Article:

James, Emma orcid.org/0000-0002-5214-0035, Gaskell, M Gareth orcid.org/0000-0001-8325-1427, Pearce, Rhiannon et al. (3 more authors) (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. 1856–1869. ISSN 1939-1285

https://doi.org/10.31234/osf.io/vm5ad

The role of prior lexical knowledge in children's and adults' incidental word learning from illustrated stories

Emma James[1,*], M. Gareth Gaskell[2], Rhiannon Pearce[2], Caroline Korell[3], Charlotte Dean[2], & Lisa M. Henderson[2]

1. Department of Experimental Psychology, University of Oxford, Oxford, UK
2. Department of Psychology, University of York, UK
3. Department of Education, University of Oxford, Oxford, UK

*Corresponding author:
Department of Experimental Psychology, Anna Watts Building, University of Oxford,
Oxford, OX2 6GG, UK
emma.james@psy.ox.ac.uk

The published version of this manuscript can now be found at:
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.* https://doi.org/10.1037/xlm0001080

Author note

Accompanying pre-registrations, materials, data, and analyses can be accessed at https://osf.io/stx6q/.

E.J. was supported by an ESRC 1+3 studentship and ESRC Postdoctoral Fellowship ES/T007524/1. The research was funded by ESRC grant ES/N009924/1 awarded to M.G.G. and L.M.H.

Abstract

Children and adults benefit from a new word's phonological neighbours during explicit vocabulary instruction, suggesting that related prior knowledge can support new learning. This study examined the influence of lexical neighbourhood structure during incidental word learning—limiting opportunities for strategically engaging prior knowledge—and tested the hypothesis that prior knowledge would provide additional support during subsequent consolidation. Children aged 8-10 years (Experiment 1) and adults (Experiment 2) were presented with 15 pseudowords embedded in a spoken story with illustrations, and were then tested on their recognition and recall of the new word-forms immediately, the next day, and one week later. The pseudowords had either no, one, or many English phonological neighbours, varying the potential connections to existing knowledge. After encountering the pseudowords in this incidental training paradigm, neither children nor adults benefitted from phonological neighbours in recall, and children were better at recognising items without neighbours. The neighbour influence did not change with opportunities for consolidation in either experiment, nor did it relate to learners' existing vocabulary ability. Exploratory analyses revealed that children experienced bigger benefits from offline consolidation overall, with adults outperforming children only for many-neighbour items one week after exposure. We discuss how the neighbour benefit in word learning may be constrained by learning context, and how the enhanced benefits of offline consolidation in childhood extend to vocabulary learning in more naturalistic contexts.

Keywords: word learning; prior knowledge; consolidation; vocabulary; learning contexts

Both children and adults face the task of learning new vocabulary from a multitude of situations. Many words are acquired explicitly and intently, for example via early language learning experiences and formal vocabulary instruction in school. However, the breadth of vocabulary knowledge far exceeds the number of words that can be actively taught (Nagy & Herman, 1987), leaving the majority of words to be encountered incidentally and learned from context (Sternberg, 1987). Understanding the factors that influence word learning in these incidental learning contexts is therefore key to understanding variability in vocabulary growth across development. In intentional learning contexts, studies suggest that children and adults can "bootstrap" new words to their existing knowledge of similar word-forms to speed new word acquisition (James et al., 2019; Vitevitch et al., 2014). In this study, we asked whether this prior lexical knowledge also supports vocabulary learning from spoken illustrated stories, an important source of new vocabulary for children (Montag et al., 2015). In doing so, we aimed to understand how learning mechanisms may be differently engaged in incidental versus intentional learning contexts. We looked at whether children and adults benefited from word-form similarities immediately after learning and after opportunities for consolidation, examining the impact of prior knowledge on each respective process in this incidental learning context.

Learning and consolidating new vocabulary

Models of learning distinguish between processes that help us to quickly acquire new words from the environment and processes that support their long-term consolidation into existing vocabulary, suggesting that variability might emerge at multiple stages of new word acquisition. Davis and Gaskell (2009) applied the Complementary Learning Systems model (McClelland et al., 1995) to vocabulary learning, describing two neural systems involved in these two aspects of new word acquisition. The hippocampal learning system enables rapid learning about a new word: its phonological form, its meaning, and syntactic properties. This newly formed representation is proposed to be stored in a sparse and relatively distinct manner. The neocortical learning system represents longer-term memory, whereby the distributed nature of lexical storage allows for robust maintenance of linguistic knowledge. Integration of a new word representation into this existing vocabulary system is a slower process requiring a prolonged period of consolidation, proposed to involve reactivation of the hippocampal representations to allow relevant connections to become strengthened within the neocortex. This reactivation can occur via retrieval practice (Antony et al., 2017) and/or by processes that occur during sleep. For example, sleep-associated benefits have been consistently observed for knowledge of word-forms (James, Gaskell, et al., 2020), but have also been seen for word meanings (e.g., McGregor et al., 2013). Further, these sleep-associated benefits have been observed across an array of explicit and incidental learning contexts (e.g., Brown et al., 2012; Williams & Horst, 2014), and even outside of experimentally controlled conditions (James, Koutraki, et al., 2020). In this study, we do not dissociate between retrieval practice and sleep-associated mechanisms in supporting consolidation, but test new word memory across multiple days to examine how predictors of consolidation may differ according to the learning context.

Recent models of word learning have considered the factors that might influence how well individuals consolidate new vocabulary in a complementary systems account. The starting point for this work is the hypothesis that new memory traces are easier to incorporate in cortical networks if they do not strongly interfere with the existing traces (e.g., Kumaran et al., 2016; McClelland, 2013; Tse et al., 2007). In cognitive terms this would mean that new memories that are similar in representation to existing ones will be integrated into long-term memory relatively swiftly. Drawing on these schema-based accounts of learning and memory, James et al. (2017) proposed that prior linguistic knowledge may facilitate rapid integration of new words into the lexicon. A rich existing vocabulary may allow more connections between new lexical items and existing word representations to be made during learning. Then, as the hippocampal representations of new words are reactivated during consolidation, shared connections can enhance the benefits of these reactivation processes (Lewis & Durrant, 2011), thus permitting more rapid strengthening of new knowledge. Below, we describe two broad approaches that have been taken to examine this hypothesis in a language learning context: (i) assessing variability in the prior knowledge that learners themselves bring to the task, and (ii) experimentally manipulating the availability of prior knowledge at the lexical level. Notably, these approaches have also tended to differ in whether they assess incidental or intentional word learning—a limitation we aimed to address in the present study.

The role of global prior knowledge in word learning

Starting with the approach of examining variability between learners, one basis for considering the role of prior knowledge in word learning was evidence that individual differences in existing vocabulary knowledge predict improvements in memory for new words during offline consolidation (e.g., Henderson et al., 2015; Horváth et al., 2015; Sénéchal et al., 1995). We refer to these separate measures of vocabulary ability as "global" prior knowledge to reflect the wealth of knowledge that the learner brings to the task. For example, Henderson and James (2018) presented 10- to 11-year-old children with novel words (e.g., *crocodol*) embedded in spoken stories with accompanying illustrations. When tested with a stem completion task ("Which word began with *cro*-?"), children with higher scores on a standardised vocabulary assessment improved more in their knowledge of the word-forms overnight than children with poorer vocabulary. This benefit was specific to learning new words in varied (and not repeated) story contexts, suggesting that children with richer vocabulary may benefit from opportunities to form multiple connections to their prior knowledge under these conditions. A similar pattern was observed for categorisation of the new words in an animacy decision task, suggesting that prior knowledge may support both form and meaning aspects of new word knowledge.

However, the mechanisms underlying this consolidation benefit are not yet clear. Notably, the studies that have demonstrated this relationship draw largely upon methods of implicit word learning—either examining words learned incidentally from context, or for very young children for whom intentional learning strategies are not yet well-developed. Speaking to this, the association between global prior knowledge and consolidation has not been replicated in studies using explicit vocabulary instruction (James et al., 2019). This could mean that the associations arise as a consequence of individual differences in broader language or comprehension ability, rather than providing support at the lexical level as proposed within the complementary learning systems framework. Alternatively, it could be that the learning mechanisms engaged differ in the context of intentional versus incidental word learning. During implicit language learning, individuals are proposed to show increased dependence on procedural memory systems, rather than the declarative systems engaged when learning under explicit instruction (Ullman, 2015, 2016). Procedural networks centred on the basal ganglia are argued to be optimised for learning predictively based on sequential structure (Ullman, 2015), and as such, predictors of learning and consolidation may vary under different learning contexts.

Local manipulations of prior knowledge in studies of word learning

The second approach to examining the contribution of prior knowledge to word learning is to manipulate potential similarities between specific items—hereafter "local" prior knowledge. Many studies have used phonological neighbours as a way of quantifying a new word's potential links to existing word knowledge. For example, studies by Storkel and colleagues (Hoover et al., 2010; Storkel, 2009; Storkel et al., 2006) taught participants pseudowords that varied in the number of real words that could be created by substituting a single phoneme. Thus, words with few phonological neighbours have more limited potential connections to existing knowledge than words with many phonological neighbours.

Experiments using this paradigm have consistently demonstrated that phonological neighbours facilitate word learning from preschool to adulthood (Storkel, 2009; Storkel et al., 2006; see Swingley & Aslin, 2007, for findings of opposite effects in younger infants) and across a range of languages (e.g., van der Kleij et al., 2016). These benefits are seen across a variety of tasks that differently capture word-form knowledge and/or the mapping to a semantic referent, and have also been observed in normed databases of vocabulary acquisition (Jones & Brandt, 2019; Siew & Vitevitch, 2020).

While phonological neighbours are consistently found to benefit new learning from early childhood, their role in consolidation is less clear. Based on the relations observed between global prior knowledge and overnight consolidation, one might predict that neighbour benefits are exacerbated during consolidation, following increased opportunities for the new words to engage with existing vocabulary knowledge. Storkel and Lee (2011) found results that were consistent with this pattern: 4-year-old children exhibited a neighbourhood density benefit for recognising non-object referents only at a one-week retention test, and not when tested immediately after learning. For these children then, prior knowledge benefits were enhanced following a period of offline consolidation.

However, a recent study indicated the opposite pattern: an immediate benefit of local prior knowledge that *diminished* after opportunities for offline consolidation. James et al. (2019) used direct vocabulary instruction to teach 7-to-9-year-old children pseudowords paired with novel object referents, and found a significant benefit of phonological neighbours in recalling the word-forms immediately after instruction but not at the one-week retention test. One explanation for this finding was that the pseudowords with strong connections to prior knowledge might engage with the neocortical system immediately: in being able to capitalise upon existing connections in the lexicon, there may be a reduced need for integrating new vocabulary offline. Conversely, phonologically distinct pseudowords that do not benefit from local prior knowledge may rely more on the hippocampal memory system at encoding, and therefore receive greater benefit from offline consolidation processes (see Havas et al., 2018; Himmer et al., 2017; Mirković & Gaskell, 2016, for similar interpretations). This study also included a measure of global vocabulary knowledge and, perhaps unsurprisingly given the early neighbour benefits, there was no evidence to support that global vocabulary knowledge predicted offline consolidation after learning in this explicit teaching context—only an association with overall memory performance.

In sum, it is clear that local connections to prior knowledge can facilitate new word learning in children and in adults under explicit vocabulary instruction, but that these benefits do not always further support consolidation. In the context of the complementary learning systems model, it may be that pseudowords with strong local connections to prior knowledge have reduced need for neocortical connections to be strengthened offline. These early benefits for phonological neighbours may be exacerbated under conditions of explicit vocabulary instruction: when participants direct their attention towards actively trying to learn the word-forms and meaning, similarities to known word-forms may become part of the learning process. Indeed, adults in James et al. (2019) reported making intentional comparisons to the words they knew, and using those similarities to make semantic connections with the novel objects being learned (from subjective reports, data unpublished).

Conversely, the young children in Storkel and Lee's (2011) study—who were perhaps less able to explicitly capitalise upon their prior knowledge to support them during initial instruction—did *not* benefit from local prior knowledge immediately after learning, but *did* show a prior knowledge benefit during consolidation. These weaker connections to prior knowledge formed during encoding may leave potential prior knowledge connections susceptible to further strengthening during later consolidation. Further, this pattern of results may better relate to the observed relationships between learners' global prior knowledge and overnight improvements in new word memory described above, primarily observed in story learning contexts. Thus, we proposed that under conditions of incidental learning from stories, both local and global prior knowledge would support offline consolidation of new vocabulary.

