condition (Table 4.5). This advantage emerged on Day 3, as reflected in the significant interaction between mapping and the Day 3 vs. Day 2 contrast (Table 4.5). None of the other interactions between mapping and day were significant.

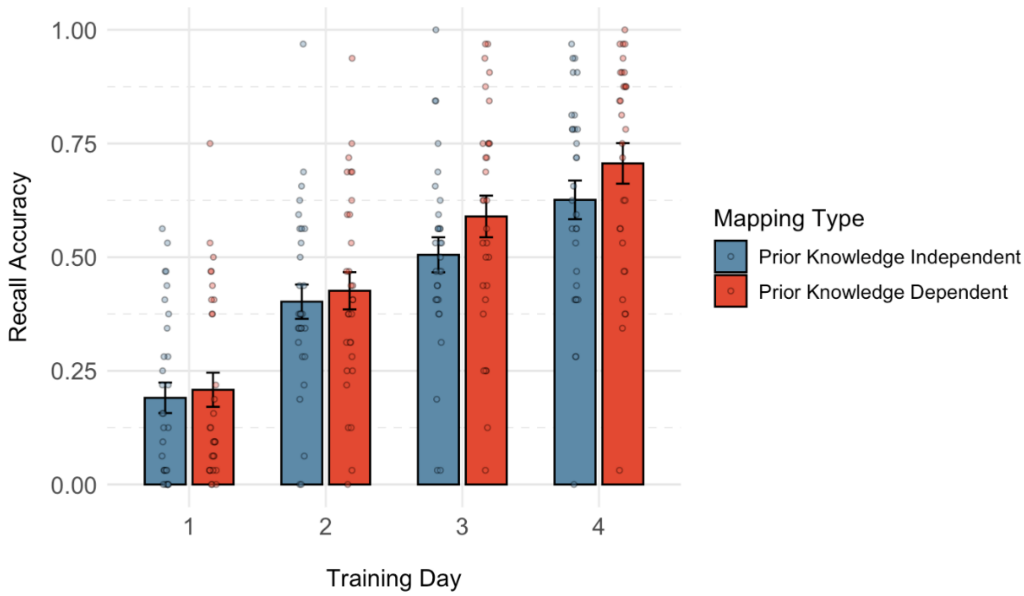

Figure 4.6 Determiner recall accuracy (proportion correct) on the Picture Naming task administered at the end of each training session. Error bars represent standard error.

Table 4.5 Regression coefficients for the fixed factors using determiner recall accuracy in the Picture Naming tasks administered at the end of each training session as the outcome.

	Estimate	Std. Error	z	p
Day2vs.Day1	**1.555**	**0.245**	**6.352**	**<.001**
Day3vs.Day2	**0.699**	**0.152**	**4.592**	**<.001**
Day4vs.Day3	**0.677**	**0.118**	**5.752**	**<.001**
Mapping	**0.163**	**0.072**	**2.277**	**.023**
Day2vs.Day1 x Mapping	0.002	0.084	0.012	.990
Day3vs.Day2 x Mapping	**0.145**	**0.073**	**1.989**	**.047**
Day4vs.Day3 x Mapping	0.046	0.077	0.601	.548

Note. Random effects included in the final model that converged: intercepts by participant and items. For participants, slopes were included for mapping and the day contrasts. Significant results are highlighted in bold.

Suffixes. As with determiners, participants' recall accuracy for suffixes improved across training sessions (Figure 4.7), with significant main effects observed for each of the day contrasts (Table 4.6). However, unlike the results for the determiners, neither the main effect of mapping nor any of the interactions between mapping and day were statistically significant.

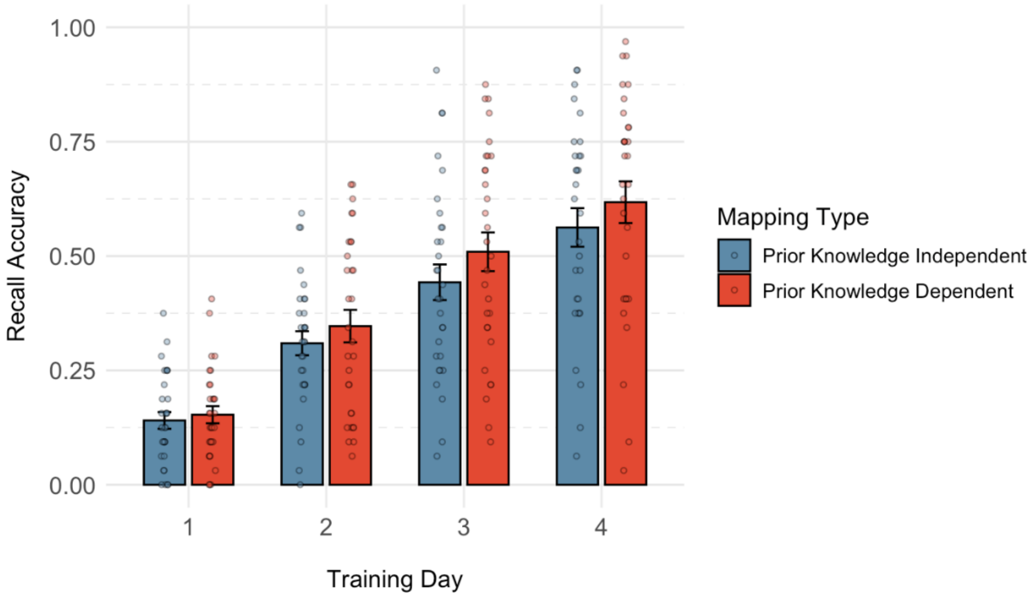

Figure 4.7. Suffix recall accuracy (proportion correct) on the Picture Naming task administered at the end of each training session. Error bars represent standard error.

Table 4.6 Regression coefficients for the fixed factors using suffix recall accuracy in the Picture Naming tasks administered at the end of each training session as the outcome.

	Estimate	Std. Error	z	p
Day2vs.Day1	**1.182**	**0.156**	**7.534**	**<.001**
Day3vs.Day2	**0.760**	**0.103**	**7.330**	**<.001**
Day4vs.Day3	**0.622**	**0.084**	**7.359**	**<.001**
Mapping	0.127	0.087	1.476	.140
Day2vs.Day1 x Mapping	0.037	0.056	0.683	.670
Day3vs.Day2 x Mapping	0.065	0.074	0.871	.384
Day4vs.Day3 x Mapping	-0.022	0.075	-0.293	.769

Note. Random effects included in the final model that converged: intercepts by participant and items. For participants, slopes were included for the day contrasts. Significant results are highlighted in bold.

4.5.2.2. Picture Naming: offline consolidation

Determiners. During each offline consolidation period, participants showed a decline in their recall accuracy of the determiners (Figure 4.8). However, there was a reduction in this decline at the third offline consolidation period relative to the second consolidation period as indicated by the significant main effect of consolidation contrast 2 (Table 4.7, Figure 4.8), which compared difference scores between the beginning of the Day 4 and the end of the Day 3 sessions with the difference scores between the beginning of the Day 3 and the end of Day 2 sessions. No other main effects or interactions were observed (see Table 4.7).

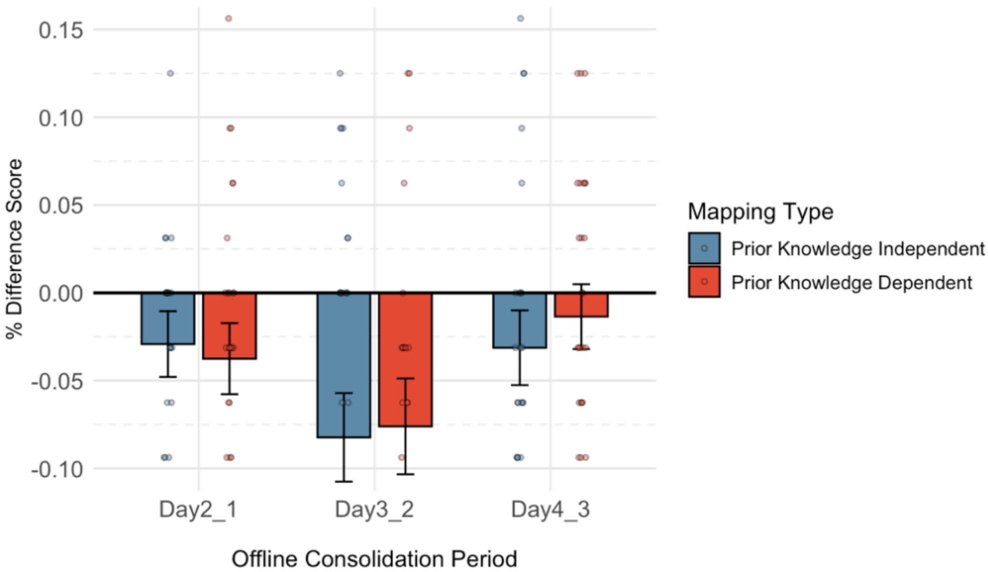

Figure 4.8. Difference scores comparing end of session vs. start of session on the following day on the Picture Naming task for determiner accuracy (proportion correct). Day2_1 = difference between the end of Day 1 and the start of Day 2; Day3_2 = difference between the end of Day 2 and the start of Day 3; Day4_3 = difference between the end of Day 3 and the start of Day 4. Positive values indicate improvements in accuracy, while negative values indicate decreases in accuracy. Error bars represent standard error.

Table 4.7 Regression coefficients for the fixed factors using mean difference scores in determiner recall accuracy in the Picture Naming task as the outcome.

	Estimate	Std. Error	t	p
Consolidation Contrast 1	-0.046	0.025	-1.808	.071
Consolidation Contrast 2	**0.057**	**0.023**	**2.498**	**.013**
Mapping	0.003	0.007	0.398	.691
Consolidation Contrast 1x Mapping	0.007	0.016	0.454	.650
Consolidation Contrast 2x Mapping	0.006	0.016	0.357	.721

Note. Random effects included in the final model that converged: intercepts by participant with random slopes for the consolidation contrasts. Significant results are highlighted in bold.

Suffixes. Like the determiners, participants exhibited a decline in their recall accuracy of the suffixes during each offline consolidation period (Figure 4.9). However, no main effects were observed for either of the consolidation contrasts, indicating that the change in performance did not significantly differ across consolidation intervals (Table 4.8). The main effect of mapping and the interactions between mapping and consolidation period were also not significant.

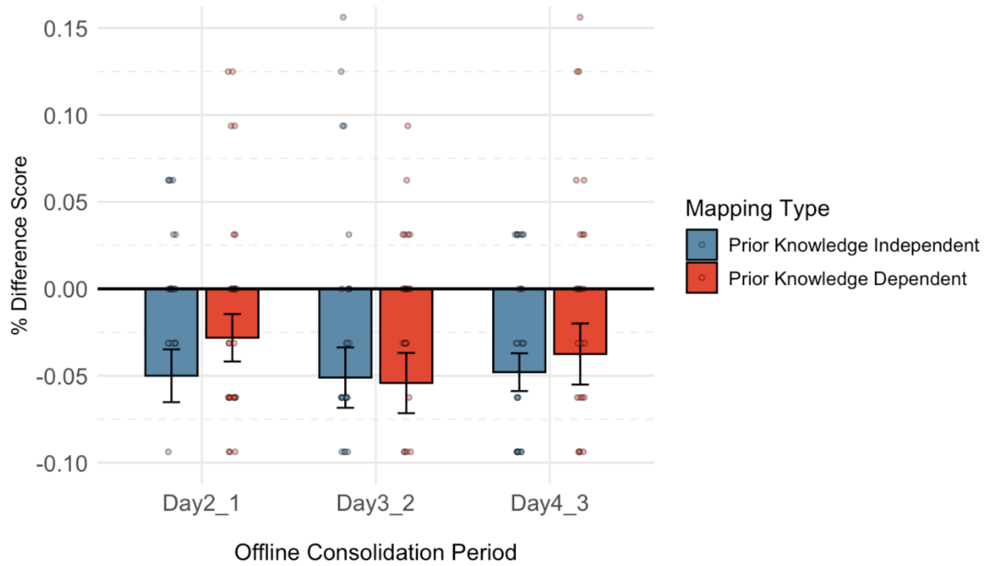

Figure 4.9. Difference scores comparing end of session vs. start of session on the following day on the Picture Naming task for suffix accuracy (proportion correct). Day2_1 = difference between the end of Day 1 and the start of Day 2; Day3_2 = difference between the end of Day 2 and the start of Day 3; Day4_3 = difference between the end of Day 3 and the start of Day 4. Positive values indicate improvements in accuracy, while negative values indicate decreases in accuracy. Error bars represent standard error.

Table 4.8 Regression coefficients for the fixed factors using mean difference scores in suffix recall accuracy in the Picture Naming task as the outcome.

	Estimate	Std. Error	t	p
Consolidation Contrast 1	-0.014	0.015	-0.913	.361
Consolidation Contrast 2	0.010	0.015	0.668	.504
Mapping	0.005	0.007	0.738	.460
Consolidation Contrast 1x Mapping	-0.013	0.015	-0.843	.399
Consolidation Contrast 2x Mapping	0.007	0.015	0.457	.648

Note. Random effects included in the final model that converged: intercepts by participant with random slopes for the consolidation contrasts.

In sum, participants showed improvements in their recall accuracy of the determiners and suffixes across training. Additionally, for the determiners, recall was greater for mappings in the prior knowledge dependent condition compared to those in the prior knowledge independent condition, with this benefit emerging on Day 3 of training. However, no such benefit was observed for the suffix mappings. Moreover, in contrast to the gains observed during training, recall accuracy declined for both determiners and suffixes during each offline consolidation period.

