Eye-tracking the time-course of novel word learning and lexical competition in adults and children.

Weighall, A.,[1] Henderson, L.M,[2] Barr, D.J.,[3] Cairney, S.A.[2] & Gaskell M.G.[2]

a.r.weighall@leeds.ac.uk; lisa-marie.henderson@york.ac.uk; d.barr@glasgow.ac.uk; scott.cairney@york.ac.uk; gareth.gaskell@york.ac.uk

[1]School of Psychology, University of Leeds, Leeds, UK, LS2 9JT

[2]Department of Psychology, University of York, York, UK, YO10 5DD

[3]Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, G12 8QB

Corresponding author:

Dr Anna Weighall

a.r.weighall@leeds.ac.uk

School of Psychology

 University of Leeds

Leeds, UK,

LS2 9JT

Please note that the data presented in this manuscript, including the code used to produce the eyetracking analyses, are publicly available at https://osf.io/wvy27

**Abstract**

Lexical competition is a hallmark of proficient, automatic word recognition. Previous research suggests that there is a delay before a new spoken word becomes engaged in this process, with sleep playing an important role. However, data from one method--the visual world paradigm--consistently show competition without a delay. We trained 42 adults and 40 children (aged 7-8) on novel word-object pairings, and employed this paradigm to measure the time-course of lexical competition. Fixations to novel objects upon hearing existing words (e.g., looks to the novel object biscal upon hearing "click on the biscuit") were compared to fixations on untrained objects. Novel word-object pairings learned immediately before testing and those learned the previous day exhibited significant competition effects, with stronger competition for the previous day pairings for children but not adults. Crucially, this competition effect was significantly smaller for novel than existing competitors (e.g., looks to candy upon hearing "click on the candle"), suggesting that novel items may not compete for recognition like fully-fledged lexical items, even after 24 hours. Explicit memory (cued recall) was superior for words learned the day before testing, particularly for children; this effect (but not the lexical competition effects) correlated with sleep-spindle density. Together, the results suggest that different aspects of new word learning follow different time courses: visual world competition effects can emerge swiftly, but are qualitatively different from those observed with established words, and are less reliant upon sleep. Furthermore, the findings fit with the view that word learning earlier in development is boosted by sleep to a greater degree.

# 1. Introduction

The relative ease with which we can learn new words, after very few exposures is well documented in both the developmental   (Bloom & Markson, 1998; Carey & Bartlett, 1978; Spiegel & Halberda, 2011)  and adult literature (e.g., Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Luce & Pisoni, 1998; Marslen-Wilson & Warren, 1994).  However, word learning is a multi-faceted process. When a new spoken word is learned we must not only recognise its phonological form but also develop a detailed semantic representation of its meaning and integrate both form and meaning with existing semantic and lexical networks. The phonological form may be learned swiftly.  However, developing a fully-fledged representation requires repeated exposures over time with the representation developing in richness with each encounter. When we perceive a sequence of speech, a lexical competition process takes place in order to identify the familiar words that most closely match the sequence (Gaskell & Marslen-Wilson, 2002; Luce & Pisoni, 1998; McClelland & Elman, 1986; Norris, 1994).Thus, a discerning measure of whether a newly acquired word has been integrated in the mental lexicon is its engagement in this automatic lexical competition process, which can only arise once it has been fully integrated with existing forms in the lexicon.  McMurray, Kapnoula and Gaskell (in press) provide a comprehensive discussion of the way in which lexical items can be conceptualised as pathways comprising dynamic and multi-faceted mappings between phonological, semantic and orthographic representations. On this view, competition between lexical items may well arise as a result of the increasing automaticity of activation of these pathways, leading to flexible and efficient word recognition.  Previous research with adults has suggested that a consolidation period, often associated with sleep, is required before novel spoken words can be accessed automatically and compete for recognition with existing words (Dumay & Gaskell, 2007; Lindsay & Gaskell, 2013; Tamminen et al., 2010).  These findings are well-explained by a complementary systems account of word learning in which novel words are initially learned via hippocampal mediation between the relevant neocortical regions. Sleep then provides an opportunity for hippocampal replay to support integration of the new mappings with existing knowledge in neocortical long-term memory (Davis & Gaskell, 2009; McClelland et al., 1995). A remarkably similar emergence of lexical competition after sleep has been reported in children (Henderson, Weighall, Brown & Gaskell, 2012), suggesting that the same framework can account for word learning in development.