The present study

In this study, we tested whether incidental learning of novel words in a storybook context leads to later-emerging benefits of local prior knowledge. A key reason for doing so was that the majority of studies using phonological neighbours have used direct vocabulary instruction, and it is not yet clear the extent to which intentional learning strategies may drive the prior knowledge benefits observed in school-aged children and adults. Conversely, when new words are encountered incidentally as part of an ongoing narrative, individuals have less opportunity to draw comparisons between new and known words in a strategic manner. Note that while previous studies of phonological neighbours in word learning have sometimes presented the items in stories (e.g., Hoover et al., 2010; Storkel et al., 2006), participants have still been made aware of the learning nature of the task, with test trials and/or explicit revision of the items interleaved with story exposures. Here, we examined the contributions of prior knowledge following *incidental* (versus intentional) learning of novel words incorporated in a storybook context, akin to shared reading activities that present a rich opportunity for vocabulary learning in children (Dawson et al., 2021; Flack et al., 2018; Montag et al., 2015). By pairing a spoken narrative with illustrations, our approach aligned with previous studies of incidental learning via stories (e.g., Henderson & James, 2018; Sénéchal et al., 1995), but it is important to note that we cannot (nor is it our goal to) isolate pure contributions of the spoken narrative to any observed prior knowledge effects. Rather, the use of spoken narratives and illustrations allow us to examine how differences in the learning context may influence the presence and longevity of prior knowledge benefits.

Importantly, by embedding local prior knowledge manipulations in an incidental word learning paradigm, we sought to bring local and global approaches to understanding prior knowledge into line—i.e., how the immediate benefits of local prior knowledge seen in explicit training studies might related to the delayed consolidation benefits of global prior knowledge observed in incidental vocabulary learning studies. To do this, we embedded a subset of 15 pseudowords from James et al. (2019) into a spoken story adapted from Henderson et al. (2015), accompanied by illustrations. The pseudowords came from one of three phonological neighbour conditions (none, one, many), enabling us to assess directly how these potential connections to existing knowledge might benefit new word acquisition for children (Experiment 1) and adults (Experiment 2). To assess the time course of these local prior knowledge benefits, we used recall and recognition tasks to test memory for the new word-forms immediately after learning, the next day, and one week later. We predicted that the benefit of phonological neighbours would be initially weaker after learning pseudowords from stories—relative to previous studies that used direct teaching (James et al., 2019)—and that these more fragile connections to prior knowledge would be *strengthened* over a period of offline consolidation. That is, we expected to see a larger neighbourhood effect emerging one day and/or one week later, resulting from the increased opportunities for new pseudowords to engage with learners' global vocabulary knowledge.

We tested our predictions in both children (Experiment 1) and adults (Experiments 2) for two reasons. First, the availability of prior knowledge is a key consideration in understanding how word learning changes across development: differences in neural processes during sleep are suggested to support enhanced offline benefits for children, whereas adults may be better able to benefit from their enhanced linguistic knowledge to support consolidation (James et al., 2017). One previous study supported the proposal that children might benefit more from offline consolidation relative to adults (James et al., 2019), but the experiments under comparison trained different numbers of items. Here, we used an identical paradigm in adults and children, allowing us to more carefully examine developmental differences. Second, children showed low levels of initial learning from stories, presenting a challenge to interpret contrasting findings between explicit training and incidental learning studies. Repeating the experiment in adults—who typically show higher levels of initial learning (e.g., Weighall et al., 2017)—thus provided further insight into whether differences from explicit training studies might be attributable to weaker learning rather than the prior knowledge mechanisms of interest. As such, each experiment set out to address the same primary research questions: 1) Do individuals show benefits of offline consolidation for pseudowords encountered incidentally in stories? 2) Do individuals show benefits of local prior knowledge (i.e., phonological neighbours) in remembering pseudowords encountered incidentally in stories? 3) Is the time course of prior knowledge benefits different when individuals learn pseudowords incidentally from stories, compared to previous studies using explicit teaching paradigms? And 4) Are benefits of local prior knowledge related to individuals' global prior knowledge? Our first experiment tested these research questions in children aged 8-10 years.

Experiment 1

Experiment 1 Methods

Participants

Six Year 4 and 5 classes were recruited from across four primary schools in North Yorkshire, with the aim of acquiring at least as many observations per condition as James et al. (2019, Exp. 3). Consent was gained from each school's headteacher, alongside parental consent on an opt-out basis. The study was approved by the Psychology Departmental Ethics Committee at the University of York (ID 18106: *How does existing vocabulary support new word learning from stories?*). From the initial 123 children, two withdrew their participation. Twenty children were excluded from analyses as they were fluent in an additional language to

English (c.f., Meade et al., 2018), and a further four children could not be entered into analyses because of absence at the time of the vocabulary assessment. These additional datasets are available online. The final analyses incorporated data from 97 children (51 male), aged between 8;06 – 10;09 years (*M* = 9;07). This age range was selected to overlap with previous studies (James et al., 2019), but was slightly older to reduce risk of floor effects when learning from stories. Three children were absent on the second day of testing, and thus contributed only two sessions of data to the final analyses.

Design and Procedure

Each child participated in three test sessions, administered individually in a quiet area of the school. On Day 1, children completed the learning phase of the experiment, in which they were presented with an illustrated spoken story containing 15 pseudowords. There were five pseudowords from each of three neighbour conditions: none, one, or many (detailed below). Children completed memory tests for the pseudowords immediately afterwards (T1), the next day (T2), and one week later (T3). All experimental tasks were programmed in OpenSesame (Mathôt et al., 2012) and administered on a laptop with headphones.

Additional assessments of expressive vocabulary and nonverbal ability were conducted across the second and third test sessions, using the Vocabulary and Matrix Reasoning subtests of the Wechsler Abbreviated Scale of Intelligence (WASI-II; Wechsler, 2011). Children were also asked about languages spoken at home, to identify those who were not native monolingual English speakers for exclusion from the analyses.

Experimental stimuli

A subset of fifteen pseudowords from James et al. (2019) was used, a stimulus set that had previously been used to demonstrate a neighbour benefit in explicit learning for children of a similar age. These pseudowords had initially been selected from the English Lexicon Project (Balota et al., 2007) for having no, one, or many orthographic neighbours in a database of 40,481 English words. Orthographic neighbours are real English words that can be created by substituting a single letter. Comparably, phonological neighbours are those that can be created by substituting a single sound. The number of phonological neighbours for each pseudoword also aligned with their categorisation into the no, one, or many neighbour conditions (CLEARPOND; Marian et al., 2012). For example, the pseudoword *femod* has no neighbours in the English language; *tabric* has the single neighbour *fabric*; whereas *dester* has many neighbours including *duster, pester,* etc. We used fewer pseudowords than in the original study by James et al. (2019) as children typically show lower levels of learning from story contexts (Henderson et al., 2015). The three lists remained matched for phoneme and letter length, and bigram probability. The subsets also did not differ from the previous lists in terms of the neighbour frequency (ps > .85; calculated using the CELEX database via N-Watch, and summed per item for pseudowords in the many neighbour condition; Baayen et al. (1996); (Davis, 2005)). All pseudowords were bisyllabic, and began with a single consonant and vowel that could be used as a cue in the stem completion task.

The pseudowords trained in James et al. (2019) were each associated with a picture of a novel unnameable object to provide a semantic referent, meaning that the key differences for this experiment were the broader semantic context (from both the spoken narrative and the scene illustrations) and absence of explicit instruction to learn the new words. We used different semantic referents in this study to fit with the narrative context, as detailed below.

Learning phase

Children heard the pseudowords embedded in a story recorded by a female native English speaker, and presented with accompanying illustrations. The story was based on one created by Henderson et al. (2015)—*Trouble at the Intergalactic Zoo*—replacing the original pseudowords with those described above. The original story contained 12 pseudowords, but we extended the story to incorporate 15 pseudowords to preserve statistical power for comparing neighbour conditions. There were five exposures to each pseudoword within the story, which were spaced in their presentation across 3-4 different paragraphs.

Each of the pseudowords referred to a concrete object in the story (e.g., a *peflin* might refer to a cactus-flavoured drink, or a *rafar* to a particular creature found at the zoo). Three versions of the story were created such that each object could be paired with a pseudoword from the no-, one-, or many- neighbour conditions (counterbalanced across participants). To facilitate engagement with the story, the objects were incorporated into cartoon-like scenes that corresponded to each paragraph. Each object featured in three of the 15 scenes, co-occurring with its corresponding pseudoword appearing in the spoken narrative (in order to maintain a coherent narrative, there was one instance of an object being presented in a scene without its naming in the paragraph, but this should not affect the results given that its assignment to a neighbour condition was counterbalanced across participants). There was always more than one pseudoword and/or referent per scene, making it challenging for participants to learn the semantic mappings without comprehending the story (although this does not rule out contributions of statistical co-occurrences across scenes; Yu & Smith, 2007). The scenes were presented alongside the spoken story using OpenSesame, changing automatically with each paragraph. All materials are available online at https://osf.io/xwcz6, and a preview of the story can also be accessed at https://gorilla.sc/openmaterials/104397 (programmed via Gorilla for Experiment 2 below).

At the start of the learning phase, children were warned that they may not know all the words in the story, but that they should keep listening until the end of the story without asking questions. The story lasted 7 minutes and 10 seconds.

Test phase

Stem completion. Recall of the new word-forms was assessed using a stem completion task. Children were cued with the first consonant and vowel sound from each pseudoword, and were asked to speak the remainder of the word that they heard during the story. Partial attempts were encouraged even if children were not certain of their responses. Items were presented in a randomised order using OpenSesame, and the experimenter transcribed the responses for scoring offline. There was no time limit.

Form recognition. Children heard each new pseudoword paired with a phonological foil (incorporating a vowel change, e.g., *peflin - peflun*), and were asked to select which word they heard during the story. They responded using keys assigned to the first or second option, and completed two practice trials (known words and foils) with feedback to adjust to the response mappings. The trial timed out after 5 s.

Form-picture recognition. To assess learning of the semantic mappings, children were presented with each novel object on screen and asked to select its name from two pseudoword options using a key press. The incorrect answer for each trial was always another pseudoword heard during the story, and remained consistent across test sessions. There was no time limit.

Analyses

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

For each analysis, we first computed a random-intercepts model with all fixed effects and interactions. If there was no indication of a three-way interaction in the model (all *p*s > .2), these were pruned to enable a more parsimonious model with better-specified random effects. We then incorporated random slopes into the model using a forward best-path approach (Barr et al., 2013), progressively adding slopes into the model and retaining only those random effects justified by the data under a liberal α-criterion ($p < .2$). In the text, we report statistics in full for significant predictors of performance, and also non-significant predictors where $p < .08$. The final model details and all statistics are presented in Supplementary Materials.

Experiment 1 Results

Stem completion

The proportion of pseudowords successfully recalled after listening to the story was very low (T1: $M = 0.03$, $SD = 0.17$), but significantly improved at later test points ($\beta = 0.68$, $SE = 0.07$, $Z = 9.18$, $p < .001$; Figure 1A). Recall performance also continued to improve substantially between T2 ($M = 0.07$, $SD = 0.26$) and T3 ($M = 0.21$, $SD = 0.41$; $\beta = 0.84$, $SE = 0.08$, $Z = 10.63$, $p < .001$), supporting the hypothesis that opportunities for consolidation would improve recall for the pseudowords.

Vocabulary ability was a significant predictor of a child's recall performance ($\beta = 0.69$, $SE = 0.16$, $Z = 4.24$, $p < .001$), suggesting a general association between global prior knowledge and new word learning which did not change with consolidation. However, more local connections to prior knowledge did not facilitate memory for the pseudowords: there was no benefit of phonological neighbours overall or in interaction with any other variable (*p*s > .11; Table S1).

Form recognition

One participant's recognition data did not save properly at T1, and is missing from the recognition analyses. Immediately after hearing the story, children could recognise the pseudowords over their phonological foils at above chance performance (T1: *M* = .65, *SD* = .48; *t*(95) = 10.21, *p* < .001; Figure 1B). Performance improved at later test points (β = 0.23, *SE* = 0.03, *Z* = 8.96, *p* < .001), and continued to improve from T2 (*M* = .75, *SD* = 0.43) to T3 (*M* = .80, *SD* = 0.40; β = 0.17, *SE* = 0.05, *Z* = 3.58, *p* < .001). There was a significant effect of phonological neighbours on performance but—in contrast to our hypothesis—pseudowords with one (*M* = .72, *SD* = .45) or more (*M* = .71, *SD* = .46) neighbours were recognised more poorly than those without neighbours (*M* = .78, *SD* = .42; β = -0.12, *SE* = 0.06, *Z* = -2.10, *p* = .036).