4.5.2.3 Eye-tracking: training

This task was used to assess the use of determiner cues predictively during real-time processing at the end of training. As shown in Table 4.9, the main effects of trial type, mapping, and their interaction was not significant, suggesting that participants were unable to use determiner cues predictively at the end of training, and prior knowledge did not modulate this ability.

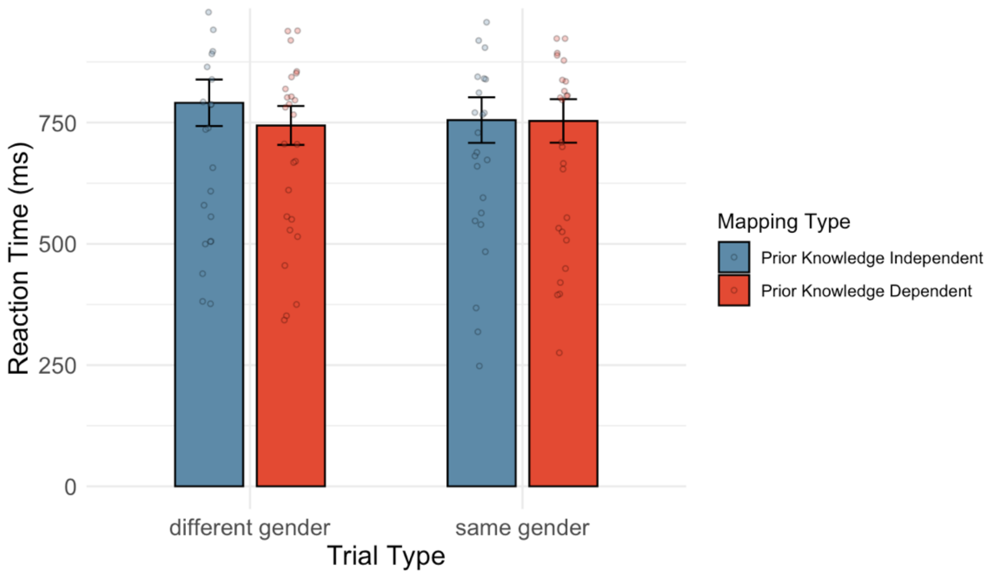

Figure 4.10. RTs for the Eye-tracking task with the trained items, administered on Day 4. Error bars represent standard error.

Table 4.9 Regression coefficients for the fixed factors using RTs for the trained items on the Eye-tracking task, administered on Day 4 as the outcome.

	Estimate	Std. Error	t	p
Trial type	-2.887	8.556	-0.337	.736
Mapping	-3.262	9.382	-0.348	.728
Trial type x Mapping	7.454	8.593	0.867	.386

Note. Random effects included in the final model that converged: intercepts by participant and item. For item, slopes were included for the trial type contrasts.

4.5.2.4 Eye-tracking: generalisation

As with the trained items, when the determiner mappings were presented with novel, untrained items, there was no evidence of using the determiner cues predictively as shown by the non-significant main effects of trial type or mapping or their interaction (Table 4.10).

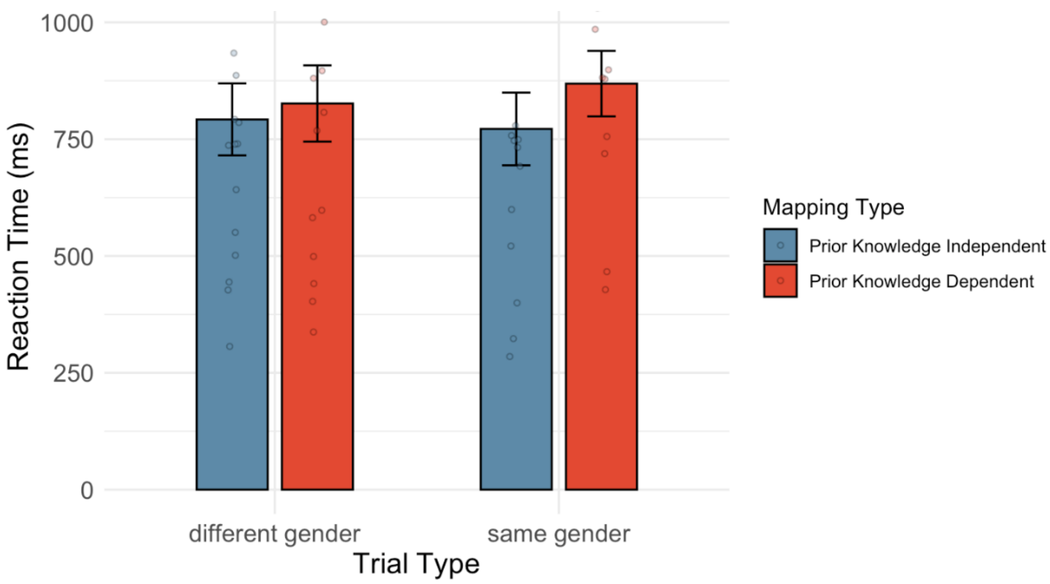

Figure 4.11. RTs for the Eye-tracking task with the generalisation items, administered on Day 4. Error bars represent standard error.

Table 4.10 Regression coefficients for the fixed factors using RTs for the generalisation items on the Eye-tracking task, administered on Day 4 as the outcome.

	Estimate	Std. Error	t	p
Trial type	17.25	34.31	0.503	.615
Mapping	10.06	28.68	0.351	.726
Trial type x Mapping	-12.86	34.22	-0.376	.707

Note. Random effects included in the final model that converged: intercepts by participant and item. For item, slopes were included for the trial type contrasts.

4.5.2.5 Word-Picture Matching: a' generalisation

The Word-Picture Matching task was used to assess participants' ability to generalise the grammatical mappings to new, untrained items on the final day of training in a task that required them to make a decision about the matching between previously unseen novel words and pictures. We also examined whether participants' ability to generalise differed based on mapping type.

One-tailed t-tests indicated that participants showed evidence of generalisation of the mappings in both the prior knowledge dependent condition (Figure 4.12; $t(30) = 2.79$, $p = .005$) and the prior knowledge independent condition (Figure 4.12; $t(29) = 2.52$, $p = .009$) in this task. There was no evidence that prior knowledge influenced generalisation as indicated by the non-significant main effect of mapping (Table 4.11).

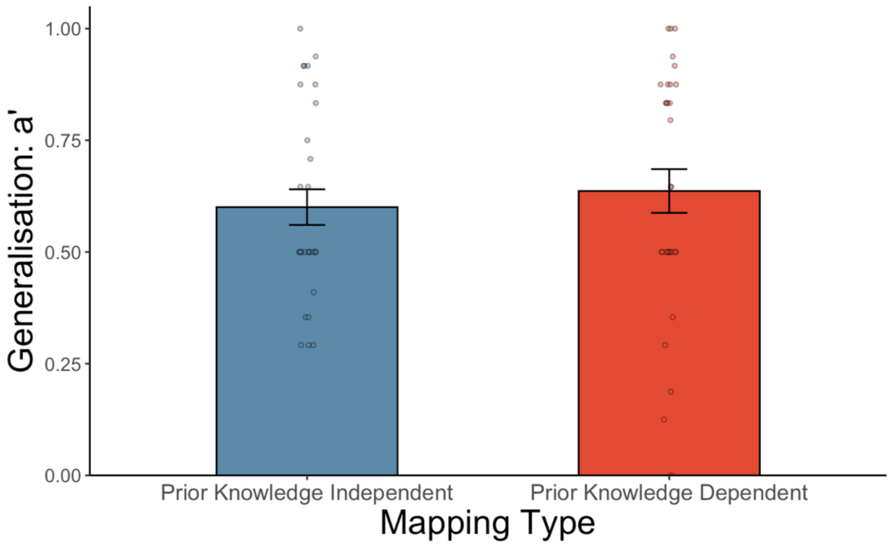

Figure 4.12. Generalisation performance for the determiner and suffix prior knowledge dependent and independent mappings. Error bars represent standard error.

Table 4.11 Regression coefficients for the fixed factors using a' scores as the outcome measure in the Word-Picture Matching task.

	Estimate	Std. Error	t	p
Mapping	0.019	0.031	0.598	.550

Note. Random effects included in the final model that converged: intercepts by participant.

In summary, participants showed improved accuracy for the determiner and suffix mappings following each training session as measured by the Picture Naming task. For the determiner mappings, a prior knowledge benefit emerged on Day 3, with participants demonstrating greater recall for determiners belonging to the prior knowledge dependent condition. In contrast, no such benefit was observed for suffix recall on any of the training days. Additionally, recall for both the prior knowledge dependent and independent determiner and suffix mappings declined over offline periods. For determiners, the rate of decline was greater during the second consolidation interval (between Day 2 and Day 3) compared to the third interval (between Day 3 and Day 4).

The eye-tracking task assessed the use of determiners as predictive cues in online processing at the final session. The results of the analysis showed that participants were unable to use the gender-marked determiners predictively to anticipate the upcoming novel word on different-gender trials, where the determiner provided an informative cue. This same pattern was also observed for the eye-tracking task where the determiner mappings were presented in the context of novel, untrained items.

Finally, participants did show evidence of generalisation when assessed using Word-Picture Matching that required them to make a decision about whether the previously unseen novel word and pictured matched, where some were consistent and some inconsistent with the newly learned regularities. However, overall performance was weak, and participants were no better at generalising the mappings in the prior knowledge dependent condition compared to the independent condition.

4.5.3 Explicit Knowledge

Explicit knowledge of the mappings was assessed using the Word–Picture Matching task in all sessions. In addition, retrospective knowledge was measured using the Explicit Knowledge Questionnaire, which was administered at the end of the final session.

4.5.3.1 Word-Picture Matching: Explicit knowledge

To assess the possible emergence of explicit grammatical knowledge across training, a zero-correlation analysis was conducted using participants' confidence ratings from the five-point response scale in the Word–Picture Matching task, which was administered on each of the four training days. On the final day (Day 4), an additional block of this task was administered using previously unseen items to assess explicit knowledge of the mappings in novel, untrained contexts.

Training. A linear mixed regression revealed a main effect of accuracy (Table 4.12), with significantly higher confidence ratings for correct compared to incorrect responses. A significant interaction between accuracy and Day2 vs. Day1 contrast was also observed, indicating that the relationship between confidence and accuracy was greater on Day 2 than on Day 1 (Figure 4.13, Table 4.12). In addition, the Day 3 vs. Day 2 contrast was significant, showing an overall increase in confidence ratings observed on Day 3 relative to Day 2 (Figure 4.13). The overall increase in confidence remained stable between Days 3 and 4, as indicated by no significant effect of the Day 4 vs. Day 3 contrast.

Confidence did not differ by mapping type as indicated by the non-significant main effect of mapping (Table 4.12). The interaction between mapping type and accuracy was also not significant, suggesting that levels of explicit knowledge (i.e., the relationship between confidence and accuracy) were similar for both the prior knowledge dependent and independent mappings. No other main effects or interactions were significant (Table 4.12).

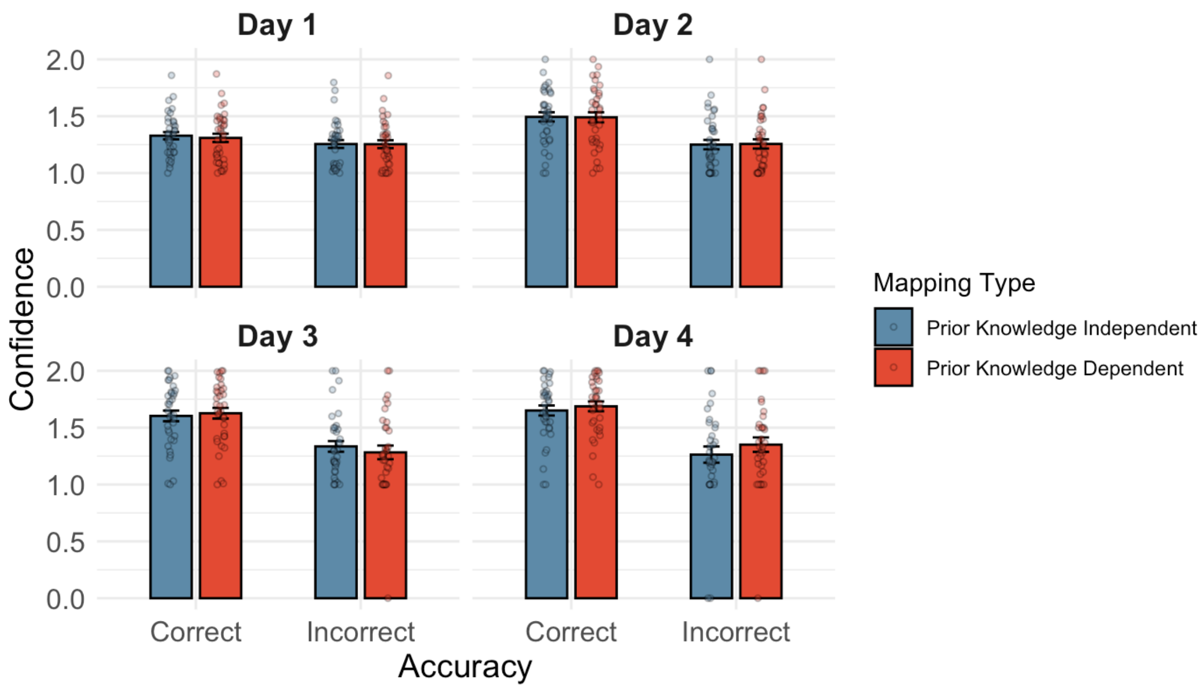

Figure 4.13. Explicit Knowledge: Mean confidence ratings for correct and incorrect responses in the Word-Picture Matching training task. Error bars represent standard error.