However, recent reports of competition effects immediately after learning have sparked debate over the extent to which offline consolidation is necessary for lexical competition effects to emerge (e.g., Coutanche & Thompson-Schill, 2014; Fernandes et al., 2009; Kapnoula et al., 2014; 2015). Such immediate effects seem to emerge under certain conditions or with particular methodologies, including when training involves extensive exposure (Fernandes et al., 2009) or promotes 'co-activation' of novel and familiar words (Coutanche & Thompson-Schill, 2014; Kapnoula et al., 2014; 2015; Lindsay & Gaskell, 2013). Thus, whilst offline consolidation plays a crucial role in lexical integration and in improving automaticity (Geukes et al., 2015; Tham et al., 2015), the emergence of lexical competition likely follows a graded trajectory dependent upon factors both intrinsic and extrinsic to the learner (McMurray, Kapnoula, & Gaskell, in press, for a review).

The present study employed the visual world paradigm, which has revealed immediate competition effects in previous studies (e.g., Creel et al., 2008; Kapnoula et al., 2015-a, 2015-b; Magnuson et al., 2003). We compared performance for items learned just before testing with those learned on the previous day. Importantly, we also sought to determine whether the magnitude of any observed competition effects would be comparable to those observed for existing lexical items, and therefore whether immediate engagement in competition is indicative of rapid neocortical learning. Finally, we compared the performance of adults to that of children aged 7- to 8- years old to examine whether the adult-like pattern of performance found in previous studies is evident when a more temporally sensitive measure of lexical competition is utilised.

A substantial number of studies provide evidence for a protracted time-course of engagement in lexical competition in adults (e.g., Bakker, Takashima, van Hell, Janzen & McQueen, 2014; Dumay & Gaskell, 2003; Gaskell & Dumay, 2007; Tamminen et al., 2010). Many of these studies used the pause detection paradigm as a measure of lexical competition (Mattys & Clark, 2002). Participants were exposed to fictitious spoken novel competitors (e.g., dolpheg) that overlapped with existing words (e.g., dolphin) and made speeded judgements on the presence/absence of a 200 ms pause inserted near the point in the word at which it deviated from the new competitor (e.g., "dolph_in"). Soon after learning there was no difference in pause detection latencies for the existing words compared with matched control words for which no close competitor had been learned; however, after a delay (particularly when the delay involved sleep; Dumay & Gaskell, 2007), a lexical competition effect emerged (i.e., the existing words were responded to more slowly than the control words). A similar sleep-associated improvement was found for recall and recognition of the

novel items, consistent with the view that sleep works to strengthen as well as integrate new lexical knowledge (Schreiner & Rasch, 2016; Rasch & Born, 2013).

Children learn thousands of words with ease; hence one might predict a less protracted time course of word learning earlier in development. Nevertheless, sleep-associated lexical competition effects have also been revealed in children, suggesting that sleep facilitates lexical integration in the developing brain (Henderson et al., 2012) in a similar manner. The same pattern of delayed lexical competition is observed  when children learn real rather than fictitious words (e.g., hippocampus competing with hippopotamus), when word meanings and picture referents are provided (Henderson, Weighall, & Gaskell, 2013) and when novel words are learned more implicitly via stories  (Henderson, Weighall, Devine & Gaskell, 2015). More recently, Horvath, Myers, Foster and Plunkett (2015) taught two novel object-word pairs to 16 month-old infants, testing lexical knowledge via a preferential looking task, both prior to and following a nap or equivalent period of wake. Whilst the nap group improved after the nap, the wake group did not change.