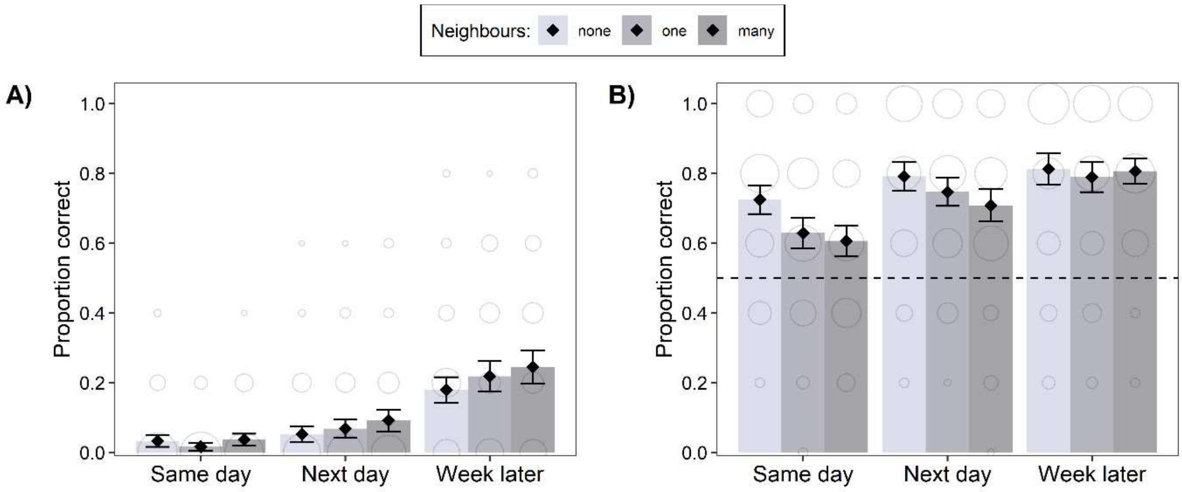
Vocabulary ability was again a significant predictor of performance for the recognition task (β = 0.30, *SE* = 0.08, *Z* = 3.88, *p* < .001). There was a trend towards an interaction with neighbour condition, suggesting that vocabulary ability better predicted performance in the no-neighbour condition. However, this was not statistically significant (β = -0.07, *SE* = 0.04, *Z* = -1.82, *p* = .069), and nor was any other interaction in the model (Table S2).

Picture-form recognition

Five participants were administered the incorrect version of this task during one session (according to their counterbalancing condition), and were excluded from this analysis. At the first test point, children could successfully select the correct name for the objects at above chance levels of performance (T1: *M* = 0.66, *SD* = 0.47; *t*(90) = 10.81, *p* < .001). Memory for these picture-form mappings improved overnight (T2: *M* = 0.72, *SD* = 0.45; β = 0.10, *SE* = 0.03, *Z* = 4.08, *p* < .001), but there were no further improvements across the week (T3: *M* = 0.72, *SD* = 0.45; *p* = .862) and performance did not differ according to neighbour condition (*p*s > .9; Table S3, Figure S1). As with the other two tasks, vocabulary ability was a significant predictor of performance overall (β = 0.35, *SE* = 0.09, *Z* = 3.96, *p* < .001). Vocabulary ability also interacted with test session in this task: children with higher vocabulary scores improved more from T1 to T2 and T3 than children with poorer vocabulary (β = 0.09, *SE* = 0.03, *Z* = 3.53, *p* < .001).

Figure 1

Proportion of items correct per condition in the A) Stem Completion, and B) Form Recognition tasks in Experiment 1.



Note. Error bars mark 95% confidence intervals. Bubble size indicates the proportion of observations at each level of accuracy. The dashed line in panel B marks chance-level performance.

Experiment 1 Discussion

In Experiment 1, we examined how children learn and consolidate pseudowords encountered incidentally in spoken stories presented with illustrations, akin to a storybook context. Children learned to recognise the word-forms and their meanings after listening to the stories, as demonstrated by their above-chance performance in the two recognition tasks. Recall performance, on the other hand, was very low at the first test point, and improved substantially by the day and week follow-up tests—consistent with models of offline consolidation, and with benefits of repeated retrieval practice.

The primary goal of this study was to determine whether individuals are influenced by their prior lexical knowledge when learning new words in this incidental learning context. Unlike previous experiments using explicit instruction (e.g., James et al., 2019; Storkel et al., 2013), there was no evidence that children benefited from local prior knowledge. It is important to note that while children of this age may not have always known the full range of pseudoword neighbours—a design choice key to examining individual variability in prior knowledge—two reasons make it unlikely that a lack of neighbour knowledge accounts for this null effect. First, 7- to 9-year-olds have previously shown robust neighbour benefits in recall following explicit training of the same pseudoword items (James et al., 2019). Second, the significant effect of neighbour condition in the form recognition task suggests that children in this experiment were influenced by at least some of the pseudoword neighbours during the experiment. However, this effect was in the opposite direction to our prediction: children were significantly worse at recognising recently encountered pseudowords that had phonological neighbours in the English language. In terms of global prior knowledge, vocabulary ability predicted performance in all three tasks, but there were no interactions with neighbour effects. There was some indication that children with good vocabulary showed greater improvements in the form-picture recognition task overnight than children with poorer vocabulary, and we return to speculate on possible explanations in the General Discussion.

The results support our prediction that children are less likely to access and benefit from prior lexical knowledge during incidental word learning from illustrated stories. However, we found limited support for the hypothesis that either local or global prior knowledge benefits emerge later with consolidation: neighbour benefits were absent throughout, and the influence of global vocabulary knowledge changed with consolidation for one task only. However, we also note that performance levels at test were low in this experiment, making it challenging to draw meaningful comparisons with prior studies of explicit training. We used 15 pseudowords to ensure comparable stimuli to a previous study (James et al., 2019) and retain adequate statistical power to detect differences between conditions, but this number is higher than would typically be included in studies of word learning from context. We therefore conducted a second experiment with adults who—like children—have previously shown benefits of phonological neighbours in explicit training conditions, but typically show better overall recall immediately after learning (e.g., Henderson et al., 2015; Weighall et al., 2017). In examining whether the (lack of) neighbour benefit remains for adults, we hoped to further inform our understanding of prior knowledge influences across word learning contexts.

Experiment 2

In light of our earlier results, we pre-registered four hypotheses on the Open Science Framework (https://osf.io/vfrw2): 1) Memory for the pseudowords will be different after opportunities for consolidation at the day and week follow-up tests compared to when tested immediately after learning. We predicted that recall in the stem completion task would improve at later test points. 2) Memory for the pseudowords will be affected by their number of phonological neighbours. We predicted that pseudowords with one/more neighbours would be better recalled than pseudowords without neighbours in the stem completion task, but did not predict a direction for this hypothesis in the recognition task. 3) The influence of phonological neighbours on memory for the pseudowords will change after opportunities for consolidation. 4) Expressive vocabulary scores will be positively associated with overall memory performance for the pseudowords, and that this association would be strongest for pseudowords with only one phonological neighbour (as in James et al., 2019).

In an initial experiment with adults (Experiment S1; pre-registered at http://osf.io/cdyrw), we incorporated written language in the design. Our reasons for doing this were threefold: 1) to provide additional orthographic support in learning the pseudowords, given that recall had been low in Experiment 1; 2) to increase the ease of administration for online testing; and 3) to make the experiment more comparable to the adult sample in James et al. (2019), who were presented with orthography during an explicit training regime and subsequent tests. We found that under these conditions, adults were influenced by local prior knowledge in learning words from illustrated stories: similar to previous studies that used explicit vocabulary instruction for the same stimuli (James et al., 2019), adult participants were better able to recall pseudowords with many neighbours in the English language than pseudowords without neighbours. However, we became concerned that the inclusion of orthography may have enabled participants to use more intentional word learning strategies. Adults can typically read at a faster pace than our spoken recording (Rayner, 1998), and may have been able to re-visit unfamiliar words and engage in explicit word learning strategies without compromising comprehension during the task—counter to our intention to study *incidental* word learning. Thus, we proceeded without orthography in Experiment 2, allowing us to compare children and adults more directly. The method and results for the experiment including orthography can be found in Supplementary Materials (Experiment S1).

Experiment 2 Methods

Participants

Experiment 2 was an online experiment. 125 adults were included in the analysis, and were recruited via Prolific Academic using the following criteria: aged 18-35 years old; native monolingual British English speakers residing in the UK; no reported visual, hearing, or literacy difficulties; had not taken part in Experiment S1; and had a working microphone that was compatible with the experiment platform. The target sample size was set to incorporate as many observations per condition as in James et al. (2019, Exp. 2), increasing the number of participants to account for the reduced number of items per neighbour condition in the present design. In line with Prolific's recommendations for longitudinal studies, we restricted recruitment to individuals who had participated in at least ten studies on the platform with a minimum 95% approval rate. The study was approved by the Psychology Departmental Ethics Committee at the University of York (ID 737: *Accessing prior knowledge when learning vocabulary from stories*). Participants received £5 for completion of all three sessions, and an additional £1 bonus if they completed each session within the same four-hour time window.

An additional 23 participants started the study but did not complete all three test sessions, and one participant failed the attention screener after listening to the story. A further 14 participants completed all three test sessions but were excluded for the following reasons: self-report of an inappropriate strategy (i.e., writing the words down; *n* = 5), failure to complete the vocabulary task properly (*n* = 7), insufficient recordings across two/more sessions (*n* = 1) or mismatch in demographic information (*n* = 1). The final sample had a mean age of 27.01 years (*SD* = 4.67), with 88 females and 37 males.

Design and procedure

The overarching procedure was identical to Experiments 1, incorporating an initial exposure and test session (T1), with follow-up sessions one day (T2) and one week later (T3). The three test sessions were programmed and hosted on the Gorilla Experiment Platform (Anwyl-Irvine et al., 2020; all tasks can be accessed at www.gorilla.sc/openmaterials/104397). The first session took approximately 20 minutes, including a sound and microphone check, collection of basic background information, presentation of the spoken story with illustrations, and the first set of memory tests (T1). As in Experiment 1, each participant was exposed to and tested on 15 pseudowords, 5 from each of three neighbour conditions (none, one, many). Like the children, adults were alerted to the presence of unfamiliar words but instructed to focus on the story. Given that the story was designed for a younger age group, adults were informed that the experiment was testing how comprehension was affected by the inclusion of different numbers of nonsense words, as are frequently encountered in children's stories.

As in Experiment 1, the second session (~5 minutes) was completed the day after the first, and involved completing the same memory tasks as in the first session (T2). The third session was completed one week after the first session, and lasted approximately 10-15 minutes. Participants completed the memory tests for a third time (T3), completed an assessment of their existing vocabulary knowledge, and filled out a questionnaire to document strategy use and technical problems. For the vocabulary assessment, participants were asked to type definitions for 13 age-appropriate items selected from the WASI-2 (Wechsler, 2011). Each item could be scored up to 2 points based on the manual guidelines, totalling a maximum score of 26.

Learning phase

As Experiment 1.

Test phase

The tasks were programmed similarly to Experiment 1, with only minor amendments to facilitate independent transition through the tasks.

Stem completion. The start of each trial was marked with “Which word began with…” in written text, followed by presentation of the spoken cue. An image of a microphone appeared to prompt participants to make a spoken response. Participants were asked to speak “pass” or “don’t know” if they could not attempt an answer, and click the “Next trial” button when they were ready to move on (timeout after 1 minute). A small percentage of trials (1.01%) could not be scored due to technical problems with the recordings, disruption, or the recording cutting off half-way through a response. These trials were assumed to be missing at random, and entered into analyses as missing data.

Form recognition. For each form recognition trial, an image of headphones was presented on screen, and participants heard the correct pseudoword and its phonological foil. The numbers “1” and “2” appeared on screen as each answer option was presented, and participants were asked to use the corresponding number keys to select whether the first or second option had been presented in the story. The trial timed out after 5 seconds, and the next one began after a 1 second interval.

Picture-form recognition. For each trial, a picture was presented on screen, and participants heard the correctly associated pseudoword and an incorrect (learned) pseudoword. As above, participants used the number keys to select whether the first or second pseudoword was the correct match. There was no timeout for this task.

Analyses

As Experiment 1.