Table 4.12 Regression coefficients for the fixed factors using confidence ratings in the Word-Picture Matching training task as the outcome.

	Estimate	Std. Error	t	p
Accuracy	**0.180**	**0.008**	**21.609**	**<.001**
Accuracy × Day2vs.Day1	**0.131**	**0.021**	**6.333**	**<.001**
Accuracy × Day3vs.Day2	0.020	0.023	0.877	.380
Accuracy × Day4vs.Day3	0.039	0.025	1.552	.121
Day2vs.Day1	0.029	0.053	0.534	.593
Day3vs.Day2	**0.093**	**0.045**	**2.082**	**.037**
Day4vs.Day3	0.014	0.039	0.350	.726
Mapping	-0.000	0.019	-0.024	.981
Accuracy × Mapping	0.015	0.012	1.317	.188

Mapping × Day2vs.Day1	-0.006	0.023	-0.261	.794
Mapping × Day3vs.Day2	-0.005	0.029	-0.188	.851
Mapping × Day4vs.Day3	0.051	0.033	1.523	.128
Accuracy Day2vs.Day1× Mapping	0.033	0.029	1.133	.257
Accuracy×Day3vs.Day2× Mapping	0.028	0.032	0.858	.390
Accuracy×Day4vs.Day3× Mapping	-0.045	0.036	-1.256	.209

Note. Random effects included in the final model that converged: intercepts by participant and items. For participants, random slopes were included for mapping and day contrasts. Significant results are highlighted in bold.

Generalisation. As shown in Table 4.13, for the Word-Picture Matching task with the generalisation items, none of the main effects or their interaction were statistically significant.

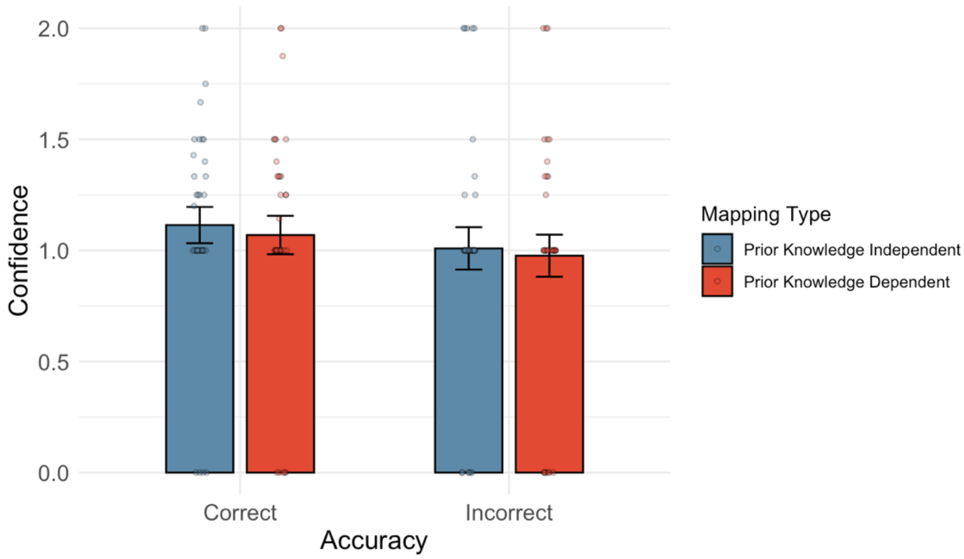

Figure 4.14 Explicit Knowledge: Mean confidence ratings for correct and incorrect responses in the Word-Picture Matching generalisation task. Error bars represent standard error.

Table 4.13 Regression coefficients for the fixed factors using confidence ratings in the Word-Picture Matching generalisation task as the outcome.

	Estimate	Std. Error	t	p

Accuracy	0.018	0.018	1.005	.315
Mapping	0.014	0.026	0.534	.593
Accuracy x Mapping	-0.007	0.018	-0.364	.716

Note. Random effects included in the final model that converged: intercepts by participant and items. For participants, random slopes were included for mapping contrasts.

Overall, these findings suggest that participants developed explicit knowledge of the mappings in their trained item contexts, with evidence of this awareness emerging by the second day of training. However, this awareness did not generalise to novel contexts, as confidence ratings were not related to accuracy in the Word–Picture Matching task with the generalisation items.

4.5.3.2 Explicit Knowledge Questionnaire

This questionnaire was administered at the end of the final training session and assessed explicit knowledge of the underlying grammatical mappings. Responses were scored separately for the determiner (both prior knowledge dependent and independent) and suffix mappings (both prior knowledge dependent and independent).

As shown in Figure 4.15, participants' self-reported explicit knowledge was low overall.

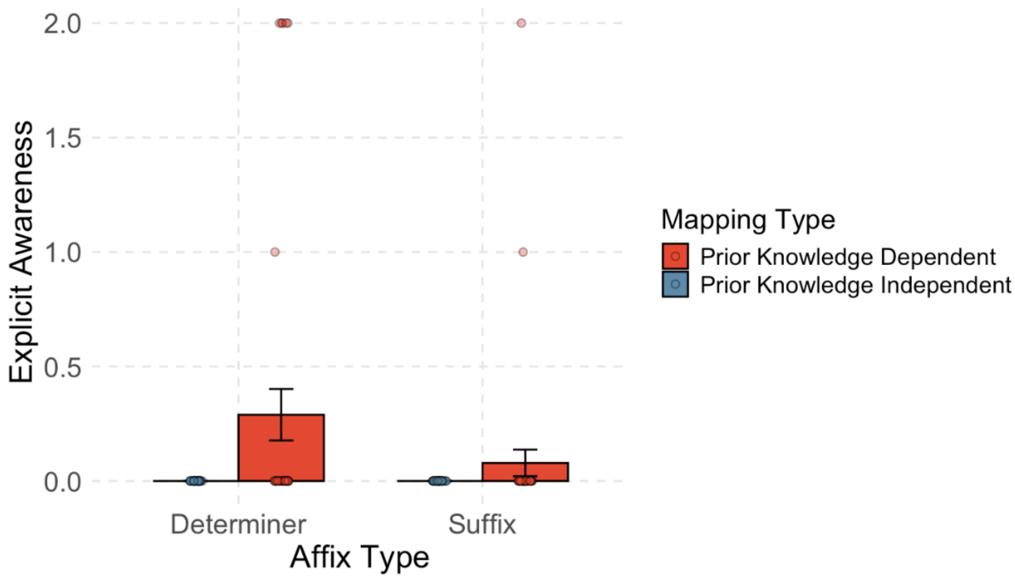

Figure 4.15 Levels of explicit awareness of the determiners and suffixes in the Explicit Knowledge Questionnaire. Error bars represent standard error.

A mixed-effects analysis found a significant main effect of mapping for the determiner mappings, with higher explicit knowledge scores observed for the determiners in the prior knowledge dependent condition (Table 4.14). In contrast, explicit knowledge for the suffix mappings was generally low, and the main effect of mapping was not significant (Table 4.15).

Table 4.14 Regression coefficients for the fixed factor using determiner responses in the Explicit Knowledge Questionnaire as the outcome.

	Estimate	Std. Error	t	p
Mapping	**0.145**	**0.056**	**2.527**	**.010**

Note. Random effects included in the final model that converged: by-participant intercepts.

Table 4.15 Regression coefficients for the fixed factor using suffix responses in the Explicit Knowledge Questionnaire as the outcome.

	Estimate	Std. Error	t	p
Mapping	0.039	0.029	1.356	.175

Note. Random effects included in the final model that converged: by-participant intercepts.

4.6 Discussion

The current study explored the contribution of implicit and explicit processes to the learning of novel grammatical mappings that were either related or unrelated to learners' existing knowledge, across a four-day training and consolidation period and using a language that matched grammatical regularities in natural languages more closely. The learning of the grammatical mappings over the four sessions was assessed using the same Picture Naming task as in Experiments 1-3. Here, the key questions were whether participants' recall of the novel words (stems and grammatical mappings) improved across training days and whether recall was greater for prior knowledge dependent grammatical mappings, as previously observed in Experiments 1, 2, and 3. We also assessed changes in recall of the stems and mappings across offline consolidation intervals as the training progressed.

To examine automatic, implicit processes in naturalistic language use and specifically the use of determiners as predictive cues, an eye-tracking task adapted from a study by Lew-Williams and Fernald (2010) was employed. The primary aim of this task was to assess whether participants could use the gender-marked determiners predictively during online processing of the novel words, and how performance on this task would differ between the two types of mappings: those related to prior knowledge and those unrelated. This task was also used to assess generalisation of the determiner mappings and was administered on the final day of training, in addition to the Word-Picture Matching task which provided a more direct measure of generalisation. Of key interest on both measures was whether participants could generalise the mappings following the four-day training period and whether they would show an advantage for the prior knowledge mappings in the more complex language.

Finally, the contribution of explicit knowledge to learning the determiner and suffix mappings was assessed using the Word-Picture Matching task as in Experiments 2 and 3. Of interest was whether explicit awareness of the mappings emerged over the course of training, and whether awareness differed between the two mapping types. On the final day of training, explicit knowledge was further assessed using a verbal report, again comparing awareness between the two mapping types.

First, we examined participants' word learning and consolidation using the same Picture Naming task used to assess word learning in Experiments 1-3. On the final day of training, the same 2AFC task that was used to assess word learning in Experiments 1-3 was also administered. Specifically, for the Picture Naming task, of interest was whether participants' recall of the stems improved across training days and how offline changes in recall compared across the three consolidation intervals. Although recall accuracy for the stems was low overall, participants improved with each training session, and by the end of training on Day 4 their recall accuracy had reached levels similar to those observed in the previous experiments. Additionally, word knowledge as measured via the 2AFC task on the final day of training was high and reached levels similar to those observed in previous experiments. Alongside word learning, the Picture Naming task was used to assess offline changes in stem recall accuracy across the three consolidation periods, and this revealed little change between training sessions (i.e., during offline intervals).

As with the stems, participants' learning of the determiner and suffix mappings was also assessed using Picture Naming. Of key interest was whether participants would show improvements in their learning of the mappings (as indicated by greater recall accuracy), across training days and whether greater recall would be observed for the mappings in the prior knowledge dependent condition, as previously observed in Experiments 1, 2, and 3. Additionally, we evaluated how recall accuracy for these mappings changed over offline consolidation periods. First, when examining recall across training sessions, we found that recall improved for both determiners and suffixes following each training session, and for the determiners specifically, greater recall was observed for prior knowledge dependent mappings compared to prior knowledge independent mappings, consistent with our previous findings. However, this advantage did not emerge until Day 3 of training. Moreover, no prior knowledge benefit was observed for the suffix mappings in the current paradigm. These results differ from Experiments 1, 2, and 3, where a prior knowledge benefit for both the determiners and suffixes was evident immediately, after just one training session.

The different findings between Experiments 1-3 and Experiment 4 may be attributed to the type of training needed for the abstraction of the grammatical regularities. For example, participants in Experiments 1 and 2 were trained on 32 words, receiving 8 exposures to each word by the end of a single training session, and participants in Experiment 3 were trained on 48 words and received 10 exposures per word within a single training session. By comparison, in the current experiment, during training on Day 1, participants would have only had 4 exposures to each of the 64 novel words (2 exposures on the eye-tracking training task and 2 on the Word-Picture training matching task). Furthermore, the Word-Picture Matching training task, included mismatched trials, where each novel word was also presented twice with an incorrect referent (once with a referent from the same semantic category and once from the opposite category). However, in the Word Repetition task which was used to train participants in our previous experiments did not include mismatched trials. Moreover, this task, unlike the Word-Picture Matching training used in the current study, required participants to repeat back the novel words thus providing additional exposure. This inconsistency during training combined with the relatively infrequent exposure to the novel words (in contrast to our previous experiments), may have made it difficult for participants to abstract the underlying regularities shared across the individual exemplars (e.g., learning that *tib* frequently co-occurs with pictures of female characters), which is undoubtedly critical if participants are going to benefit from their connections to existing grammatical knowledge.

Moreover, for the suffixes in particular, for which no prior knowledge advantage was observed, more exposure was likely needed to support the process of abstraction, as unlike the determiners (of which there were only two: *tib* and *ked*), participants had to learn twice as many suffixes per semantic category in the current study (e.g., novel words assigned to pictures of female characters ended either in *-eem* or *-esh*, and those assigned to male characters ended either in *-aff* or *-ool*).

In addition to evaluating changes in learning, the Picture Naming task was also used to examine offline changes in recall accuracy for the grammatical mappings, comparing performance across the three consolidation intervals. For both the determiners and suffixes, participants exhibited a decline in their recall accuracy during each consolidation interval and this was the case for both types of determiner and suffix mappings: those with the novel words in the prior knowledge dependent condition and those with the novel words in the prior knowledge independent condition. Upon comparing changes in performance across the consolidation intervals, for determiners, this decline was more pronounced during the second consolidation interval (between Day 2 and Day 3) compared to the third interval (between Day 3 and Day 4). For suffixes, the rate of decline did not differ across the three intervals. Overall, the general decline in recall accuracy across offline consolidation intervals for both the prior knowledge dependent and independent grammatical mappings contrasts with the findings for the stems, which showed little change in recall across the same periods. Although the current experiment did not specifically test the influence of sleep-related consolidation, this pattern, maintenance of the more arbitrary stem–meaning mappings versus the decline of the more systematic determiner and suffix mappings, is broadly consistent with the findings from Experiment 1. In Experiment 1, recall accuracy of the stems was maintained following a 12-hour consolidation period that included sleep, whereas accuracy for the determiner and suffix mappings declined over the same interval. Taken together, the findings from

Experiment 4, like those from Experiment 1, are consistent with and provide further support for the interpretation of the findings by Mirković and Gaskell (2016): that when language learners encounter novel words that combine arbitrary and systematic properties, offline consolidation may prioritise the more hippocampally dependent arbitrary mappings.

Moreover, the findings from the Picture Naming task in the current experiment are also broadly consistent with our previous findings regarding the role of prior knowledge in the initial learning of the systematic mappings. Specifically, the prior knowledge benefit observed for recall of the determiners in the current study replicates the findings of Experiments 1, 2, and 3, further supporting the theoretical predictions of the Complementary Learning Systems (CLS; McClelland, 2013) and Schema-Linked Interactions between Medial prefrontal and Medial temporal regions (SLIMM; Van Kesteren et al., 2012). These models predict that prior knowledge related information should benefit from its connections to existing knowledge during learning, resulting in faster consolidation into long-term memory than when new information is unrelated.

In addition to the above, the current experiment examined implicit learning and generalisation of the determiner mappings using a new eye-tracking task. Following Lew-Williams and Fernald (2010), it was reasoned that if participants had learned the determiner mappings, they would be able to use them predictively during online processing of the novel words, orienting their gaze more quickly toward the correct referent on different-gender trials, where the determiner provided a predictive cue, compared to same-gender trials, where it did not. If participants were able to use the gender-marked determiners predictively, of key interest was whether they would show an advantage for the determiners in the prior knowledge dependent condition (e.g., would they be quicker to orient their gaze to the target referent on different gender trials when the novel word belonged to the prior knowledge dependent condition). However, for the eye-tracking task with the trained items, only a small number of trials met the inclusion criteria on each day; therefore, only Day 4 was included in the analysis, as it had the highest proportion of trials meeting the criteria, although these were still relatively low (e.g., prior knowledge dependent: 29% different-gender, 25% same-gender; prior knowledge independent: 26% different-gender, 24% same-gender). Similarly, the proportion of trials included for the eye-tracking task with the generalisation items on Day 4 was also low (e.g., prior knowledge dependent: 33% different-gender, 28% same-gender; prior knowledge independent: 39% different-gender, 29% same-gender).

By the end of training on Day 4 participants showed no evidence of being able to use the gender-marked determiners predictively, as no differences in reaction times were observed between the same-gender and different-gender trials. This was the case for both training (i.e., when the determiner mappings were presented in their trained item contexts) and generalisation (when the determiner mappings were presented in novel, untrained contexts). Nonetheless, the lack of evidence for predictive use of the determiners in this task should be taken cautiously given that, as outlined above, only a small proportion of trials met the inclusion criteria for analysis. This is likely because we applied the same inclusion criteria as Lew-Williams and Fernald (2010), who used the task to examine predictive processing of gender-marked determiners in a natural language, with participants who were already of high to moderate proficiency. By contrast, participants in the current experiment were trained on a large and complex artificial language for which they had only received four days of training. Thus, the Lew-Williams and Fernald (2010) inclusion criteria may have been too stringent for the participants in our study, which may explain why we ended up with only a small number of trials for analysis and hence why no differences in reaction times were observed between the same-gender and different-gender trials.

Furthermore, the proficiency of the participants in our experiment may also explain why we found no evidence that they were able to use the gender-marked determiners predictively.

For example, Lew-Williams and Fernald (2010) showed that while L1 speakers of Spanish could use gender-marked determiners predictively during online processing, L2 speakers who were only moderately proficient in Spanish were less successful. They managed to do so only when the experimenters controlled for frequency of exposure by training and testing both groups on a small set of novel nouns and objects (just four items), rather than real Spanish nouns and objects. Only L1 speakers were able to use the determiners predictively when both groups were tested using real Spanish nouns and objects and only L1 speakers were able to generalise when the novel nouns were presented with different determiners at test (e.g., *un/una* at training, *el/la* at test). Thus, given that the L2 speakers in Lew-Williams and Fernald's (2010) study, who were moderately proficient in Spanish, managed to use the determiners predictively only when trained on a very small set of 4 novel nouns and objects, it is perhaps not surprising that participants in the current study, who were trained and tested on 64 novel words across only four days, were unable to take advantage of the gender-marked determiners to support online processing.

Indeed, previous research has shown that learners' ability to use gender marking predictively during online processing critically depends on the stability of their lexical gender representations. For example, in a visual world eye-tracking paradigm, similar to the one used in the current study, Hopp (2013) tested whether L1 and L2 speakers of German were faster in shifting their gaze to a target object on trials where gender-marked determiners provided an informative cue regarding the upcoming noun. The results revealed that only the L2 speakers who consistently assigned the correct gender to almost all nouns in a separate production task were able to use the gender-marked articles predictively during online processing. By contrast, L2 speakers who showed inconsistencies in their assignment of gender in the production task were less likely to use gender marking on determiners to support online processing. Based on these findings, Hopp (2013) suggested that language learners must have stable, accurate and consistent knowledge about the grammatical gender of nouns in order to be able to use gender marking to make predictions about them during online comprehension. Those with less stable or less accurate gender representations will perceive gender-marked determiners as less reliable and therefore will be less likely to find them informative during online processing. This is broadly in line with other research which has shown that language proficiency is a critical factor in determining learners' ability to use gender marking predictively during online processing (e.g., Dussias et al., 2013). Taking these findings into consideration, it is possible that, due to the large number of novel words and relatively little exposure across training days, participants in the current study only formed weak or inconsistent lexical gender representations which could not be relied upon to support processing in real time.

In addition to assessing generalisation of the determiner mappings in the eye-tracking task, we also assessed participants' ability to generalise the determiner and suffix mappings using the same Word-Picture Matching task which was used to assess generalisation in Experiments 2 and 3. This task was administered on the final day of training (Day 4) and provided an additional, more direct measure of generalisation. Unlike the more implicit measure provided by the eye-tracking task, the Word-Picture Matching task revealed evidence of generalisation. However, overall performance was weak, and participants showed no advantage in generalising mappings from the prior knowledge dependent condition relative to the prior knowledge independent condition. Taken together with the results from the Picture Naming task, which revealed a prior knowledge benefit for determiner mappings following training, these findings align with Experiments 2 and 3, demonstrating that while prior knowledge can support the initial learning of the grammatical mappings, it does not facilitate their generalisation to novel contexts.

Alongside evaluating learning and consolidation, another key aim of the current study was to assess participants' explicit knowledge of the grammatical mappings. To test this, the same two tasks used previously in Experiments 2 and 3, the Word-Picture Matching task and a retrospective verbal report, were employed. First, the Word-Picture Matching task was administered on each of the four training days and assessed participants' explicit knowledge of the grammatical mappings when presented within their trained item contexts. Of interest was whether participants would develop explicit knowledge of these mappings across training days, and whether explicit knowledge differed between the two mapping types. The results of the analysis for this task revealed evidence of a role for explicit knowledge which emerged on Day 2 of training. However, participants did not demonstrate greater explicit awareness for the prior knowledge dependent mappings in this task. On the final day of training, we also examined explicit knowledge in the Word-Picture Matching task, when the mappings were presented in untrained, novel contexts. Unlike with the trained items, participants' accuracy was not related to their confidence in their Word-Picture Matching judgments for the novel items, suggesting an absence of explicit knowledge for the mappings when presented in novel contexts.

Additionally, self-reported explicit knowledge was also tested on the final day of training via the retrospective verbal report measure. In contrast to the findings for the Word-Picture Matching task, where explicit knowledge did not differ between the two mapping types, self-reported knowledge was greater for the prior knowledge dependent determiner mappings relative to the prior knowledge independent determiner mappings. However, awareness was low overall and no differences between the mapping types were observed for the suffixes. Overall, the findings observed for both the Word-Picture Matching and self-report measure are consistent with the findings observed for these measures in Experiment 2 and 3 (and also Experiment 1, in the case of the self-reported explicit knowledge) in that greater awareness for the prior knowledge related (determiner) mappings was only found in the self-report measure. While evidence of explicit knowledge was observed in the Word-Picture Matching task when the mappings were presented in their trained item contexts, this knowledge was not greater for the prior knowledge dependent mappings. Nonetheless, the greater levels of explicit knowledge for the prior knowledge dependent determiner mappings observed in the retrospective self-report should be interpreted cautiously given the very low levels of knowledge overall and the fact that the measure was administered only at the very end of the experiment (final task on Day 4).

Finally, it should be acknowledged that some of the findings discussed above must be considered in light of several limitations. For instance, no a priori power analysis was conducted for Experiment 4, and the small number of participants (N = 36), together with the large number of excluded trials in some tasks, suggest that some of the findings outlined above should be interpreted with caution. This is particularly relevant to the results of the eye-tracking task, for which only a small number of trials satisfied the inclusion criteria for analysis, potentially reducing statistical power and possibly contributing to the null effects observed.

One other limitation to be considered is that the methodological changes introduced in Experiment 4 limit the extent to which the current findings, particularly those from the new eye-tracking task can inform and extend the interpretation of the findings observed in Experiments 1–3. For instance, in the current experiment, the artificial language was made both larger and more complex than in Experiments 1–3. Specifically, participants were trained on 64 words, compared to 32 words in Experiments 1 and 2 and 48 words in Experiment 3. In addition, the semantic categories associated with the prior knowledge independent mappings were changed to accommodate the larger training set, resulting in a more wide-ranging set of categories than in the earlier experiments. Whereas these mappings had previously involved insects and sea animals, in Experiment 4 *tib* words were paired with domesticated animals and insects, and *ked* words were paired with wild animals and reptiles/amphibians. The suffix mappings were also made more complex with each determiner combining with two possible suffixes rather than one, as was the case in the previous experiments. These changes were introduced to make the training language more closely resemble the complexity of natural languages and thereby allow us to test whether the previously observed prior knowledge benefit for recalling the determiner and suffix mappings would replicate in a more complex language.

In addition to the changes made to the complexity of the artificial language, some key changes were also made to the training procedure. In Experiments 1-3, participants' training involved a word repetition task, however, in the current experiment, the eye-tracking and Word-Picture Matching tasks provided training on the artificial language, and these training tasks were administered across four consecutive days rather than within a single session. The key motivation for this change was that these tasks, unlike the word repetition task, allowed us to track the possible emergence of implicit and explicit knowledge over the course of training. However, as outlined above, because the current experiment made several changes to the artificial language and training procedure, the findings observed, particularly for the new eye-tracking measure are less able to build directly on the findings from Experiment 1-3.

In summary, the experiment presented in this chapter explored how implicit and explicit processes contribute to learning and generalising new grammatical mappings, which were either related or unrelated to participants' prior knowledge using a more complex language and over a longer training period. The results from the Picture Naming task indicated participants learned the determiner and suffix mappings over the four-day training period and to a similar degree as in Experiments 1-3. However, the eye-tracking task, which was introduced to provide a more implicit measure of learning and generalisation of the determiner mappings revealed that, after four training sessions, participants were unable to use the determiners predictively during online processing. Moreover, although we did replicate the prior knowledge benefit for the determiner mappings in the Picture Naming task (i.e., at recall), where weak generalisation was observed (in the more direct Word-Picture Matching task), prior knowledge did not appear to support greater generalisation of related mappings. This is consistent with what was observed in Experiments 1, 2 and 3. Finally, while participants did demonstrate evidence of emergent explicit knowledge for the mappings when presented in their trained item contexts, in line with Experiments 2 and 3, explicit knowledge did not appear to play a significant role in the learning of the prior knowledge related mappings.

Chapter 5. General Discussion

This section summarises the key findings and conclusions from the experiments described in Chapters 1–4, which examined the role of prior knowledge in the initial learning and consolidation of novel grammatical regularities.

5.1 Summary of Key Findings

5.1.1 The Role of Offline Consolidation

The primary aim of Experiment 1 was to investigate the contribution of offline processes during sleep compared to wakefulness in consolidating systematic types of linguistic knowledge (i.e., grammatical mappings). This aim was motivated by previous findings. For example, while a significant body of evidence has demonstrated a clear role for sleep in the consolidation of arbitrary knowledge (i.e., novel words; Schimke et al., 2021), research examining the role of offline processes in the consolidation of more systematic types of knowledge has produced mixed results. For instance, a study by Mirković and Gaskell (2016), which trained participants on a fully fledged artificial language that comprised both arbitrary and systematic mappings, observed a beneficial effect of sleep on the consolidation of arbitrary mappings but found no such effect in the generalisation of systematic mappings. In contrast, research by Tamminen et al. (2012; 2015), which examined the contribution of offline processes to systematic knowledge in isolation (i.e., not in parallel to arbitrary/word knowledge), observed improvements in the generalisation of systematic mappings following an offline consolidation period.