Experiment 2 Results

Stem completion

Participants recalled .11 (*SD* = .31) of the pseudowords in the first session. Recall was lowest at the first test point ($\beta = 0.34$, *SE* = 0.03, *Z* = 9.99, $p < .001$), and continued to improve between T2 (*M* = .16, *SD* = .37) and T3 (*M* = .25, *SD* = .45; $\beta = 0.38$, *SE* = 0.05, *Z* = 7.90, $p < .001$; Figure 2A).

Pseudowords with many neighbours showed the highest recall (*M* = .23, *SD* = .42), relative to pseudowords with no neighbours (*M* = .14, *SD* = .35) or one neighbour (*M* = .15, *SD* = .36). However, this influence of phonological neighbours was not statistically significant (*p*s > .2). The influence of neighbours appeared to strengthen slightly with consolidation (Figure 2A), but this interaction was not statistically significant (β =0.07, *SE* = 0.04, *Z* = 1.81, *p* = .071). Vocabulary ability remained a strong predictor of performance ($\beta = 0.33$, *SE* = 0.12, *Z* = 2.86, *p* = .004), but did not interact with any other variable (Table S4).

Form recognition

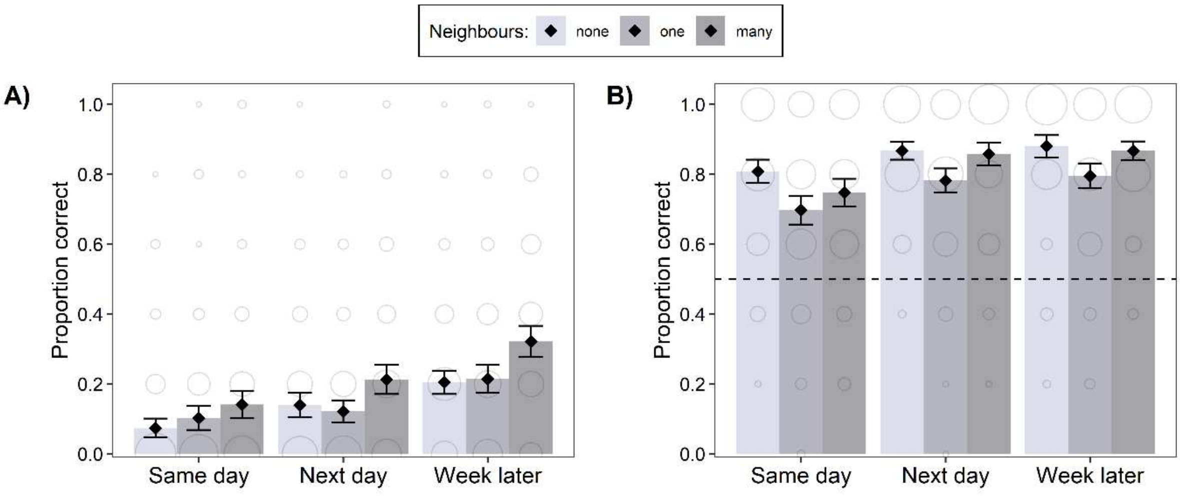
Despite low levels of recall, participants could successfully recognise the correct word-forms at above chance levels after learning (*M* = .75, *SD* = .43; *t*(124) = 18.41, $p < .001$).

Performance was lowest at this first test relative to the subsequent tests (β = 0.21, *SE* = 0.02, *Z* = 8.65, *p* < .001), but did not significantly change between the day (*M* = .84, *SD* = .37) and week (*M* = .85, *SD* = .36) tests (*p* = .251). Vocabulary ability predicted overall performance (β = 0.25, *SE* = 0.07, *Z* = 3.62, *p* < .001).

As with children's performance on this task, there was some indication that neighbours interfered with new learning (Figure 2B): performance was higher for pseudowords without neighbours (*M* = .85, *SD* = .36) than pseudowords with one (*M* = .76, *SD* = .43) or many neighbours (*M* = .82, *SD* = .38), but this was not statistically significant (β = -0.16, *SE* = 0.09, *Z* = -1.82, *p* = .069). The presence of one neighbour appeared to be more detrimental than many neighbours, particularly at later time points, but neither the main contrast (β = 0.29, *SE* = 0.15, *Z* = 1.91, *p* = .057) nor interaction (β = 0.05, *SE* =0.03, *Z* = 1.87, *p* = .061) were statistically significant (Table S5).

Figure 2

Proportion of items correct per condition in the A) Stem Completion, and B) Form Recognition tasks in Experiment 2.



Note. Error bars mark 95% confidence intervals. Bubble size indicates the proportion of observations at each level of accuracy. The dashed line in panel B marks chance-level performance.

Picture-form recognition

Participants could also identify the correct pseudoword meanings at above chance levels after learning (*M* = .81, *SD* = .39; *t*(124) = 26.85, $p < .001$). For this task, performance remained stable over time (*p*s > .29). Only vocabulary ability predicted performance (β = 0.26, *SE* = 0.09, *Z* = 2.88, *p* = .004), whereas there was no influence of phonological neighbours (*p*s > .54) and no further interactions (Table S6; Figure S2).

Experiment 2 Discussion

In Experiment 2, we tested whether adults were influenced by prior lexical knowledge when learning words encountered incidentally in spoken and illustrated stories, using a near-identical experimental protocol to Experiment 1 conducted with children (differing only in the online presentation of the task, and sample-appropriate background measures). As in Experiment 1, adults improved in their recall of pseudowords with opportunities for consolidation. However, without orthographic support (cf. Experiment S1), adults were also less able to benefit from local prior knowledge in recalling the pseudowords. Bolstering this, similar to the children, they showed a slight tendency towards interference from phonological neighbours in recognising the new word-forms, although this was not statistically significant.

Exploratory analyses

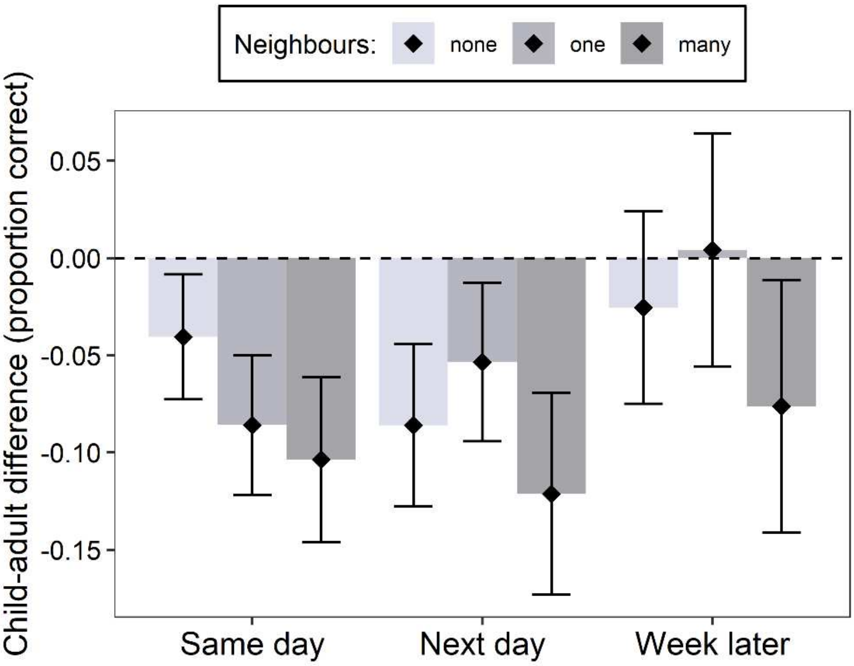
In James et al. (2019), an exploratory analysis showed that children benefited more from opportunities for offline consolidation than adults. The benefit for offline consolidation was greatest for the pseudowords without neighbours, leaving an overall prior knowledge benefit for the adults only. We conducted an additional analysis to explore whether this is similarly the case for pseudowords encountered incidentally in illustrated stories. For example, it may be that children's superior offline consolidation is diminished under such low levels of initial learning, and/or that adults' global prior knowledge allows them to benefit more from the semantic context compared to children (Henderson & James, 2018; Horst, 2013). We compared children's and adults' recall of the pseudowords by analysing the stem completion data from each experiment. Test session, neighbour condition, and experimental group (children vs. adults) were entered as fixed effects into the model, alongside their corresponding interactions (Table S7). As would be expected, adults performed significantly better than the children ($\beta = 0.57$, $SE = 0.11$, $Z = 5.17$, $p < .001$). However, this difference interacted with test session for both contrasts (*T1 vs T2&3*: $\beta = -0.15$, $SE = 0.03$, $Z = -4.49$, $p < .001$; *T2 vs T3*: $\beta = -0.22$, $SE = 0.04$, $Z = -5.22$, $p < .001$). To better understand the nature of this interaction, we used the *emmeans* package (Lenth et al., 2018) to compare children and adults in each test session. Adults recalled significantly more pseudowords than children at the first ($\beta = 1.77$, $SE = 0.28$, $Z = 6.25$, $p < .001$) and second ($\beta = 1.28$, $SE = 0.25$, $Z = 5.16$, $p < .001$) test sessions. The difference was smaller at the final test session, and was not statistically significant ($\beta = 0.40$, $SE = 0.22$, $Z = 1.82$, $p = .069$). That is, as in James et al. (2019), children seemed to improve more with consolidation than adults, gradually closing the gap in performance across the week. Note that while adults performed higher on average, they were still far from task ceiling at the final test point ($M = .25$, $SD = .43$).

Also in line with James et al. (2019), there was a three-way interaction between group, test session (*T1 vs T2&3*) and neighbour condition (*none vs one&many;* $\beta = -0.05$, $SE = 0.02$, $Z = -2.28$, $p = .023$), which we further explored using pairwise comparisons for each neighbour condition separately for each session (Figure 3). In the first two test sessions, adults were outperforming children in all three neighbour conditions (all *p*s < .03). However, by the third test session, adults only retained an advantage over children in the many neighbour condition ($\beta = 0.67$, $SE = 0.31$, $Z = 2.19$, $p = .029$), whereas the group differences in recall for the no ($\beta = 0.49$, $SE = 0.32$, $Z = 1.54$, $p = .124$) and one ($\beta = 0.05$, $SE = 0.30$, $Z = 0.18$, $p = .856$) neighbour conditions were much smaller and not statistically significant. That is, the three-way interaction appeared to be primarily driven by the many neighbour condition retaining its benefit in adults.

This pattern is consistent with previous suggestions that items with limited connections to local prior knowledge improve most with consolidation, and that children are better able to capitalise upon offline mechanisms to support new learning.

Figure 3

Difference in Stem Completion performance between children and adults in each condition.



Note. Error bars mark 95% confidence intervals. The dashed line marks no difference in performance between children and adults, with negative values depicting lower performance for children.

General Discussion

We examined the extent to which children and adults use prior lexical knowledge to support word learning from illustrated spoken stories, by using pseudowords that varied in their number of phonological neighbours. In both experiments, participants became familiar with presented pseudowords immediately after hearing the story, and showed improvements in their recall of the new word-forms across the week. While previous studies have shown a consistent benefit of phonological neighbours in supporting intentional word learning (Hoover et al., 2010; James et al., 2019; Storkel et al., 2006), we did not find such benefits during incidental word learning from stories. In an exploratory analysis however, we found that adults retained superior recall of many-neighbour items relative to children one week after learning, whereas children became as good as adults at recalling pseudowords with limited connections to prior knowledge. In sum, we consider that the benefits for local prior knowledge are at least more fragile under incidental learning conditions compared to explicit training, but that children's superior ability to acquire phonologically distinct items holds across learning contexts.

Learning and consolidating new words encountered in stories

In both experiments, participants showed above-chance performance in recognising the pseudowords and their referents encountered in the stories. This learning occurred despite no explicit instruction to learn the new words, and with relatively few exposures to each item. While producing the new word-forms was much more challenging—particularly for children—both groups showed improvements across the course of the week. This improvement is consistent with a role for consolidation processes in strengthening new representations in longer-term storage, likely resulting from sleep-associated processes (e.g., Dumay & Gaskell, 2007; Henderson et al., 2012) as well as opportunities for retrieval practice during each test session (Roediger & Karpicke, 2006). It should also be noted the repeated tests in the present study involved re-exposure to the word-forms during the recognition tasks, which may have additionally supported improvements in word-form knowledge across the week. However, children also showed offline improvements in their ability to map the words to the correct referent despite no further input, suggesting that re-exposures cannot entirely account for the improvements seen. Furthermore, studies that have directly examined consolidation with and without intervening tests have not found significant differences in the magnitude of overnight improvements (Henderson et al., 2013), and repeated tests separated by a period of sleep show greater benefits than those separated by an equivalent period of wake (e.g., James, Gaskell, et al., 2020). Thus, while retrieval practice and re-exposure may contribute to the improvements seen here, offline consolidation mechanisms are likely playing a key role.

It is also important to acknowledge that we cannot isolate independent contributions of the spoken narrative and illustrations to supporting learning and consolidation in the present experiments. Both elements provided broader contextual information to the pseudowords and their referents than have been included in explicit training studies, and we did not set out to determine their independent contributions to supporting new word acquisition. Indeed, learning in this context may be supported by distributional statistics that aid in associating pseudowords with novel referents across scenes, as well as the semantic information provided in the narrative (cf. studies of cross-situational word learning; Yu & Smith, 2007). How the narrative context and illustrations each influence new word memory remains an interesting question for future research, but the current discussion focuses on how the learning context provided by the two together (i.e., in an incidental learning setting) may affect the extent to which learners benefit from prior lexical knowledge in new learning.

The influence of local prior knowledge on learning words from stories

Several studies have shown that words with phonological neighbours in the learner's language are learned more easily (James et al., 2019; Meade et al., 2018; Storkel et al., 2013), particularly if those neighbours are highly frequent (Vitevitch et al., 2014). However, these studies explicitly instructed participants to learn the new words, providing opportunities for them to make strategic links between new and known words during learning. Here, we asked whether learners benefit from local prior knowledge when new words are encountered incidentally when listening to stories presented with illustrations. By presenting the stories in this format and emphasising comprehension in the task instructions, we aimed to minimise opportunities for learners to make strategic comparisons to known words as memory aids. Embedding the pseudowords in stories also provides a richer semantic context than learning the items in isolation, although neighbour benefits have previously been observed for words encountered in stories when participants were aware of the learning task (Han et al., 2019; Hoover et al., 2010; Storkel et al., 2006; Storkel & Hoover, 2011). The focus of this study was on accessing prior knowledge at the *lexical* level, but broader *semantic* connections to prior knowledge may make substantial contributions to word learning from stories that were not thoroughly examined in these experiments (cf. Henderson & James, 2018).

Under these incidental learning conditions, the presence of phonological neighbours did not significantly benefit learners, suggesting that activation of local prior knowledge is at least weaker during incidental word learning. It remains possible that the scope for identifying neighbour benefits in the recall task was limited by the very low levels of performance (at least in Experiment 1)—likely attributable to the high number of pseudowords we incorporated into the story learning context. As such, it is plausible that neighbour benefits would re-emerge with more exposures to the items and fewer demands on learning overall. However, it seems that with limited opportunities or resources to make explicit connections with prior knowledge during learning, the neighbour benefit is less robust. This highlights the importance of supporting incidental vocabulary learning with explicit instruction in educational settings, providing learners with opportunities to draw connections with words they already know. Repetition of stories may also increase the cognitive resources available for word learning (Horst, 2013), and provide increased opportunities for strengthening connections with prior knowledge either implicitly or explicitly.

It is also important to consider that participants were not at floor levels of performance at the later test points, leaving more variability in performance that could have allowed for neighbour-related differences to emerge. We had predicted that—under these conditions of weak neighbour benefits during learning—local prior knowledge would play a supportive role during consolidation and benefit recall at later test points. This prediction was based on the enhanced opportunities for the pseudowords to interact with the learner's global prior knowledge during consolidation processes, as described by the CLS model. The data showed limited evidence in support of this hypothesis: the influence of phonological neighbours did not change across the course of the week in either of the experiments (nor Experiment S1). Broadly speaking then, our conclusions regarding the role of local prior knowledge in the CLS model seem to support those of James et al. (2019): where neighbour-related benefits are observed (i.e., following explicit training, and for our supplementary experiment incorporating orthography), they emerge early and do not require offline processes. We find here that offline processes have no further role to play in capitalising upon neighbour benefits even when those initial benefits are weak.

While children were not supported by local prior knowledge in recalling the pseudowords, they still appeared to access it during the experiment: there was a significant effect of neighbour condition in recognition of the new word-forms. However, this effect was in the opposite direction to previous findings: children showed *poorer* recognition performance for pseudowords with one/more neighbours than pseudowords without neighbours. A similar effect was observed for adults, but was not statistically significant. In many respects this seemingly conflicting result reflects the broader word recognition and production literature: real words are *recognised* more quickly if they have few competing neighbours (e.g., Metsala, 1997) but are *produced* more accurately if they have many neighbours (e.g., German & Newman, 2004). However, in an experiment using identical pseudowords and test tasks to the present study, children showed a benefit from neighbours in *both* recall and recognition tasks (James et al., 2019; Experiment 3). We suggest that the divergent findings likely reflect the learning context in addition to the test demands: perhaps drawing explicit attention to similarities and differences to known words enables individuals to benefit in forming a new representation, whereas implicit activation of word-form similarities otherwise causes interference. Speaking to this explanation, Swingley and Aslin (2007) showed that very young infants—who cannot yet engage explicit learning strategies—were poorer at learning phonological neighbours of words they knew (e.g., *tog,* a neighbour in *dog*) than non-neighbours (e.g., *meb*).

Despite experimental evidence of interference from phonological neighbours in young infants, other approaches to studying vocabulary acquisition have still indicated a neighbour benefit early in language development. For example, Jones and Brandt (2019) analysed communication inventory data (parental reports of word knowledge) and found that infants showed a phonological neighbourhood advantage in early language production. Similar conclusions were drawn by Siew and Vitevitch (2020), who analysed data from age of acquisition norms to show that children initially learn words from dense neighbourhoods. Given that these analyses are based on *acquired* words, it may be that—with enough exposure and repetition—the interference we observed early in the process of word learning could still lead to downstream benefits in word knowledge (cf. Mak et al., 2020, for similar interpretations of semantic diversity benefits). Within this broader context, the results again highlight a way in which additional explicit vocabulary instruction may be particularly helpful in drawing out these longer-term benefits.

Relations between local and global prior knowledge

A key motivation for these experiments was to understand how individuals with more vocabulary knowledge might use their knowledge to support their learning and/or consolidation of new words. In both experiments, global vocabulary knowledge predicted performance across tasks. However, we found no evidence that individuals with better vocabulary were more able to benefit from phonological neighbours. Thus, consistent with

James et al.'s (2019) study using explicit vocabulary instruction, phonological similarity does not appear to drive the relationship between learners' global vocabulary knowledge and new word learning.

There was also very little evidence that the relationship between existing vocabulary knowledge and memory for the pseudowords changed across test points, as had been found in earlier studies of word learning from stories (e.g., Henderson et al., 2015). In Experiment 1 only, we found that children with good vocabulary knowledge showed greater overnight improvements in the picture-form recognition task. This result perhaps indicates that the benefit of global prior knowledge in learning relates more to the semantic connections made during the story, rather than connections with existing form knowledge. However, we are cautious in over-interpreting this result given that we did not find a comparable relationship in either of the subsequent adult experiments. More broadly, it may be that the challenging nature of our learning task placed such strong demands on vocabulary ability from the outset that there was no further variability to be accounted for. For comparison, children in Henderson et al.'s (2015) study were exposed to fewer pseudowords in the story, and had an additional explicit exposure to them before the story began. As such, their recall performance was higher at the first test point ($M = 0.12$) than in Experiment 1 here ($M = 0.03$). When individuals learn from contexts tailored to their ability level then, there may be greater variability during subsequent consolidation that is not already accounted for.

Developmental differences in consolidating vocabulary encountered in stories

We additionally analysed whether children and adults differed in their benefits from prior knowledge and consolidation processes. In James et al.'s (2019) previous study of explicit word learning, children showed greater improvements in their recall of pseudowords across the week than adults. These offline processes particularly improved recall of words *without* neighbours, minimising the benefits of local prior knowledge on long-term memory. For adults, offline benefits were smaller, and they retained an overall benefit for local prior knowledge. These results were suggested to reflect complementary mechanisms of offline consolidation and prior knowledge in supporting new word memory, the relative contributions of which change across development (James et al., 2017; Wilhelm et al., 2012). While adults have typically accumulated more prior knowledge that might support new learning, children show differences in neural activity during sleep that may support consolidation in the context of more limited prior knowledge.

Our exploratory analysis tested whether these relative strengths hold for incidental vocabulary learning—considering that children showed much weaker learning in this study that could limit their consolidation benefits, and that adults' superior language skills may allow them to better capitalise upon the story context to support learning. Despite these differences, children still showed greater improvements in recall across the course of the week, reducing the difference with adults' recall at each test point. At the week test, adults only retained an advantage over children for recalling pseudowords with many neighbours, suggesting that adults might still benefit from their greater amounts of prior knowledge in this condition. These findings align with recent studies demonstrating enhanced benefits of offline processes for children versus adults (e.g., Peiffer et al., 2020), often attributed to developmental differences in sleep quality (most notably in slow-wave sleep; Wilhelm et al., 2012; Wilhelm et al., 2013). Alternatively, it may be that children are better able to benefit from retrieval practice, a prediction that has not to our knowledge been directly tested. Thus, this result from our exploratory analysis highlights the importance of examining models of learning and memory in the context of development.

Conclusions

In sum, we show that the benefits of local prior knowledge connections are less robust when new words are encountered incidentally via spoken stories with illustrations, relative to previous studies of intentional word learning. The lack of clear neighbour benefit under these learning conditions suggests that previous studies of neighbourhood effects in word learning may reflect—at least in part—strategic engagement of prior knowledge during learning, and may be more constrained by task demands than previously acknowledged. These findings emphasise the importance of examining the extent to which lab-based findings generalise to paradigms that are closer to how individuals learn in real-world contexts. From an applied perspective, our findings suggest that capitalising upon prior lexical knowledge may be one route by which explicit vocabulary instruction provides additional support over incidental word learning.

References

Antony, J. W., Ferreira, C. S., Norman, K. A., & Wimber, M. (2017). Retrieval as a fast route to memory consolidation. *Trends in cognitive sciences*, *21*(8), 573-576. https://doi.org/10.1016/j.tics.2017.05.001

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*(1), 388-407. https://doi.org/10.3758/s13428-019-01237-x

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1996). The CELEX lexical database (cd-rom).

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English lexicon project. *Behavior research methods*, *39*(3), 445-459. https://doi.org/10.3758/BF03193014

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1-48. https://doi.org/doi:10.18637/jss.v067.i01

Brown, H., Weighall, A., Henderson, L. M., & Gaskell, M. G. (2012). Enhanced recognition and recall of new words in 7-and 12-year-olds following a period of offline consolidation. *Journal of experimental child psychology*, *112*(1), 56-72. http://ac.els-cdn.com/S0022096511002670/1-s2.0-S0022096511002670-main.pdf?_tid=dc29264e-190c-11e6-88a6-00000aab0f01&acdnat=1463145550_6de09c0028872404268c34402b5cd0a8

Davis, C. J. (2005). N-Watch: A program for deriving neighborhood size and other psycholinguistic statistics. *Behavior research methods*, *37*(1), 65-70. https://doi.org/10.3758/BF03206399

Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning: neural and behavioural evidence. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *364*(1536), 3773-3800. https://doi.org/10.1098/rstb.2009.0111

Dawson, N., Hsiao, Y., Wei Ming Tan, A., Banerji, N., & Nation, K. (2021). Features of lexical richness in children's books: Comparisons with child-directed speech. *Language Development Research.*

Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science*, *18*(1), 35-39. https://doi.org/10.1111/j.1467-9280.2007.01845.x

Flack, Z. M., Field, A. P., & Horst, J. S. (2018). The effects of shared storybook reading on word learning: A meta-analysis. *Developmental Psychology*, *54*(7), 1334. https://doi.org/10.1037/dev0000512

German, D. J., & Newman, R. S. (2004). The impact of lexical factors on children's word-finding errors. *Journal of Speech, Language, and Hearing Research*, *47*(3), 624-636. https://doi.org/10.1044/1092-4388(2004/048)

Han, M. K., Storkel, H., & Bontempo, D. E. (2019). The effect of neighborhood density on children's word learning in noise. *Journal of child language*, *46*(1), 153-169. https://doi.org/10.1017/S0305000918000284

Havas, V., Taylor, J., Vaquero, L., de Diego-Balaguer, R., Rodríguez-Fornells, A., & Davis, M. H. (2018). Semantic and phonological schema influence spoken word learning and overnight consolidation. *The Quarterly Journal of Experimental Psychology*, *71*(6), 1469-1481. https://doi.org/10.1080/17470218.2017.1329325