Mirković and Gaskell (2016) suggested that the absence of a role for sleep in the consolidation of the systematic mappings in their study was likely because arbitrary and systematic knowledge competed for consolidation, with arbitrary mappings being prioritised due to their greater reliance on the hippocampus at encoding. However, in Experiment 1, two alternative explanations for these conflicting findings were considered: the influence of prior linguistic knowledge and the tests used to assess generalisation of the systematic mappings. Firstly, the systematic mappings in Mirković and Gaskell's (2016) study closely aligned with participants' existing grammatical knowledge. According to the Complementary Learning Systems model (CLS; McClelland, 2013) and the Schema-Linked Interactions between Medial prefrontal and Medial temporal regions framework (SLIMM; Van Kesteren et al., 2012), when new knowledge is related to existing knowledge, it can be integrated more efficiently into neocortical networks. This results in comparatively limited hippocampal involvement during encoding, potentially reducing the need for hippocampal reactivation to occur offline (e.g., during post-learning sleep), compared to when information is unrelated to or incongruent with prior knowledge. Thus, from this perspective, it was considered that the absence of a sleep-related consolidation benefit for the systematic mappings in the study by Mirković and Gaskell (2016) could be explained by their relatedness to existing knowledge.

It was also considered that the conflicting findings in the studies outlined above could be the result of differences in task sensitivity to sleep-related consolidation effects. For example, Mirković and Gaskell (2016) used a non-speeded task to assess the generalisation of systematic mappings. In contrast, Tamminen et al. (2012, 2015), who employed both speeded and non-speeded measures, found evidence of a role for offline consolidation in generalisation only when using a speeded task. In the non-speeded task, participants were able to generalise immediately after learning, and their performance was not further influenced by offline consolidation. Therefore, the absence of a sleep-related consolidation benefit for the systematic mappings in Mirković and Gaskell (2016) may also have been explained by the fact that a non-speeded generalisation test was used.

Using an artificial language similar to that used by Mirković and Gaskell (2016), which incorporated both arbitrary and systematic mappings, Experiment 1 aimed to address these conflicting findings and clarify the role of offline consolidation in grammar learning. Specifically, we used a combination of non-speeded and speeded tests to examine the contribution of sleep versus wake in the consolidation of two types of systematic mappings: those related to learners' existing grammatical knowledge and those that were unrelated. In addition to the systematic mappings, we also examined the contribution of sleep versus wake in the consolidation of the arbitrary mappings.

First, to examine how offline consolidation processes during sleep versus wakefulness might contribute to the consolidation of the prior knowledge dependent and independent systematic mappings, recall accuracy in the non-speeded Picture Naming task and response times in the speeded shadowing task were compared immediately after training, following a 12-hour consolidation period involving sleep or wake, and again after a 1-week delay. First, in the Picture Naming task, for both the determiners and suffixes, we observed a decline in recall accuracy after the initial 12-hour consolidation interval, followed by a further decline after one week. This pattern was evident for mappings in both the prior knowledge dependent and prior knowledge independent conditions and occurred regardless of whether participants slept or remained awake during the 12-hour interval. Additionally, in the speeded shadowing, we observed an overall increase in response times (i.e., participants were slower to repeat the novel words) after the 12-hour consolidation interval, followed by a decrease in response times (i.e., participants were faster to repeat the words) across the subsequent week. While the degree of slowing, after the 12-hour interval and speeding up after 1 week, appeared to vary by group and mapping type, follow-up analyses indicated that these differences were likely driven by baseline differences in performance between the two groups at Session 1. Furthermore, although we only had the data available for Sessions 1 and 2, when the mappings were presented with novel, untrained stems (speeded shadowing generalisation), no differences in response times were observed between the two groups or mapping types after the 12-hour interval. Thus, using both non-speeded and speeded tasks, we found no evidence that sleep influenced the consolidation of the systematic mappings, regardless of their relationship to participants' existing knowledge.

In contrast to the systematic mappings, evidence for a role of sleep was observed in the consolidation of the arbitrary mappings (i.e., stem–meaning mappings). Specifically, in the Picture Naming cued recall task, participants in the sleep group maintained their knowledge of the stems after a 12-hour interval, whereas those in the wake group showed a decline over the same period. Although both groups exhibited a decrease in their recall accuracy of the stems after 1 week, the maintenance in word recall over 12 hours of sleep relative to wake contrasts with the overall decline in recall of the systematic mappings across the same interval. Thus, the findings from the Picture Naming task in Experiment 1 highlight a clear dissociation in the role of sleep in the consolidation of the arbitrary versus systematic mappings. Moreover, although we did not specifically test the influence of sleep-related consolidation in Experiment 4, a similar dissociation was also observed here in that recall accuracy for the stems remained relatively stable across offline consolidation periods, whereas accuracy for the more systematic determiner and suffix mappings declined. Overall, these findings extend those of Mirković and Gaskell (2016) by showing that, even when systematic knowledge is unrelated to learners' existing knowledge, the arbitrary components are still prioritised during offline consolidation.

5.1.2 How Does Prior Knowledge Influence the Initial Learning of the Systematic Mappings?

Experiment 1 was the first study, to our knowledge, to explore the role of prior knowledge in grammar learning. Therefore, in addition to examining its influence on the consolidation of the mappings, a further key aim was to investigate how prior knowledge affected their initial learning. For instance, experimental evidence from the word learning literature has shown that, while prior knowledge can support the initial acquisition of novel word-forms (James et al., 2019), related semantic knowledge can interfere with the learning of novel words and their meanings (James et al., 2023; Tamminen et al., 2013). Thus, Experiment 1 also sought to determine whether existing grammatical knowledge would support or hinder the initial encoding of the related mappings.

Upon assessing learning of the systematic mappings immediately after training, we found evidence of a role for prior knowledge in supporting recall of related mappings. Specifically, in the non-speeded Picture Naming task, participants in both groups showed higher recall accuracy for the systematic mappings (determiners and suffixes) that were related to their existing grammatical knowledge compared to those mappings that were unrelated. However, this benefit did not extend beyond recall. For example, in the speeded shadowing task, which required participants to repeat the novel words as quickly and accurately as possible, participants were no faster at repeating the novel words with the mappings in the prior knowledge dependent condition compared to the words with the mappings in the independent condition for either group when tested immediately after training. This was the case for both speeded shadowing training (i.e., the version of the task where the mappings were presented with the original word stems encountered during training) and the speeded shadowing generalisation (when the mappings were presented with new, untrained stems).

Based on these findings, it was hypothesised that explicit knowledge may have contributed to the prior knowledge benefit observed in the Picture Naming task. For instance, a key distinction between the non-speeded Picture Naming task and the speeded shadowing task lies in the retrieval processes they engage, with the speeded tasks limiting opportunities for reflection and, consequently, the engagement of explicit memory processes. By contrast, the non-speeded nature of the Picture Naming task likely provided participants with the opportunity to reflect on and recall connections between their existing knowledge and new, related knowledge. Therefore, in Experiments 2 and 3, we further focused on the influence of prior knowledge on the initial encoding of grammatical mappings, aiming to build on the findings of Experiment 1 by replicating the prior knowledge benefit at cued recall and examining the potential contribution of explicit knowledge to this effect. In Experiment 4, these same aims were addressed but within the context of an artificial language that more closely simulated the complexity of natural languages. Moreover, in Experiment 4, recall accuracy and explicit knowledge were assessed across four days, whereas in Experiments 2 and 3, all training and testing took place within a single session.

First, the prior knowledge benefit observed in the Picture Naming cued recall task in Experiment 1 was replicated in Experiments 2 and 3 for both the determiner and suffix mappings. However, in Experiment 4, this benefit at recall was observed only for the determiners and did not emerge until the third day of training. Unlike in Experiments 1–3, participants had minimal exposure to the novel words on each training day and a larger set of words to learn. Moreover, the artificial language in Experiment 4 contained twice as many suffixes as determiners and twice as many suffixes as the language used in Experiments 1–3. Thus, it was considered that participants in Experiment 4, particularly early on in training, may not have learned enough words to abstract the underlying regularities, and therefore could not yet benefit from their connections to existing grammatical knowledge. Taken together, while the findings from Experiments 1–3 align with the theoretical predictions of the CLS (McClelland, 2013) and SLIMM (Van Kesteren et al., 2012) models of memory consolidation in showing that learners can benefit from their connections to existing knowledge at encoding, the findings from Experiment 4 indicate that, in a paradigm more consistent with real-world language learning, this benefit is unlikely to emerge straight away. Instead, language learners may first need to acquire a sufficient number of words in order to abstract the mappings before they can benefit from their connections to existing grammatical knowledge.

5.1.3 Types of Knowledge Underlying the Prior Knowledge Benefit.

As outlined above, in addition to replicating the prior knowledge benefit at cued recall, Experiments 2 and 3 also aimed to build on the findings of Experiment 1 by exploring the possibility of an explicit knowledge contribution to the observed prior knowledge benefit. This was primarily motivated by the fact that, in Experiment 1, evidence of a role for prior knowledge in the initial learning of the grammatical mappings was only observed in a task permitting explicit retrieval (the non-speeded Picture Naming task). To investigate the role of explicit knowledge, two measures were employed: a retrospective verbal report that was also administered at the end of the final session in Experiment 1, and a Word-Picture Matching task with an integrated confidence rating measure. This task was introduced to provide a more sensitive, within-task measure of explicit knowledge, as verbal reports alone may not capture the full range of explicit knowledge (e.g., Rebuschat, 2013). Both the verbal report measure and the Word-Picture Matching task were also used to assess participants' explicit knowledge of the mappings in Experiment 4.

Across Experiments 2, 3 and 4, although participants demonstrated some explicit knowledge of the underlying grammatical mappings in both tasks, contrary to what was hypothesised in Experiment 1, it did not appear to play a significant role in the learning of the prior knowledge related mappings. For example, while participants' self-reported explicit knowledge (as measured by the retrospective verbal report) was greater for the prior knowledge dependent determiner mappings, explicit knowledge was low overall, and no differences were observed between the two mapping types for the suffixes. This pattern of results was observed across Experiments 1-4. Additionally, although we found evidence of a role for explicit knowledge in Experiments 2, 3, and 4 in the Word-Picture Matching task when the mappings were presented in their trained-item contexts, this knowledge was no greater for the prior knowledge dependent mappings than for the prior knowledge independent mappings. Thus, when considered together, these findings challenge the interpretation of the findings from Experiment 1, where it was considered that the presence of a prior knowledge benefit in the non-speeded Picture Naming task, but not in the speeded tasks, was because the non-speeded task afforded participants the opportunity to recall connections between their existing grammatical knowledge and new, related knowledge.

Nonetheless, future research using more sensitive measures would help to clarify these findings. For instance, in the Word-Picture Matching task used in the experiments outlined above, participants' explicit knowledge about the underlying grammatical mappings was inferred from their confidence ratings when making word-picture matching judgements (e.g., "I am certain / I think / I guess they go well together" and "I am certain / I think / I guess they do not go well together"). However, although confidence ratings are considered a more sensitive measure of explicit knowledge than verbal reports, they still allow participants to set their own criteria for reporting knowledge and therefore do not fully overcome the problem of response bias (Dienes, 2007; 2008). For example, more conservative respondents may only indicate high confidence when they are completely certain, while more liberal respondents may express high levels of confidence at the slightest intuition (Rebuschat, 2013). Taking this into consideration, a more sensitive measure of explicit knowledge may be required to tease apart the possible differences in explicit awareness between the prior knowledge related and prior knowledge unrelated mappings in the current paradigm. One alternative to confidence rating measures would be to use EEG. Indeed, in a study by Batterink et al. (2015), EEG was one measure used to examine the contribution of explicit knowledge to grammar learning in adults. Specifically, the Late Positive Component (LPC), an ERP typically associated with the recollection of previously learned information, was used as an index of explicit memory retrieval. Thus, incorporating a similar EEG measure into the current paradigm could provide a more sensitive index of explicit knowledge, free from the influence of response bias.

In addition to examining explicit knowledge contributions, Experiment 4 also explored the role of implicit knowledge. As a first attempt to investigate this, a new eye-tracking measure was introduced to examine the contribution of automatic, implicit processes to the learning of the systematic mappings. Of key interest was whether participants could use the determiners predictively during online processing, and, if so, whether we would observe an advantage for the prior knowledge dependent mappings in this task. The procedure was adapted from Lew-Williams and Fernald (2010), who examined the predictive processing of gender-marked articles/determiners in a natural language, and we analysed our data using the same approach. However, because their exclusionary criteria were developed for a natural language paradigm, applying them to the data from our artificial language paradigm meant that relatively few trials qualified for analysis. Thus, perhaps unsurprisingly, we found no evidence that participants were able to use the gender-marked determiners predictively in either mapping condition.

5.1.4 Does prior knowledge influence the generalisation of systematic mappings?

Alongside exploring the possibility of an explicit knowledge contribution to the learning of the prior knowledge related mappings, in Experiments 2 and 3, we also set out to examine generalisation of the mappings using a non-speeded test to determine whether the observed prior knowledge benefit extends beyond item-specific recall to support the generalisation of grammatical knowledge to novel, untrained items. In Experiment 1, despite observing a prior knowledge benefit for the trained item mappings at cued recall, participants did not show greater generalisation for these mappings when tested under speeded conditions. However, we considered that if participants were relying on explicit knowledge to support their learning of the prior knowledge dependent mappings, a prior knowledge benefit at generalisation would be more likely to emerge in a non-speeded task, as this would afford participants the opportunity to consciously retrieve and apply their knowledge to support performance. To test this, the same Word-Picture Matching used to assess explicit knowledge of the mappings was administered. In both Experiments 2 and 3, although we found evidence of generalisation, it was weak overall, and participants were no better at generalising the prior knowledge dependent mappings than the independent ones. This same pattern of results was also observed for this task on the fourth and final day of training in Experiment 4. Thus, as previously discussed, it seems unlikely that differences in the retrieval conditions between the non-speeded Picture Naming task and the speeded shadowing task in Experiment 1 can account for the absence of a prior knowledge benefit in generalisation.