Henderson, L. M., Devine, K., Weighall, A., & Gaskell, G. (2015). When the daffodat flew to the intergalactic zoo: Off-line consolidation is critical for word learning from stories. *Developmental Psychology*, *51*(3), 406. https://doi.org/10.1037/a0038786

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. https://doi.org/10.1016/j.jecp.2017.09.017

Henderson, L. M., Weighall, A., & Gaskell, G. (2013). Learning new vocabulary during childhood: Effects of semantic training on lexical consolidation and integration. *Journal of experimental child psychology*, *116*(3), 572-592. https://doi.org/10.1016/j.jecp.2013.07.004

Henderson, L. M., Weighall, A. R., Brown, H., & Gaskell, M. G. (2012). Consolidation of vocabulary is associated with sleep in children. *Developmental Science*, *15*(5), 674-687. https://doi.org/10.1111/j.1467-7687.2012.01172.x

Himmer, L., Müller, E., Gais, S., & Schönauer, M. (2017). Sleep-mediated memory consolidation depends on the level of integration at encoding. *Neurobiology of Learning and Memory*, *137*, 101-106. https://doi.org/10.1016/j.nlm.2016.11.019

Hoover, J. R., Storkel, H. L., & Hogan, T. P. (2010). A cross-sectional comparison of the effects of phonotactic probability and neighborhood density on word learning by preschool children. *Journal of Memory and Language*, *63*(1), 100-116. https://doi.org/10.1016/j.jml.2010.02.003

Horst, J. S. (2013). Context and repetition in word learning. *Frontiers in Psychology*, *4*, 149. https://doi.org/10.3389/fpsyg.2013.00149

Horváth, K., Myers, K., Foster, R., & Plunkett, K. (2015). Napping facilitates word learning in early lexical development. *Journal of Sleep Research*, *24*(5), 503-509. https://doi.org/10.1111/jsr.12306

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*, e12776. https://doi.org/10.1111/desc.12776

James, E., Gaskell, M. G., & Henderson, L. M. (2020). Sleep-dependent consolidation in children with comprehension and vocabulary weaknesses: it'll be alright on the night? *Journal of Child Psychology and Psychiatry*, *61*(10), 1104-1115. https://doi.org/10.1111/jcpp.13253

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. https://doi.org/10.1016/j.neubiorev.2017.01.054

James, E., Koutraki, Y. G., & Tickle, H. (2020). Sleep-associated consolidation in app-based language learning. *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society* 1178-1184.

Jones, S. D., & Brandt, S. (2019). Do children really acquire dense neighbourhoods? *Journal of child language*, *46*(6), 1260-1273. https://doi.org/10.1017/S0305000919000473

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. https://doi.org/10.1016/j.tics.2016.05.004

Lenth, R., Singmann, H., & Love, J. (2018). Emmeans: Estimated marginal means, aka least-squares means. *R package version*, *1*(1).

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

Mak, M. H., Hsiao, Y., & Nation, K. (2020). Anchoring and contextual variation in the early stages of incidental word learning during reading. https://doi.org/10.31219/osf.io/kf96e

Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND: Cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PloS one*, *7*(8), e43230. https://doi.org/10.1371/journal.pone.0043230

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314-324. https://doi.org/10.3758/s13428-011-0168-7

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. https://doi.org/10.1037/a0033812

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. https://doi.org/10.1037/0033-295X.102.3.419

McGregor, K. K., Licandro, U., Arenas, R., Eden, N., Stiles, D., Bean, A., & Walker, E. (2013). Why words are hard for adults with developmental language impairments. *Journal of Speech, Language, and Hearing Research*, *56*(6), 1845-1856. https://doi.org/10.1044/1092-4388(2013/12-0233)

Meade, G., Midgley, K. J., Dijkstra, T., & Holcomb, P. J. (2018). Cross-language neighborhood effects in learners indicative of an integrated lexicon. *Journal of Cognitive Neuroscience*, *30*(1), 70-85. https://doi.org/10.1162/jocn_a_01184

Metsala, J. L. (1997). An examination of word frequency and neighborhood density in the development of spoken-word recognition. *Memory & cognition*, *25*(1), 47-56. https://doi.org/10.3758/BF03197284

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. https://doi.org/10.1371/journal.pone.0152489

Montag, J. L., Jones, M. N., & Smith, L. B. (2015). The words children hear: Picture books and the statistics for language learning. *Psychological Science*, *26*(9), 1489-1496. https://doi.org/10.1177/0956797615594361

Nagy, W. E., & Herman, P. A. (1987). Breadth and Depth of Vocabulary Knowledge: Implications for Acquisition and Instruction. In M. G. McKeown & C. M. E (Eds.), *The Nature of Vocabulary Acquisition* (pp. 19-35).

Peiffer, A., Brichet, M., De Tiège, X., Peigneux, P., & Urbain, C. (2020). The power of children's sleep-Improved declarative memory consolidation in children compared with adults. *Scientific Reports*, *10*(1), 9979. https://doi.org/10.1038/s41598-020-66880-3

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, *124*(3), 372. https://doi.org/10.1037/0033-2909.124.3.372

Roediger, H. L. I., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249-255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Sénéchal, M., Thomas, E., & Monker, J.-A. (1995). Individual differences in 4-year-old children's acquisition of vocabulary during storybook reading. *Journal of Educational Psychology*, *87*(2), 218. https://doi.org/10.1037/0022-0663.87.2.218

Siew, C. S., & Vitevitch, M. S. (2020). An investigation of network growth principles in the phonological language network. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0000876

Sternberg, R. J. (1987). Most Vocabulary is Learned From Context. In M. G. McKeown & C. M. E (Eds.), *The Nature of Vocabulary Acquisition* (pp. 89-105).

Storkel, H. L. (2009). Developmental differences in the effects of phonological, lexical and semantic variables on word learning by infants. *Journal of child language*, *36*(02), 291-321. https://doi.org/10.1017/S030500090800891X

Storkel, H. L., Armbrüster, J., & Hogan, T. P. (2006). Differentiating phonotactic probability and neighborhood density in adult word learning. *Journal of Speech, Language, and Hearing Research*, *49*(6), 1175-1192. https://doi.org/10.1044/1092-4388(2006/085)

Storkel, H. L., Bontempo, D. E., Aschenbrenner, A. J., Maekawa, J., & Lee, S.-Y. (2013). The effect of incremental changes in phonotactic probability and neighborhood density on word learning by preschool children. *Journal of Speech, Language, and Hearing Research*, *56*(5), 1689-1700. https://doi.org/10.1044/1092-4388(2013/12-0245)

Storkel, H. L., & Hoover, J. R. (2011). The influence of part-word phonotactic probability/neighborhood density on word learning by preschool children varying in expressive vocabulary. *Journal of child language*, *38*(3), 628-643. https://doi.org/10.1017/S0305000910000176

Storkel, H. L., & Lee, S.-Y. (2011). The independent effects of phonotactic probability and neighbourhood density on lexical acquisition by preschool children. *Language and Cognitive Processes*, *26*(2), 191-211. https://doi.org/10.1080/01690961003787609

Swingley, D., & Aslin, R. N. (2007). Lexical competition in young children's word learning. *Cognitive Psychology*, *54*(2), 99-132. https://doi.org/10.1016/j.cogpsych.2006.05.001

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

Ullman, M. T. (2015). The declarative/procedural model. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 135-158). Routledge.

Ullman, M. T. (2016). The declarative/procedural model: a neurobiological model of language learning, knowledge, and use. In *Neurobiology of language* (pp. 953-968). Elsevier. https://doi.org/10.1016/B978-0-12-407794-2.00076-6

van der Kleij, S. W., Rispens, J. E., & Scheper, A. R. (2016). The effect of phonotactic probability and neighbourhood density on pseudoword learning in 6-and 7-year-old children. *First Language*, *36*(2), 93-108. https://doi.org/10.1177/0142723715626064

Vitevitch, M. S., Storkel, H. L., Francisco, A. C., Evans, K. J., & Goldstein, R. (2014). The influence of known-word frequency on the acquisition of new neighbours in adults: Evidence for exemplar representations in word learning. *Language, cognition and neuroscience*, *29*(10), 1311-1316. https://doi.org/10.1080/23273798.2014.912342

Wechsler, D. (2011). *WASI-II: Wechsler Abbreviated Scale of Intelligence*. Pearson.

Weighall, A., Henderson, L., Barr, D., Cairney, S., & Gaskell, M. (2017). Eye-tracking the time-course of novel word learning and lexical competition in adults and children. *Brain and Language*, *167*, 13-27. https://doi.org/10.1016/j.bandl.2016.07.010

Wickham, H. (2016). *ggplot2: elegant graphics for data analysis*. Springer.

Wilhelm, I., Prehn-Kristensen, A., & Born, J. (2012). Sleep-dependent memory consolidation–What can be learnt from children? *Neuroscience & Biobehavioral Reviews*, *36*(7), 1718-1728. https://doi.org/10.1016/j.neubiorev.2012.03.002

Wilhelm, I., Rose, M., Imhof, K. I., Rasch, B., Büchel, C., & Born, J. (2013). The sleeping child outplays the adult's capacity to convert implicit into explicit knowledge. *Nature neuroscience*, *16*(4), 391-393. https://doi.org/10.1038/nn.3343

Williams, S. E., & Horst, J. S. (2014). Goodnight book: Sleep consolidation improves word learning via storybooks. *Frontiers in Psychology*, *5*(184), 1-12.

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, *18*(5), 414-420.

Supplementary Materials

The first part of this supplementary materials file includes the full statistical models for each analysis presented in the manuscript, and figures for the picture-form recognition tasks.

The second part (p11+) details the methods, results, and a brief discussion for Experiment S1: an initial experiment that was conducted with adults, which also included written presentation of the stories.

All stimuli, experimental tasks, raw data, and analysis scripts can be accessed via the Open Science Framework: https://osf.io/stx6q

Table S1

Final model for Experiment 1 stem completion accuracy

Predictor	β	*SE*	Z	*p*
(Intercept)	-3.44	0.30	-11.50	<.001
delay1	**0.68**	**0.07**	**9.18**	**<.001**
delay2	**0.84**	**0.08**	**10.63**	**<.001**
neighb1	0.29	0.20	1.40	.160
neighb2	0.14	0.34	0.40	.688
vocab	**0.69**	**0.16**	**4.24**	**<.001**
delay1:neighb1	0.02	0.06	0.32	.750
delay2:neighb1	-0.06	0.06	-0.98	.328
delay1:neighb2	-0.07	0.08	-0.87	.386
delay2:neighb2	-0.03	0.09	-0.30	.766
delay1:vocab	-0.01	0.07	-0.10	.920
delay2:vocab	0.00	0.08	-0.05	.963
neighb1:vocab	-0.08	0.09	-0.93	.350
neighb2:vocab	0.09	0.14	0.64	.519
delay1:neighb1:vocab	0.08	0.05	1.61	.107
delay2:neighb1:vocab	0.02	0.06	0.38	.707
delay1:neighb2:vocab	-0.02	0.09	-0.28	.783
delay2:neighb2:vocab	0.03	0.09	0.37	.712

Note. Model formed from 4320 observations, from 97 participants and 15 items. Factorial predictors used the following contrasts: *delay1* compared T1 vs. T2&3; *delay2* compared T2 vs. T3; *neighb1* contrast compared no vs. one&many; *neighb2* compared one vs. many. The model incorporated by-participant random slopes for the effect of neighbour condition, and by-item slopes for vocabulary.

Table S2

Final model for Experiment 1 form recognition accuracy

Predictor	β	*SE*	Z	*p*
(Intercept)	1.16	0.10	11.84	<.001
delay1	**0.23**	**0.03**	**8.96**	**<.001**
delay2	**0.17**	**0.05**	**3.58**	**<.001**
neighb1	**-0.12**	**0.06**	**-2.10**	**.036**
neighb2	-0.03	0.10	-0.27	.789
vocab	**0.30**	**0.08**	**3.88**	**<.001**
delay1:neighb1	0.03	0.02	1.71	.088
delay2:neighb1	0.04	0.03	1.30	.194
delay1:neighb2	0.01	0.03	0.35	.724
delay2:neighb2	0.08	0.06	1.43	.153
neighb1:vocab	-0.07	0.04	-1.82	.069
neighb2:vocab	-0.01	0.06	-0.12	.906
delay1:vocab	0.04	0.03	1.42	.156
delay2:vocab	0.06	0.05	1.30	.192

Note. Model formed from 4305 observations, from 97 participants and 15 items. Factorial predictors used the following contrasts: *delay1* compared T1 vs. T2&3; *delay2* compared T2 vs. T3; *neighb1* contrast compared no vs. one&many; *neighb2* compared one vs. many. Three-way interactions were pruned during analysis with no reduction in model fit (χ^2 = 0.53, *p* = .970). The model incorporated by-item random slopes for vocabulary.

Table S3

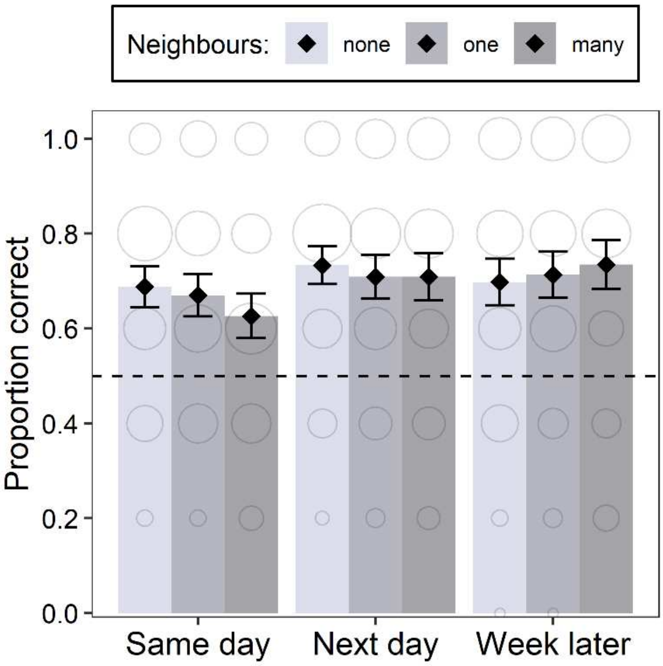
Final model for Experiment 1 picture-form recognition accuracy

Predictor	β	*SE*	Z	*p*
(Intercept)	0.97	0.09	10.90	<.001
delay1	**0.10**	**0.03**	**4.08**	**<.001**
delay2	-0.01	0.05	-0.17	.862
neighb1	0.00	0.05	-0.10	.920
neighb2	0.00	0.09	0.02	.987
vocab	**0.35**	**0.09**	**3.96**	**<.001**
delay1:neighb1	0.03	0.02	1.44	.149
delay2:neighb1	0.05	0.03	1.47	.142
delay1:neighb2	0.05	0.03	1.63	.103
delay2:neighb2	0.03	0.06	0.57	.571
neighb1:vocab	0.04	0.05	0.89	.376
neighb2:vocab	0.14	0.09	1.61	.108
delay1:vocab	**0.09**	**0.03**	**3.53**	**<.001**
delay2:vocab	0.02	0.05	0.47	.639

Note. Model formed from 4091 observations, from 92 participants and 15 items. Factorial predictors used the following contrasts: *delay1* compared T1 vs. T2&3; *delay2* compared T2 vs. T3; *neighb1* contrast compared no vs. one&many; *neighb2* compared one vs. many. Three-way interactions were pruned during analysis with no reduction in model fit ($\chi^2 = 1.90$, $p = .754$). The model incorporated by-participant random slopes for neighbour condition, and by-item slopes for vocabulary.

Figure S1

Accuracy in Experiment 1 picture-form recognition



Note. Bars mark mean proportion correct for each neighbour condition in each test session; error bars mark 95% confidence intervals. Bubbles are used to indicate the spread of the data: the larger the bubble, the higher the proportion of observations at that level of the dependent variable.

Table S4

Final model for Experiment 2 stem completion accuracy

Predictor	β	*SE*	Z	*p*
(Intercept)	-2.25	0.22	-10.05	<.001
delay1	**0.34**	**0.03**	**9.99**	**<.001**
delay2	**0.38**	**0.05**	**7.90**	**<.001**
neighb1	0.17	0.14	1.21	.225
neighb2	0.31	0.24	1.25	.210
vocab	**0.33**	**0.12**	**2.86**	**.004**
delay1:neighb1	-0.04	0.03	-1.44	.149
delay2:neighb1	0.03	0.04	0.80	.422
delay1:neighb2	0.07	0.04	1.81	.071
delay2:neighb2	-0.02	0.06	-0.28	.777
delay1:vocab	0.01	0.03	0.33	.742
delay2:vocab	0.03	0.05	0.54	.588
neighb1:vocab	-0.05	0.04	-1.29	.198
neighb2:vocab	0.05	0.07	0.66	.507
delay1:neighb1:vocab	0.00	0.02	0.09	.927
delay2:neighb1:vocab	0.04	0.03	1.22	.222
delay1:neighb2:vocab	0.01	0.04	0.25	.805
delay2:neighb2:vocab	-0.09	0.06	-1.60	.109

Note. Model formed from 5568 observations, from 125 participants and 15 items. Factorial predictors used the following contrasts: *delay1* compared T1 vs. T2&3; *delay2* compared T2 vs. T3; *neighb1* contrast compared no vs. one&many; *neighb2* compared one vs. many. The model incorporated by-participant random slopes for the effect of neighbour condition.

Table S5

Final model for Experiment 2 form recognition accuracy

Predictor	β	*SE*	Z	*p*
(Intercept)	1.73	0.14	12.66	<.001
delay1	**0.21**	**0.02**	**8.65**	**<.001**
delay2	0.05	0.05	1.15	.251
neighb1	-0.16	0.09	-1.82	.069
neighb2	0.29	0.15	1.90	.057
vocab	**0.25**	**0.07**	**3.62**	**<.001**
delay1:neighb1	0.01	0.02	0.80	.423
delay2:neighb1	-0.01	0.03	-0.21	.832
delay1:neighb2	0.05	0.03	1.87	.061
delay2:neighb2	0.00	0.06	0.03	.979
neighb1:vocab	-0.01	0.03	-0.28	.780
neighb2:vocab	-0.03	0.05	-0.68	.496
delay1:vocab	0.00	0.02	-0.11	.909
delay2:vocab	0.03	0.05	0.69	.491

Note. Model formed from 5625 observations, from 125 participants and 15 items. Factorial predictors used the following contrasts: *delay1* compared T1 vs. T2&3; *delay2* compared T2 vs. T3; *neighb1* contrast compared no vs. one&many; *neighb2* compared one vs. many. Three-way interactions were pruned during analysis with no reduction in model fit (χ^2 = 1.05, *p* = .902). The model incorporated by-participant random slopes for neighbour condition.

Table S6

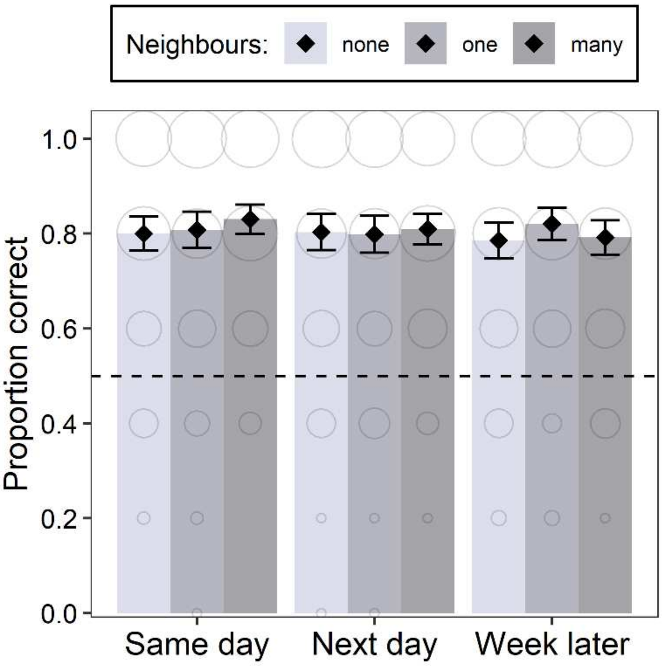
Final model for Experiment 2 picture-form recognition accuracy

Predictor	β	*SE*	Z	*p*
(Intercept)	1.79	0.15	12.07	<.001
delay1	-0.03	0.03	-1.04	.298
delay2	-0.01	0.04	-0.26	.796
neighb1	0.06	0.09	0.61	.544
neighb2	-0.02	0.16	-0.15	.879
vocab	**0.26**	**0.09**	**2.88**	**.004**
delay1:neighb1	-0.01	0.02	-0.42	.678
delay2:neighb1	0.03	0.03	0.82	.411
delay1:neighb2	-0.04	0.03	-1.29	.196
delay2:neighb2	-0.08	0.05	-1.43	.154
neighb1:vocab	0.05	0.04	1.27	.205
neighb2:vocab	0.00	0.06	0.03	.976
delay1:vocab	0.02	0.03	0.59	.558
delay2:vocab	0.02	0.04	0.37	.715

Note. Model formed from 5625 observations, from 125 participants and 15 items. Factorial predictors used the following contrasts: *delay1* compared T1 vs. T2&3; *delay2* compared T2 vs. T3; *neighb1* contrast compared no vs. one&many; *neighb2* compared one vs. many. Three-way interactions were pruned during analysis with no reduction in model fit ($\chi^2 = 0.64$, $p = .958$). The model incorporated by-participant random slopes for neighbour condition.

Figure S2

Accuracy in Experiment 2 picture-form recognition



Note. Bars mark mean proportion correct for each neighbour condition in each test session; error bars mark 95% confidence intervals. Bubbles are used to indicate the spread of the data: the larger the bubble, the higher the proportion of observations at that level of the dependent variable.

Table S7

Final model for comparison of children's (Experiment 1) and adults' (Experiment 2) stem completion accuracy.

Predictor	β	*SE*	Z	*p*
(Intercept)	-2.83	0.24	-11.61	<.001
delay1	**0.49**	**0.03**	**14.24**	**<.001**
delay2	**0.60**	**0.04**	**14.32**	**<.001**
neighb1	0.19	0.16	1.18	.237
neighb2	0.21	0.28	0.76	.448
group	**0.57**	**0.11**	**5.17**	**<.001**
delay1:neighb1	0.02	0.02	0.65	.515
delay2:neighb1	0.00	0.03	-0.11	.912
delay1:neighb2	-0.01	0.04	-0.14	.892
delay2:neighb2	-0.02	0.05	-0.43	.671
delay1:group	**-0.15**	**0.03**	**-4.49**	**<.001**
delay2:group	**-0.22**	**0.04**	**-5.22**	**<.001**
neighb1:group	0.00	0.05	-0.03	.975
neighb2:group	0.08	0.09	0.91	.363
delay1:neighb1:group	**-0.05**	**0.02**	**-2.28**	**.023**
delay2:neighb1:group	0.03	0.03	1.13	.256
delay1:neighb2:group	0.08	0.04	1.79	.073
delay2:neighb2:group	0.00	0.05	-0.08	.938

Note. Model formed from 9888 observations, from 222 participants and 15 items. Factorial predictors used the following contrasts: *delay1* compared T1 vs. T2&3; *delay2* compared T2 vs. T3; *neighb1* contrast compared no vs. one&many; *neighb2* compared one vs. many. The model incorporated by-participant random slopes for neighbour condition and by-item slopes for group.

Experiment S1

Experiment S1 was an additional experiment conducted with adults, prior to Experiment 2 presented in the main manuscript. It incorporated written language in both the exposure phase (written narrative) and test phase (orthographic presentation and typed responses). The hypotheses were identical to those for Experiment 2 in the main manuscript, and were pre-registered at https://osf.io/cdyrw.

Experiment S1 Methods

Participants

Experiment S1 was an online experiment. 130 adults were included in the analysis, and were recruited via Prolific Academic according to the same criteria as Experiment 2, except that a working microphone was not a requirement for this version of the experiment.

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

Design and procedure

The overall design and procedure was the same as Experiment 2, except that we incorporated written language throughout (detailed below).

Experimental stimuli

As Experiments 1 and 2.

Learning phase

As Experiments 1 and 2, with the addition of written text below each picture for participants to read along with the story. Our reasons for doing this were threefold: 1) to provide additional orthographic support in learning the pseudowords, given that recall had been low in Experiment 1; 2) to bring the encoding procedure in line with the written testing format for this study; and 3) to make the experiment more comparable to the adult sample in James et al.

(2019), who were presented with orthography during an explicit training regime and subsequent tests.

Test phase

Stem completion. Recall of the new word-forms was again tested using a stem completion task, but in orthographic form. Participants were provided with the written cue (first consonant and vowel) alongside the spoken cue, and were required to type their responses. Answers were scored as accurate if they read as phonologically correct.