Instead, in Chapter 3, we considered the possibility that the representations of the grammatical mappings formed during training were not abstract enough to support their generalisation to novel contexts. For instance, abstractionist theories of category learning posit that, in order for learners to be able to generalise their knowledge to novel situations, they must first form an integrated summary or average of the individual category members (i.e., the trained exemplars; Posner & Keele, 1968). However, this process is dependent on the degree of within-category similarity or coherence of the trained exemplars. This is because when category members have high perceptual coherence at training, it is easier for learners to extract the central tendency of the category and form a prototype representation. This form of generalisation is likely to underlie performance in speeded contexts.

For example, Tamminen et al. (2015) showed that participants' ability to generalise novel affixes in speeded tasks was influenced by factors that enhanced the discovery of shared features across a number of different exemplars of a given category, and one such factor was the semantic consistency of the novel affixes during training. That is, when participants were trained on two types of novel affixes: semantically consistent and semantically inconsistent, only the affixes in the semantically consistent condition supported generalisation in speeded contexts. Based on these findings, we considered that the absence of a prior knowledge benefit in the speeded shadowing task in Experiment 1 was likely due to the fact that the training language consisted of grammatical mappings that were semantically inconsistent. For example, in the training language used in Experiments 1-3, the determiners and suffixes each predicted two distinct semantic categories, with *tib* and *eem* referring to both female characters and insects and *ked* and *ool* referring to both male characters and sea animals. This inconsistency likely made it difficult for participants to form an integrated, summary representation (i.e., a prototype representation) of the prior knowledge dependent and independent mappings at training, which they could then use to support generalisation in the speeded task.

Moreover, it was also considered that the semantic inconsistency of the determiner and suffix mappings during training could account for the weak generalisation and lack of a prior knowledge benefit in the non-speeded task in Experiments 2 and 3. For instance, in the study by Tamminen et al. (2015), although semantic inconsistency prevented generalisation in speeded contexts, it also diminished performance in a non-speeded explicit congruency judgement task. Thus, semantic inconsistency at training may have reduced participants' generalisation of both mapping types in Experiments 2 and 3 and in doing so, reduced our ability to observe any differences between the two conditions.

5.2 Limitations

The findings presented across the empirical chapters should also be considered in light of several limitations. The first concerns the potential weakness of the prior knowledge manipulation of the systematic mappings. In the paradigm used across this thesis, the systematic mappings were designed to be consistent with participants' existing grammatical knowledge such that the semantic category assignment of male and female was encoded in the phonological properties of the novel words (i.e. the determiners and suffixes), the same way English encodes biological sex in the pronouns she/her and he/him. However, this manipulation is not fully consistent with grammar in English, as such distinctions are not marked systematically across determiners or noun endings. In addition, it ought to be considered that the category assignment of male and female may be more salient to participants than semantic categories within animal types (sea animals vs. insects). Thus, the prior knowledge benefit observed across the experiments presented in this thesis could also reflect the fact that the related mappings were associated with more salient semantic categories for participants. Future research could address this by using the same paradigm in speakers of languages with a more fully grammaticalised gender system where gender is marked systematically on determiners and noun endings, and by using semantic categories that are more closely matched in salience across the prior knowledge dependent and independent conditions.

A second limitation relevant to the interpretation of the findings observed in Experiments 1 and 4 is the fact that no a priori power analysis was conducted. This should be taken into account when interpreting some of the null effects observed in these experiments, as the studies may not have been sufficiently powered to detect small or even moderate effects. This issue is particularly relevant in Experiment 4, where the small sample size and large number of excluded trials in the eye-tracking task may have contributed to the null effects observed.

One final point to consider concerns the limitations associated with running complex experiments online. Experiments 1 and 3 were conducted remotely, which meant that there was less control over the environments in which participants completed the sessions and, in Experiment 1, no direct control over participants' activities during the 12-hour sleep and wake consolidation intervals. Although it should be noted that some checks were implemented. For example, engagement with training was checked by listening to the audio recordings of participants' responses in the training task. Similarly, in Experiment 1, at the start of their second session, participants were asked questions about their activities during the 12-hour consolidation interval (e.g., the wake group were asked whether they napped). In addition, at the end of both Experiments 1 and 3, participants were asked if they had written any of the novel words down during training and if they had avoided completing any tasks. Nonetheless, these checks mostly relied on participants' honesty. As such, some reduction in experimental control remained unavoidable in the online studies.

5.3 Conclusions

Overall, across the four experiments outlined in this thesis, we observed a role for prior knowledge in supporting the initial learning of the systematic mappings. These findings are consistent with the theoretical predictions outlined in the CLS (McClelland, 2013) and SLIMM (Van Kesteren et al., 2012) models, which propose that existing schematic knowledge should facilitate the neocortical integration of related information at encoding, resulting in faster consolidation into long-term memory networks.

In addition, the findings presented in this thesis further support those of Mirković & Gaskell (2016) by showing that, when language learners encounter a fully-fledged artificial language comprising both arbitrary and systematic mappings, unlike the arbitrary components, the systematic aspects of the language do not benefit from offline consolidation. Importantly, we also extend this work by showing that this was true even for systematic knowledge unrelated to existing knowledge, precisely the type of information that the CLS (McClelland, 2013) and SLIMM (Van Kesteren et al., 2012) models predict should benefit most from offline consolidation. According to these models, when new knowledge is unrelated to or incongruent with existing knowledge, it is more likely to be dependent on the hippocampus at encoding and, as a result, subject to stronger offline consolidation (i.e., hippocampal reactivation during sleep) than prior knowledge related information. Thus, if the hippocampus is involved in the learning of systematic knowledge, we would expect to see a role for sleep in the consolidation of the systematic mappings without connections to existing knowledge. However, this is not what we found.

One possible interpretation of these findings is that the hippocampus is generally less involved in the learning of systematic knowledge, particularly in a fully-fledged artificial language that incorporates both arbitrary and systematic mappings, which may compete for consolidation (Mirković & Gaskell, 2016). This perspective is consistent with more recent proposals by Gaskell (2024), which advocate a more nuanced view of hippocampal and neocortical contributions to word learning, moving away from the strictly dichotomous framework proposed in the original Davis and Gaskell (2009) model. In the original model's characterisation of word learning, the hippocampus supported the encoding of novel words, while the neocortex provided a robust means of retaining and processing well-established words. Gaskell (2024), however, moves away from this hard division, suggesting that both the hippocampus and the neocortex can contribute to the learning of new words, but the degree of involvement of each depends on the properties of the word. Specifically, arbitrary properties should have the strongest hippocampal representation at encoding and, therefore, be more likely to benefit from hippocampal reactivation during sleep. By contrast, for systematic properties, the bulk of learning will be cortical, making these components less likely to benefit from sleep.

When considered in this light, the findings presented in this thesis raise a more mechanistic question: how might the systematic mappings have been learned via the neocortical route? One possible explanation comes from connectionist models, which have often been used to model the gradual, distributed learning associated with the neocortex (McClelland et al., 1995). For example, connectionist models of inflectional morphology in natural languages have shown that, in the course of learning mappings between meaning and form, the system gradually extracts higher-order regularities such as gender marking and past-tense inflection from semantic and morphophonological regularities across individual words (e.g., Joanisse & Seidenberg, 1999; Mirković et al., 2005). Thus, from a connectionist view, higher-order grammatical regularities can be understood as emergent properties of learned semantic–phonological mappings.

This emergent perspective provides a useful framework for interpreting several aspects of the findings observed across this thesis. First, it may help explain why the prior knowledge advantage was task dependent. If higher-order grammatical regularities emerge through learning relationships between meaning and phonological form, then a benefit for the prior knowledge related systematic mappings should be most evident in tasks that require those semantic-phonological relationships to be retrieved. This may explain why a prior knowledge advantage for the systematic mappings was observed in Picture Naming, where participants had to retrieve the phonological form from a semantic cue, but not in speeded shadowing, where performance placed greater emphasis on reproducing the phonological form.

Moreover, this emergent view may also offer some explanation for the delayed prior knowledge benefit observed for the determiner mappings in Picture Naming in Experiment 4. A key difference between Experiments 1–3 and Experiment 4 was that Experiment 4 involved a larger training set and fewer exposures to each word within a single training session. If, as connectionist models suggest, a neocortical learning system gradually extracts higher-order regularities during the learning of individual semantic-phonological mappings, then the amount of exposure to these mappings is likely to influence how reliably those regularities are extracted. From this perspective, the larger training set and fewer exposures to each word in Experiment 4 may have limited participants' ability to extract the systematic mappings early in training. This may explain why the relationship between the systematic mappings and existing grammatical knowledge did not support performance until the third day.

The same logic may also apply to the weak generalisation observed across the thesis. As outlined above, if a certain level of training or exposure is needed before the systematic mappings can be extracted reliably enough for their connections to prior knowledge to support performance in trained contexts, then further training or exposure may be needed before these mappings can be extracted in a way that is robust enough to support their generalisation to novel contexts.

Appendices

Table A1. Training items used in Experiments 1, 2 and 3.

Word	Picture	Mapping Type
tib vedeem	airhostess	prior knowledge dependent
tib zimbeem	ant	prior knowledge independent
ked craunool	astronaut	prior knowledge dependent
tib shemeem	bee	prior knowledge independent
ked garrool	builder	prior knowledge dependent
ked cassool	butcher	prior knowledge dependent
tib morcheem	caterpillar	prior knowledge independent
tib sarbeem	centipede	prior knowledge independent
ked dowthool	conductor	prior knowledge dependent
ked clarrool	eel	prior knowledge independent
ked goomool	fisherman	prior knowledge dependent
tib scoiffeem	fly	prior knowledge independent
tib smowneem	fortuneteller	prior knowledge dependent
ked geshool	jellyfish	prior knowledge independent

ked vaddool	lobster	prior knowledge independent
ked tormool	mechanic	prior knowledge dependent
tib zeapeem	mermaid	prior knowledge dependent
tib doudeem	mosquito	prior knowledge independent
tib ropeem	moth	prior knowledge independent
tib twalleem	nun	prior knowledge dependent
tib prilleem	nurse	prior knowledge dependent
tib zarfeem	queen	prior knowledge dependent
ked zugool	seahorse	prior knowledge independent
tib regeem	seamstress	prior knowledge dependent
ked shreelool	shark	prior knowledge independent
ked gniddool	soldier	prior knowledge dependent
tib phlaveem	spider	prior knowledge independent
ked treggool	squid	prior knowledge independent
ked vevool	starfish	prior knowledge independent
ked furgool	stingray	prior knowledge independent

ked jorool	vampire	prior knowledge dependent
tib viffeem	waitress	prior knowledge dependent

Table A2. Generalisation items used in Experiments 1, 2 and 3.

Word	Picture	Mapping Type
tib mofeem	fairy	prior knowledge dependent
tib blupeem	cheerleader	prior knowledge dependent
ked zurrool	priest	prior knowledge dependent
ked wilnool	king	prior knowledge dependent
tib heefeem	wasp	prior knowledge independent
tib gatcheem	cockroach	prior knowledge independent
ked roivool	whale	prior knowledge independent
ked plounool	octopus	prior knowledge independent
tib kigeem	ballerina	prior knowledge dependent
tib jerleem	teacher	prior knowledge dependent
ked zomool	boxer	prior knowledge dependent
ked tremool	cowboy	prior knowledge dependent

tib dreteem	ladybird	prior knowledge independent
tib flarneem	beetle	prior knowledge independent
ked migool	fish	prior knowledge independent
ked larshool	dolphin	prior knowledge independent
tib dupeem	witch	prior knowledge dependent
tib gatcheem	bride	prior knowledge dependent
ked snarool	pirate	prior knowledge dependent
ked shillool	firefighter	prior knowledge dependent
tib cherkeem	dragonfly	prior knowledge independent
tib norneem	butterfly	prior knowledge independent
ked jurbool	prawn	prior knowledge independent
ked rukeool	crab	prior knowledge independent

Table A3. Training items used in Experiment 4.

Word	Picture	Mapping Type
tib lerthesh	airhostess	prior knowledge dependent
tib vedeem	angel	prior knowledge dependent
tib shemeem	ant	prior knowledge independent
ked thetaff	astronaut	prior knowledge dependent
tib scoiffesh	ballerina	prior knowledge dependent
ked dowthaff	bear	prior knowledge independent
ked soidool	boxer	prior knowledge dependent
tib viffeem	bride	prior knowledge dependent
ked snarool	builder	prior knowledge dependent
tib darleem	butterfly	prior knowledge independent
tib hormeem	centipede	prior knowledge independent
ked heedfaff	conductor	prior knowledge dependent
tib treggesh	cow	prior knowledge independent
ked migool	cowboy	prior knowledge dependent
ked goitaff	crab	prior knowledge independent
ked shillool	crocodile	prior knowledge independent

ked norkaff	deer	prior knowledge independent
ked narpaff	doctor	prior knowledge dependent
ked senool	dolphin	prior knowledge independent
tib curresh	duck	prior knowledge independent
tib zimbesh	eagle	prior knowledge independent
ked rournool	eel	prior knowledge independent
ked rishaff	elephant	prior knowledge independent
tib sarbesh	fairy	prior knowledge dependent
ked shintaff	firefighter	prior knowledge dependent
tib mofeem	flamingo	prior knowledge independent
ked tormool	footballer	prior knowledge dependent
tib zeapeem	fortuneteller	prior knowledge dependent
ked jorool	frog	prior knowledge independent
ked cherkaff	giraffe	prior knowledge independent
tib thrairesh	goat	prior knowledge independent
ked neentaff	gorilla	prior knowledge independent

tib rutchesh	grasshopper	prior knowledge independent
tib litchesh	gymnast	prior knowledge dependent
tib bisesh	hen	prior knowledge independent
tib chusheem	iceskater	prior knowledge dependent
ked zomaff	king	prior knowledge dependent
ked vruraff	lion	prior knowledge independent
tib gatcheem	maid	prior knowledge dependent
ked larshool	mechanic	prior knowledge dependent
tib jurbesh	mermaid	prior knowledge dependent
tib phlaveem	nun	prior knowledge dependent
tib chebesh	nurse	prior knowledge dependent
tib thileem	parrot	prior knowledge independent
tib doudeem	pig	prior knowledge independent
ked megaff	pilot	prior knowledge dependent
ked sirdool	policeman	prior knowledge dependent
ked lupaff	priest	prior knowledge dependent
tib lekeem	queen	prior knowledge dependent

ked borchool	sailor	prior knowledge dependent
ked roivool	shark	prior knowledge independent
tib norneem	sheep	prior knowledge independent
ked felchool	soldier	prior knowledge dependent
tib vangeem	spider	prior knowledge independent
tib jitesh	swan	prior knowledge independent
tib gortesh	tennisplayer	prior knowledge dependent
ked shegool	tiger	prior knowledge independent
ked cassool	tortoise	prior knowledge independent
ked chowthaff	vampire	prior knowledge dependent
tib hegeem	waitress	prior knowledge dependent
tib garresh	wasp	prior knowledge independent
ked zarfaff	whale	prior knowledge independent
tib nertesh	witch	prior knowledge dependent
ked geechool	wolf	prior knowledge independent

Table A4. Generalisation items used in Experiment 4.

Word	Picture	Mapping Type	Consistency
tib bupeem	canary	prior knowledge independent	consistent
tib tercheem	seamstress	prior knowledge dependent	consistent
tib dretesh	mosquito	prior knowledge independent	consistent
tib plounesh	cheerleader	prior knowledge dependent	consistent
ked furgaff	snake	prior knowledge independent	consistent
ked pessaff	fencer	prior knowledge dependent	consistent
ked shubool	zebra	prior knowledge independent	consistent
ked tarbool	judge	prior knowledge dependent	consistent
ked blorchaff	dove	prior knowledge independent	inconsistent
ked tremaff	librarian	prior knowledge dependent	inconsistent
ked dranool	fly	prior knowledge independent	inconsistent

ked pilmool	teacher	prior knowledge dependent	inconsistent
tib flarneem	camel	prior knowledge independent	inconsistent
tib taseem	butcher	prior knowledge dependent	inconsistent
tib shreelesh	cheetah	prior knowledge independent	inconsistent
tib perrbesh	fisherman	prior knowledge dependent	inconsistent

References

Achermann, P., & Borbély, A. A. (2003). Mathematical models of sleep regulation. *Frontiers in Bioscience, 8,* s683-s693.

Ackermann, S., Hartmann, F., Papassotiropoulos, A., de Quervain, D. J.-F., & Rasch, B. (2015). No associations between interindividual differences in sleep parameters and episodic memory consolidation. *Sleep, 38*(6), 951-959.

Altmann, G. T., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science, 33*(4), 583-609.

Alvarez, P., & Squire, L. R. (1994). Memory consolidation and the medial temporal lobe: A simple network model. *Proceedings of the National Academy of Sciences, 91*(15), 7041-7045.

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods, 52*, 388-407.

Aslin, R. N., & Newport, E. L. (2012). Statistical learning: From acquiring specific items to forming general rules. *Current Directions in Psychological Science, 21*(3), 170-176.

Bakker, A., Kirwan, C. B., Miller, M., & Stark, C. E. (2008). Pattern separation in the human hippocampal CA3 and dentate gyrus. *Science, 319*(5870), 1640-1642.

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.

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.

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

Batterink, L. J., Westerberg, C. E., & Paller, K. A. (2017). Vocabulary learning benefits from REM after slow-wave sleep. *Neurobiology of Learning and Memory, 144*, 102-113.

Berron, D., Schütze, H., Maass, A., Cardenas-Blanco, A., Kuijf, H. J., Kumaran, D., & Düzel, E. (2016). Strong evidence for pattern separation in human dentate gyrus. *Journal of Neuroscience, 36*(29), 7569-7579.

Born, J., & Wilhelm, I. (2012). System consolidation of memory during sleep. *Psychological Research, 76*(2), 192-203.

Bornstein, A. M., & Daw, N. D. (2012). Dissociating hippocampal and striatal contributions to sequential prediction learning. *European Journal of Neuroscience, 35*(7), 1011-1023.

Bowman, C. R., & Zeithamova, D. (2020). Training set coherence and set size effects on concept generalization and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(8), 1442-1464.

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

Brill-Schuetz, K. A., and Morgan-Short, K (2014) The role of procedural memory in adult second language acquisition. *Proceedings of the 36th Annual Conference of the Cognitive Science Society,* 260-265.

Brown, H., Smith, K., Samara, A., & Wonnacott, E. (2022). Semantic cues in language learning: An artificial language study with adult and child learners. *Language, Cognition and Neuroscience, 37*(4), 509-531.

Buysse, D. J., Reynolds III, C. F., Monk, T. H., Berman, S. R., & Kupfer, D. J. (1989). The Pittsburgh Sleep Quality Index: A new instrument for psychiatric practice and research. *Psychiatry Research, 28*(2), 193-213.

Cassidy, K. W., & Kelly, M. H. (1991). Phonological information for grammatical category assignments. *Journal of Memory and Language, 30*(3), 348-369.

Chambers, K. E., Onishi, K. H., & Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experience. *Cognition, 87*(2), B69-B77.

Channon, S., Shanks, D., Johnstone, T., Vakili, K., Chin, J., & Sinclair, E. (2002). Is implicit learning spared in amnesia?: Rule abstraction and item familiarity in artificial grammar learning. *Neuropsychologia, 40*(12), 2185-2197.

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

Cordi, M. J., & Rasch, B. (2021). How robust are sleep-mediated memory benefits? Current Opinion in *Neurobiology, 67*, 1-7.

Damasio, H., Grabowski, T. J., Tranel, D., Hichwa, R. D., & Damasio, A. R. (1996). A neural basis for lexical retrieval. *Nature, 380*(6574), 499-505.

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

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*(4), 803-820.

Dew, I. T., & Cabeza, R. (2011). The porous boundaries between explicit and implicit memory: Behavioral and neural evidence. *Annals of the New York Academy of Sciences, 1224*(1), 174-190.

Diekelmann, S., & Born, J. (2010). The memory function of sleep. *Nature Reviews Neuroscience, 11*(2), 114-126.

Dienes, Z. (2007). Subjective measures of unconscious knowledge. *Progress in Brain Research, 168,* 1-45.

Dienes, Z., & Berry, D. (1997). Implicit learning: Below the subjective threshold. *Psychonomic Bulletin & Review, 4*, 3-23.

Dienes, Z., Altmann, G., Kwan, L., & Goode, A. (1995). Unconscious knowledge of artificial grammars is applied strategically. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(5), 1322-1338.

Dinges, D. F., & Powell, J. W. (1985). Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behavior Research Methods, Instruments, & Computers, 17*(6), 652-655.

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*(1), 35-39.

Dussias, P. E., Kroff, J. R. V., Tamargo, R. E. G., & Gerfen, C. (2013). When gender and looking go hand in hand: Grammatical gender processing in L2 Spanish. *Studies in Second Language Acquisition, 35*(2), 353-387.

Fitneva, S. A., Christiansen, M. H., & Monaghan, P. (2009). From sound to syntax: Phonological constraints on children's lexical categorization of new words. *Journal of Child Language, 36*(5), 967-997.

Franco, A., Cleeremans, A., & Destrebecqz, A. (2011). Statistical learning of two artificial languages presented successively: How conscious? *Frontiers in Psychology, 2*, 1-12.

Frankland, P. W., & Bontempi, B. (2005). The organization of recent and remote memories. *Nature Reviews Neuroscience, 6*(2), 119-130.

Friedrich, M., Wilhelm, I., Born, J., & Friederici, A. D. (2015). Generalization of word meanings during infant sleep. *Nature Communications, 6*(1), 1-9.

Gabrieli, J. D. (1998). Cognitive neuroscience of human memory. *Annual Review of Psychology, 49*(1), 87-115.

Gais, S., Lucas, B., & Born, J. (2006). Sleep after learning aids memory recall. *Learning & Memory, 13*(3), 259-262.

Gaskell, M. G. (2024). EPS mid-career prize: An integrated framework for the learning, recognition and interpretation of words. *Quarterly Journal of Experimental Psychology, 77*(12), 2365-2384.

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

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

Gluck, M. A., & Myers, C. E. (1993). Hippocampal mediation of stimulus representation: A computational theory. *Hippocampus, 3*(4), 491-516.

Goldman-Rakic, P. S., Selemon, L. D., & Schwartz, M. L. (1984). Dual pathways connecting the dorsolateral prefrontal cortex with the hippocampal formation and parahippocampal cortex in the rhesus monkey. *Neuroscience, 12*(3), 719-743.

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

Gómez, R. L., & Edgin, J. O. (2016). The extended trajectory of hippocampal development: Implications for early memory development and disorder. *Developmental Cognitive Neuroscience, 18,* 57-69.

Gómez, R. L., Bootzin, R. R., & Nadel, L. (2006). Naps promote abstraction in language-learning infants. *Psychological Science, 17*(8), 670-674.

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

Graf, P., & Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*(3), 501-518.

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution, 7*(4), 493-498.

Hardt, O., Nader, K., & Nadel, L. (2013). Decay happens: The role of active forgetting in memory. *Trends in Cognitive Sciences, 17*(3), 111-120.

Harrison, L. M., Duggins, A., & Friston, K. J. (2006). Encoding uncertainty in the hippocampus. *Neural Networks, 19*(5), 535-546.

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. R., Brown, H., & Gareth Gaskell, M. (2012). Consolidation of vocabulary is associated with sleep in children. *Developmental Science, 15*(5), 674-687.

Henderson, L., Devine, K., Weighall, A., & Gaskell, G. (2015). When the daffodat flew to the intergalactic zoo: Offline consolidation is critical for word learning from stories. *Developmental Psychology, 51*(3), 406-417.

Hickey, A. J. (2022). *Grammatical Generalisation in Statistical Learning: The Contribution of Vocabulary, Explicit and Implicit Knowledge* [Doctoral dissertation, University of York].

Hockett, C. F., & Hockett, C. D. (1960). The origin of speech. *Scientific American, 203*(3), 88-97.

Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., & Dement, W. C. (1973). Quantification of sleepiness: A new approach. *Psychophysiology, 10*(4), 431-436.

Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research, 29*(1), 33-56.

James, E., Gaskell, M. G., & Henderson, L. M. (2019). Offline consolidation supersedes prior knowledge benefits in children's (but not adults') word learning. *Developmental Science, 22*(3), e12776.

James, E., Gaskell, M. G., Murphy, G., Tulip, J., & Henderson, L. M. (2022). Word learning in the context of semantic prior knowledge: Evidence of interference from feature-based neighbours in children and adults. *Language, Cognition and Neuroscience*, 1-18.

James, E., Gaskell, M. G., Pearce, R., Korell, C., Dean, C., & Henderson, L. M. (2021). The role of prior lexical knowledge in children's and adults' incidental word learning from illustrated stories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 47*(11), 1856-1869.

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.

Jenkins, J. G., & Dallenbach, K. M. (1924). Obliviscence during sleep and waking. The American *Journal of Psychology, 35*(4), 605-612.

Joanisse, M. F., & Seidenberg, M. S. (1999). Impairments in verb morphology after brain injury: A connectionist model. *Proceedings of the National Academy of Sciences of the United States of America, 96*(13), 7592-7597.

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

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

Kumaran, D., & McClelland, J. L. (2012). Generalization through the recurrent interaction of episodic memories: A model of the hippocampal system. *Psychological Review, 119*(3), 573-616.

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*(7), 512-534.

Kumle, L., Võ, M. L. H., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: An open introduction and tutorial in R. *Behavior Research Methods, 53*(6), 2528-2543.

Kurdziel, L. B., & Spencer, R. M. (2016). Consolidation of novel word learning in native English-speaking adults. *Memory, 24*(4), 471-481.

Lahl, O., Wispel, C., Willigens, B., & Pietrowsky, R. (2008). An ultra-short episode of sleep is sufficient to promote declarative memory performance. *Journal of Sleep Research, 17*(1), 3-10.

Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language, 63*(4), 447-464.

Lewis, P. A., & Durrant, S. J. (2011). Overlapping memory replay during sleep builds cognitive schemata. *Trends in Cognitive Sciences, 15*(8), 343-351.

Lindsay, S., & Gaskell, M. G. (2013). Lexical integration of novel words without sleep. Journal of Experimental Psychology: *Learning, Memory, and Cognition, 39*(2), 608-622.

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology, 6*, 1171.

Marr, D. (1971). Simple memory: A theory for archicortex. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 262*(841), 23-81.

Mattys, S. L., & Clark, J. H. (2002). Lexical activity in speech processing: Evidence from pause detection. *Journal of Memory and Language, 47*(3), 343-359.

McClelland, J. L. (2013). Incorporating rapid neocortical learning of new schema-consistent information into complementary learning systems theory. *Journal of Experimental Psychology: General, 142*(4), 1190-1210.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*(1), 1-86.

McClelland, J. L., & Patterson, K. (2002). Rules or connections in past-tense inflections: What does the evidence rule out?. *Trends in Cognitive Sciences, 6*(11), 465-472.