Recognition. We administered only a single recognition task, combining the options from the form- and meaning-recognition tasks into a single 4-AFC trial for each item (as in James et al., 2019). Participants were provided with each picture and asked to choose which of four orthographically presented options its name was. The options consisted of the correct answer, a phonological foil for the correct answer, an alternative pseudoword that had been presented in the story (but was not the correct semantic mapping), and the phonological foil for the incorrect option. Participants could hear each option spoken by clicking a speaker.

Analyses

As Experiments 1 and 2.

Experiment S1 Results

Stem completion

Recall performance was higher than in Experiments 1 and 2: adults successfully recalled a mean proportion of .20 of the pseudowords (*SD* = .40) in the first session, and improved in later test sessions ($\beta = 0.06$, *SE* = 0.03, *Z* = 2.31, *p* = .021; Figure 2A). Performance continued to improve from T2 (*M* = .21, *SD* =.41) to T3 (*M* = .24, *SD* = .43; $\beta = 0.14$, *SE* = 0.05, *Z* = 2.97, *p* = .003).

In this experiment, there were benefits of both global and local prior knowledge. Vocabulary ability was a significant predictor of recall performance ($\beta = 0.48$, *SE* = 0.13, *Z* = 3.58, $p < .001$): adults with better vocabulary were better at recalling the new words. Unlike in the experiments which used only the spoken modality, pseudowords that had many neighbours were better recalled (*M* = .31, *SD* = .46) than words with only one neighbour (*M* = .19, *SD* = .39; $\beta = 0.47$, *SE* = 0.22, *Z* = 2.16, *p* = .031). However, the contrast between pseudowords with and without neighbours overall (no vs. one&many) was not significant (*p* = .19), suggesting that pseudowords with only one neighbour did not benefit from these more limited connections compared to pseudowords without neighbours (*M* = .16, *SD* = .37; Figure S3A). There was also no interaction between vocabulary ability and neighbour benefit (*p* = .18) and no evidence of a three-way interaction (pruned from model; *p* = .70), suggesting that all participants benefited from local connections to prior knowledge consistently over sessions (Table S8).

Recognition

Recognition performance was highest immediately after story exposure (*M* = .74, *SD* = .44), with performance clearly above chance (*t*(129) = 28.74, *p* < .001; Figure S3B). Performance significantly declined by the later tests (β = -0.11, *SE* = 0.02, *Z* = -4.66, *p* < .001), but the decrease in performance between the day (T2: *M* = .70, *SD* = .46) and week (T3: *M* = .68, *SD* = .47) tests was not statistically significant. Vocabulary ability was again a positive predictor of performance (β = 0.35 *SE* = 0.10, *Z* = 3.54, *p* < .001), but there was no effect of phonological neighbours or any further interactions (Table S9).

Exploratory analysis: written vs. combined modality

We conducted an additional exploratory analysis on the stem completion data to test for potential differences between Experiment S1 (combined written and spoken modality) and Experiment 2 presented in the main manuscript (spoken modality only; Table S10). Recall performance was marginally higher overall following combined presentation (β = 0.20, *SE* = 0.10, *Z* = 1.97, *p* = .049), and interacted further with test session (*delay1:* β = -0.14, *SE* = 0.02, *Z* = -6.66, *p* < .001; *delay2:* β = -0.12, *SE* = 0.03, *Z* = -3.79, *p* < .001). That is, performance differed at the first test point (*p* < .001) but recall in the spoken modality improved more with consolidation to minimise differences the following day (*p* = .09) and week (*p* = .56). The combined datasets did not show an overall neighbour effect (*neighb1*: *p* = .17; *neighb2:* β = 0.37, *SE* = .21, *Z* = 1.76, *p* = .078) and did not support an interaction between modality and neighbour influence (*p*s > .40).

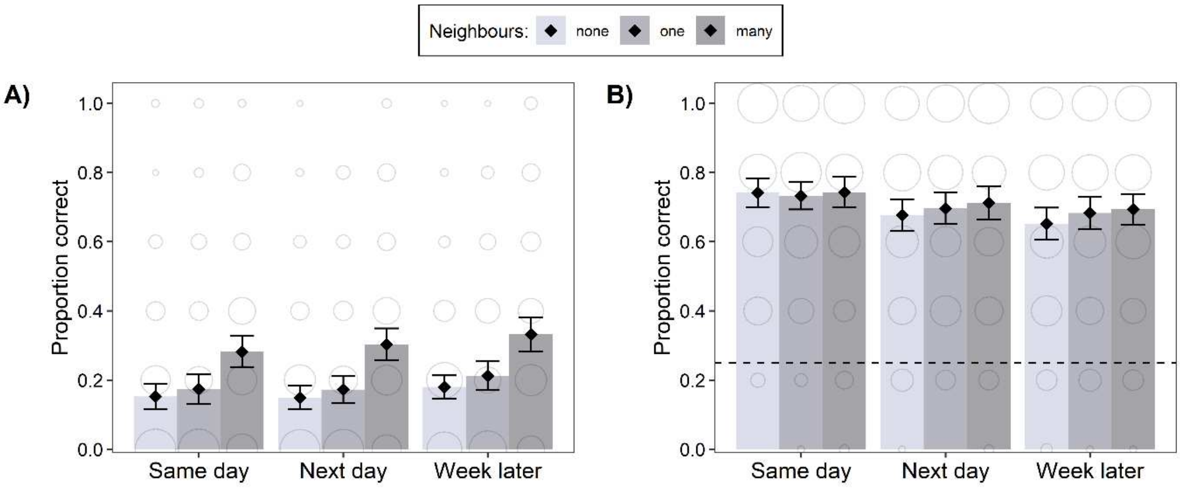
Experiment S1 Discussion

In this version of the experiment with combined written and spoken presentation, adults were able to access and benefit from local prior knowledge in learning words from stories. As when receiving explicit vocabulary instruction for the same stimuli (James et al., 2019), adult participants were better able to recall pseudowords with many neighbours in the English language than pseudowords without neighbours. This was the case for all participants regardless of vocabulary ability: again, individuals with good vocabulary learned more pseudowords overall, but they were no different in their ability to consolidate the new word-forms or benefit from phonological neighbours. Thus, adults appear more able to benefit from neighbours under written presentation conditions, compared to when the story and test tasks were presented in the spoken modality. However, an exploratory analysis comparing the two experiments did not support an interaction between modality and neighbour influence (and only weak evidence of a neighbour influence overall), suggesting that these differences may not be robust.

It is important to note that the inclusion of orthography in this experiment was not a manipulation of interest in relation to our hypotheses. Rather, it was included for ease of administering the experiment online, and for comparability to explicit training study presented in James et al. (2019) which also used the written modality. Despite not finding clear differences in the exploratory analysis, it is worth considering why neighbour benefits might have been more likely to emerge in Experiment S1. First, it may be that orthography provides an additional route to existing lexical knowledge, supporting the relevant connections at the word-level. Although a plausible contributor, it is unlikely that the inclusion of orthography is the sole driver of the neighbour benefit, given that the majority of previous studies have examined purely spoken word learning (e.g., Storkel et al., 2006), and that neighbour benefits have been observed regardless of spoken/written presentation using the same stimuli (James et al., 2019). Second, it may be that the presentation of written stories allowed participants to approach word learning more strategically. Adults can typically read at a faster pace than spoken language, providing participants with an opportunity to revisit unfamiliar words without necessarily disrupting comprehension. As such, it may be that Experiment S1 became more of an explicit learning task for some (perhaps more motivated) participants, counter to our intentions for the present study. Thus, Experiment 2 presented in the main manuscript was conducted to resolve this conflict, rendering the adult experiment more comparable to the initial experiment with children.

Figure S3

Proportion of items correct per condition in the A) Stem Completion, and B) Recognition tasks in Experiment S1.



Note. Error bars mark 95% confidence intervals. Bubble size indicates the proportion of observations at each level of accuracy. The dashed line in panel B marks chance-level performance.

Table S8

Final model for Experiment S1 stem completion accuracy

Predictor	β	*SE*	Z	*p*
(Intercept)	-1.91	0.21	-9.07	<.001
delay1	**0.06**	**0.03**	**2.31**	**.021**
delay2	**0.14**	**0.05**	**2.97**	**.003**
neighb1	0.17	0.13	1.32	.186
neighb2	**0.47**	**0.22**	**2.16**	**.031**
vocab	**0.48**	**0.13**	**3.58**	**<.001**
delay1:neighb1	0.01	0.02	0.54	.590
delay2:neighb1	0.00	0.03	0.01	.992
delay1:neighb2	0.01	0.03	0.44	.657
delay2:neighb2	-0.04	0.05	-0.67	.501
delay1:vocab	-0.02	0.03	-0.62	.538
delay2:vocab	0.03	0.05	0.65	.515
neighb1:vocab	-0.02	0.05	-0.47	.636
neighb2:vocab	-0.12	0.09	-1.34	.180

Note. Model formed from 5850 observations, from 130 participants and 15 items. Factorial predictors used the following contrasts: *delay1* compared T1 vs. T2&3; *delay2* compared T2 vs. T3; *neighb1* contrast compared no vs. one&many; *neighb2* compared one vs. many. Three-way interactions were pruned during analysis with no reduction in model fit ($\chi^2 = 2.21$, $p = .697$). The model incorporated by-participant random slopes for neighbour condition, and by-item slopes for vocabulary.

Table S9

Final model for Experiment S1 recognition accuracy

Predictor	β	*SE*	Z	*p*
(Intercept)	1.17	0.15	7.73	<.001
delay1	**-0.11**	**0.02**	**-4.66**	**<.001**
delay2	-0.06	0.04	-1.45	.147
neighb1	0.07	0.08	0.86	.388
neighb2	0.09	0.15	0.62	.536
vocab	**0.35**	**0.10**	**3.54**	**<.001**
delay1:neighb1	0.02	0.02	1.26	.208
delay2:neighb1	0.01	0.03	0.21	.833
delay1:neighb2	0.00	0.03	-0.07	.943
delay2:neighb2	-0.01	0.05	-0.20	.840
delay1:vocab	0.01	0.02	0.35	.727
delay2:vocab	0.00	0.04	0.06	.950
neighb1:vocab	0.01	0.03	0.34	.736
neighb2:vocab	0.00	0.06	0.02	.981

Note. Model formed from 5850 observations, from 130 participants and 15 items. Factorial predictors used the following contrasts: *delay1* compared T1 vs. T2&3; *delay2* compared T2 vs. T3; *neighb1* contrast compared no vs. one&many; *neighb2* compared one vs. many. Three-way interactions were pruned during analysis with no reduction in model fit ($\chi^2 = 1.18$, $p = .881$). The model incorporated by-participant random slopes for neighbour condition.

Table S10

Final model for exploratory analysis comparing Experiment S1 (combined written/spoken modality) and Experiment 2 (spoken modality only) stem completion accuracy

Predictor	β	*SE*	Z	*p*
(Intercept)	-2.07	0.19	-10.87	<.001
delay1	**0.20**	**0.02**	**9.42**	**<.001**
delay2	**0.26**	**0.03**	**8.01**	**<.001**
neighb1	0.17	0.12	1.37	.170
neighb2	0.37	0.21	1.76	.078
modality	**0.20**	**0.10**	**1.97**	**.049**
delay1:neighb1	-0.01	0.02	-0.77	.441
delay2:neighb1	0.02	0.02	0.74	.460
delay1:neighb2	0.04	0.02	1.76	.079
delay2:neighb2	-0.03	0.04	-0.79	.431
delay1:modality	**-0.14**	**0.02**	**-6.66**	**<.001**
delay2:modality	**-0.12**	**0.03**	**-3.79**	**<.001**
neighb1:modality	0.01	0.05	0.31	.754
neighb2:modality	0.07	0.08	0.83	.407
delay1:neighb1:modality	0.02	0.02	1.45	.147
delay2:neighb1:modality	-0.02	0.02	-0.80	.421
delay1:neighb2:modality	-0.03	0.02	-1.13	.257
delay2:neighb2:modality	-0.01	0.04	-0.16	.871

Note. Model formed from 11418 observations, from 255 participants and 15 items. Factorial predictors used the following contrasts: *delay1* compared T1 vs. T2&3; *delay2* compared T2 vs. T3; *neighb1* contrast compared no vs. one&many; *neighb2* compared one vs. many; *modality* compared spoken vs. combined written/spoken. The model incorporated by-participant random slopes for neighbour condition, and by-item slopes for modality.