McClelland, J. L., McNaughton, B. L., & Lampinen, A. K. (2020). Integration of new information in memory: New insights from a complementary learning systems perspective. *Philosophical Transactions of the Royal Society B, 375*(1799), 20190637.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102*(3), 419-457.

Mirković, J., & Gaskell, M. G. (2016). Does sleep improve your grammar? Preferential consolidation of arbitrary components of new linguistic knowledge. *PLOS ONE, 11*(4), e0152489.

Mirković, J., Forrest, S., & Gaskell, G. (2011). Semantic regularities in grammatical categories: Learning grammatical gender in an artificial language. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (vol.33, pp. 324-329).

Mirković, J., MacDonald, M. C., & Seidenberg, M. S. (2005). Where does gender come from? Evidence from a complex inflectional system. *Language and Cognitive Processes, 20*(1–2), 139-167.

Mirković, J., Vinals, L., & Gaskell, M. G. (2019). The role of complementary learning systems in learning and consolidation in a quasi-regular domain. *Cortex, 116*, 228-249.

Mirković, J., Yee, M., Kennedy, M., & Hayiou-Thomas, M. E. (2021). *Cognitive mechanisms underpinning statistical learning of grammatical regularities in child and adult learners: Evidence for a multicomponential view*. [Preprint]. OSF.

Mölle, M., Marshall, L., Gais, S., & Born, J. (2002). Grouping of spindle activity during slow oscillations in human non-rapid eye movement sleep. *Journal of Neuroscience, 22*(24), 10941-10947.

Monaghan, P., Christiansen, M. H., & Fitneva, S. A. (2011). The arbitrariness of the sign: Learning advantages from the structure of the vocabulary. *Journal of Experimental Psychology: General, 140*(3), 325-347.

Morgan-Short, K., Faretta-Stutenberg, M., Brill-Schuetz, K. A., Carpenter, H., & Wong, P. C. (2014). Declarative and procedural memory as individual differences in second language acquisition. *Bilingualism: Language and Cognition, 17*(1), 56-72.

Newbury, C. R., Crowley, R., Rastle, K., & Tamminen, J. (2021). Sleep deprivation and memory: Meta-analytic reviews of studies on sleep deprivation before and after learning. *Psychological Bulletin, 147*(11), 1215-1235.

Newman, A. J., Pancheva, R., Ozawa, K., Neville, H. J., & Ullman, M. T. (2001). An event-related fMRI study of syntactic and semantic violations. *Journal of Psycholinguistic Research, 30*(3), 339-364.

Nieuwenhuis, I. L., Folia, V., Forkstam, C., Jensen, O., & Petersson, K. M. (2013). Sleep promotes the extraction of grammatical rules. *PLOS ONE, 8*(6), e65046.

O'Reilly, R. C., & Norman, K. A. (2002). Hippocampal and neocortical contributions to memory: Advances in the complementary learning systems framework. *Trends in Cognitive Sciences, 6*(12), 505-510.

O'Reilly, R. C., & Rudy, J. W. (2001). Conjunctive representations in learning and memory: Principles of cortical and hippocampal function. *Psychological Review, 108*(2), 311-345.

Ohala, J. J. (1984). An ethological perspective on common cross-language utilization of F_0 of voice. *Phonetica, 41*(1), 1-16.

Onnis, L., Waterfall, H., & Edelman, S. (2008). *Variation sets facilitate artificial language learning. In Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 971-976). Cognitive Science Society.

Pacton, S., Perruchet, P., Fayol, M., & Cleeremans, A. (2001). Implicit learning out of the lab: The case of orthographic regularities. *Journal of Experimental Psychology: General, 130*(3), 401-426.

Palma, P., & Titone, D. (2021). Something old, something new: A review of the literature on sleep-related lexicalization of novel words in adults. *Psychonomic Bulletin & Review, 28*(1), 96-121.

Perniss, P., Thompson, R. L., & Vigliocco, G. (2010). Iconicity as a general property of language: Evidence from spoken and signed languages. *Frontiers in Psychology, 1,* 227.

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

Poldrack, R. A., Clark, J., Paré-Blagoev, E. J., Shohamy, D., Creso Moyano, J., Myers, C., & Gluck, M. A. (2001). Interactive memory systems in the human brain. *Nature, 414*(6863), 546-550.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77*(3, Pt. 1), 353-363.

Postle, B. R., & Corkin, S. (1998). Impaired word-stem completion priming but intact perceptual identification priming with novel words: Evidence from the amnesic patient HM. *Neuropsychologia, 36*(5), 421-440.

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

Rasch, B., & Born, J. (2013). About sleep's role in memory. *Physiological Reviews, 93*(2), 681-766.

Rasch, B., & Born, J. (2015). In search of a role of REM sleep in memory formation. *Neurobiology of Learning and Memory, 122*, 1-3.

Rasch, B., Büchel, C., Gais, S., & Born, J. (2007). Odor cues during slow-wave sleep prompt declarative memory consolidation. *Science, 315*(5817), 1426-1429.

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

Reber, P. J. (2013). The neural basis of implicit learning and memory: A review of neuropsychological and neuroimaging research. *Neuropsychologia, 51*(10), 2026-2042.

Rebuschat, P. (2013). Measuring implicit and explicit knowledge in second language research. *Language Learning, 63*(3), 595-626.

Rockland, K. S., & Van Hoesen, G. W. (1999). Some temporal and parietal cortical connections converge in CA1 of the primate hippocampus. *Cerebral Cortex, 9*(3), 232-237.

Sadeh, T., Ozubko, J. D., Winocur, G., & Moscovitch, M. (2014). How we forget may depend on how we remember. *Trends in Cognitive Sciences, 18*(1), 26-36.

Saffran, J. R., & Wilson, D. P. (2003). From syllables to syntax: Multilevel statistical learning by 12-month-old infants. *Infancy, 4*(2), 273-284.

Sawi, O. M., & Rueckl, J. (2019). Reading and the neurocognitive bases of statistical learning. *Scientific Studies of Reading, 23*(1), 8-23.

Schapiro, A. C., Gregory, E., Landau, B., McCloskey, M., & Turk-Browne, N. B. (2014). The necessity of the medial temporal lobe for statistical learning. *Journal of Cognitive Neuroscience, 26*(8), 1736-1747.

Schapiro, A. C., Turk-Browne, N. B., Botvinick, M. M., & Norman, K. A. (2017). Complementary learning systems within the hippocampus: A neural network modelling approach to reconciling episodic memory with statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences, 372*(1711), 20160049.

Schendan, H. E., Searl, M. M., Melrose, R. J., & Stern, C. E. (2003). An FMRI study of the role of the medial temporal lobe in implicit and explicit sequence learning. *Neuron, 37*(6), 1013-1025.

Schimke, E. A., Angwin, A. J., Cheng, B. B., & Copland, D. A. (2021). The effect of sleep on novel word learning in healthy adults: A systematic review and meta-analysis. *Psychonomic Bulletin & Review, 28*(6), 1811-1838.

Schreiner, T., & Rasch, B. (2015). Boosting vocabulary learning by verbal cueing during sleep. *Cerebral Cortex, 25*(11), 4169-4179.

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery, and Psychiatry, 20*(1), 11-21.

Seidenberg, M. S., & MacDonald, M. C. (2018). The impact of language experience on language and reading: A statistical learning approach. *Topics in Language Disorders, 38*(1), 66-83.

Seidenberg, M. S., & Plaut, D. C. (2014). Quasiregularity and its discontents: The legacy of the past tense debate. *Cognitive science, 38*(6), 1190-1228.

Shanks, D. R., & Berry, C. J. (2012). Are there multiple memory systems? Tests of models of implicit and explicit memory. *Quarterly Journal of Experimental Psychology, 65*(8), 1449-1474.

Shanks, D. R., & John, M. F. S. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences, 17*(3), 367-395.

Shohamy, D., & Turk-Browne, N. B. (2013). Mechanisms for widespread hippocampal involvement in cognition. *Journal of Experimental Psychology: General, 142*(4), 1159-1170.

Siapas, A. G., & Wilson, M. A. (1998). Coordinated interactions between hippocampal ripples and cortical spindles during slow-wave sleep. *Neuron, 21*(5), 1123-1128.

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

Smalle, E. H. M., Page, M. P. A., Duyck, W., Edwards, M., & Szmalec, A. (2018). Children retain implicitly learned phonological sequences better than adults: A longitudinal study. *Developmental Science, 21*(5), e12634.

Squire, L. R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory, 82*(3), 171-177.

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, 18*(11), 1679-1686.

Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature, 437*(7063), 1272-1278.

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*(2), 306-320.

Sweegers, C. C., & Talamini, L. M. (2014). Generalization from episodic memories across time: A route for semantic knowledge acquisition. *Cortex, 59*, 49-61.

Szmalec, A., Page, M. P. A., & Duyck, W. (2012). The development of long-term lexical representations through Hebb repetition learning. *Journal of Memory and Language, 67*(3), 342-354.

Tamminen, J., & Gaskell, M. G. (2008). Short article: Newly learned spoken words show long-term lexical competition effects. *Quarterly Journal of Experimental Psychology, 61*(3), 361-371.

Tamminen, J., Davis, M. H., & Rastle, K. (2015). From specific examples to general knowledge in language learning. *Cognitive Psychology, 79*, 1-39.

Tamminen, J., Davis, M. H., Merkx, M., & Rastle, K. (2012). The role of memory consolidation in generalisation of new linguistic information. *Cognition, 125*(1), 107-112.

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. *Journal of Neuroscience, 30*(43), 14356-14360.

Tamminen, J., Ralph, M. A. L., & Lewis, P. A. (2013). The role of sleep spindles and slow-wave activity in integrating new information in semantic memory. *Journal of Neuroscience, 33*(39), 15376-15381.

Tanenhaus, M. K., & Brown-Schmidt, S. (2008). Language processing in the natural world. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*(1493), 1105-1122.

Timmermans, B., & Cleeremans, A. (2015). How can we measure awareness? An overview of current methods. In *Behavioural Methods in Consciousness Research* (pp. 21-46).

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*(5821), 76-82.

Tse, D., Takeuchi, T., Kakeyama, M., Kajii, Y., Okuno, H., Tohyama, C., ... & Morris, R. G. (2011). Schema-dependent gene activation and memory encoding in neocortex. *Science, 333*(6044), 891-895.

Ullman, M. T. (2001). A neurocognitive perspective on language: The declarative/procedural model. *Nature Reviews Neuroscience, 2*(10), 717-726.

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition, 92*(1-2), 231-270.

Van Kesteren, M. T., Ruiter, D. J., Fernández, G., & Henson, R. N. (2012). How schema and novelty augment memory formation. *Trends in Neurosciences, 35*(4), 211-219.

Venables, W. N., & Ripley, B. D. (2002). Modern applied statistics with S (4th ed.). Springer. https://doi.org/10.1007/978-0-387-21706-2.

Vertes, R. P., & Eastman, K. E. (2000). The case against memory consolidation in REM sleep. *Behavioral and Brain Sciences, 23*(6), 867-876.

Vlach, H. A., & Sandhofer, C. M. (2011). Developmental differences in children's context-dependent word learning. *Journal of Experimental Child Psychology, 108*(2), 394-401.

Voeten, C. C. (2021). buildmer: *Stepwise elimination and term reordering for mixed-effects regression* (Version 2.4) [R package]. CRAN. https://cran.r-project.org/package=buildmer.

Vujović, M., Ramscar, M., & Wonnacott, E. (2021). Language learning as uncertainty reduction: The role of prediction error in linguistic generalization and item-learning. *Journal of Memory and Language, 119*, 104231.

Vuong, L. C., Meyer, A. S., & Christiansen, M. H. (2016). Concurrent statistical learning of adjacent and nonadjacent dependencies. *Language Learning, 66*(1), 8–30.

Wagner, U., Hallschmid, M., Verleger, R., & Born, J. (2003). Signs of REM sleep dependent enhancement of implicit face memory: A repetition priming study. *Biological Psychology, 62*(3), 197-210.

Walker, M. P., & Stickgold, R. (2004). Sleep-dependent learning and memory consolidation. *Neuron, 44*(1), 121-133.

Walker, M. P., & Stickgold, R. (2010). Overnight alchemy: Sleep-dependent memory evolution. *Nature Reviews Neuroscience, 11*(3), 218-218.

Wang, H. C., Savage, G., Gaskell, M. G., Paulin, T., Robidoux, S., & Castles, A. (2017). Bedding down new words: Sleep promotes the emergence of lexical competition in visual word recognition. *Psychonomic Bulletin & Review, 24*(4), 1186–1193.

Werchan, D. M., & Gómez, R. L. (2014). Wakefulness (not sleep) promotes generalisation of word learning in 2.5-year-old children. *Child Development, 85*(2), 429-436.

Wilson, M. A., & McNaughton, B. L. (1994). Reactivation of hippocampal ensemble memories during sleep. *Science, 265*(5172), 676-679.

Witt, A., Puspitawati, I., & Vinter, A. (2013). How explicit and implicit test instructions in an implicit learning task affect performance. *PLOS ONE, 8*(1), e53296.

Yi, H. G., Maddox, W. T., Mumford, J. A., & Chandrasekaran, B. (2016). The role of corticostriatal systems in speech category learning. *Cerebral Cortex, 26*(4), 1409-1420.

Zion, D. B., Nevat, M., Prior, A., & Bitan, T. (2019). Prior knowledge predicts early consolidation in second language learning. *Frontiers in Psychology, 10*, 2312.