These findings, from across development, are consistent with a complementary learning systems account of vocabulary acquisition (Davis & Gaskell., 2009; McClelland, McNaughton & O'Reilly, 1995), as well as active systems models of sleep-dependent consolidation (Born & Wilhelm, 2012; Diekelmann & Born, 2010; Rasch & Born, 2013). According to the complementary systems account, acquiring new words too quickly can disrupt memory for similar items already in long-term memory. Hence, in order to protect existing items from "catastrophic interference", new words are encoded using short-term hippocampal mediation, before a long-term neocortical memory representation is strengthened via consolidation (see Davis, Di Betta, MacDonald & Gaskell, 2009, for fMRI evidence in support of this theory). Consistent with this, the active systems model posits that slow oscillations in sleep drive the transfer of initially hippocampally mediated memory traces to neocortical sites for long-term storage. Slow oscillations comprise up-states with wake-like levels of firing activity, and down-states of neuronal silence. The up-states of slow oscillations are temporally synchronised with two key EEG events – thalamocortical spindles and sharp-wave ripples – which are proposed to signal recently learned memory reactivations from the hippocampus and facilitate integration into neocortical storage sites (e.g., Molle & Born, 2011). For example, Tamminen et al (2010) reported positive correlations between overnight increases in lexical competition and sleep spindles in adults, and between slow-wave sleep (SWS) duration and increases in recognition speed to newly learned words. Not

only does this support an active role of sleep in the consolidation of newly acquired words, it also hints at a multiplicity of method underlying different aspects of word learning.

Children sleep more than adults, display more SWS, and show increased levels of slow oscillation activity, peaking at 10-12 years (Kurth et al., 2010; Ohayon et al., 2004). In this light, it is not surprising that children have been reported to show enhanced sleep-dependent consolidation for explicit aspects of declarative memory than compared to adults (Wilhelm et al., 2013). Despite the remarkably similar time-course of lexical competition for novel words in children (e.g., Henderson et al., 2012; although cf. Brown et al., 2012), there have been important in the magnitude of the competition effects, with children showing larger effects than adults even when baseline RT is controlled (Henderson et al., 2013). Furthermore, Henderson et al (2015) reported larger overnight increases in lexical competition for children relative to adults, although such an effect appeared to be due to baseline differences in response speed which was slower in the children. Thus, whether we see enhanced sleep-dependent consolidation in childhood, remains to be determined.

The visual world paradigm (VWP) is also sensitive to lexical competition effects in adults and children, with eye movements around a visual scene closely time-locked to incoming speech (e.g., Dahan et al., 2001; Tanenhaus at al., 1995). Allopenna et al (1998) presented adult participants with four pictures on a computer display and asked them to follow spoken instructions to move the depicted objects with the mouse. The target object (e.g., beaker) was contrasted with a phonological competitor (e.g., beetle) or a phonologically unrelated distracter (e.g., carriage). Robust competition was evident in the eye-movement record as participants fixated both the target and the competitor significantly more than the distracter soon after target word onset. This indicates that lexical competition arises from the phonological overlap between the target and competitor.

Previous studies have found that the VWP is sensitive to the properties of an artificial lexicon and can detect competition between novel words after minimal training (Magnuson et al., 2003; Pirog-Revill et al. 2008). For instance, over the course of two consecutive days Magnuson et al (2003) trained adults on novel bisyllabic names, each of which was associated with a different novel shape (e.g., pibo, pibu, dibo, dibu). They found that artificial lexical items were processed incrementally in adults, such that competition effects similar to those shown by Allopenna et al. (1988) were found. This, along with similar later studies

(e.g., Creel et al., 2006, 2008), shows that the VWP used with artificial lexicons can provide a sensitive and revealing model of spoken word recognition.

Given that such competition effects can be found with little or no time between the artificial lexicon training and the VWP test, these studies might provide evidence that in fact engagement in competition does not require time or sleep as previous literature suggests. However, two points are worth noting. First, these studies have examined competition between new words rather than between new and existing words. If new words are retained initially in episodic memory then they may compete with each other more easily than they compete with existing words stored in long-term memory. In fact Magnuson et al. (2003) also tested whether neighbours of novel words in a participant's existing lexicon might influence fixations to novel word referents in their study (in effect the reverse of our current goal), but found little evidence that this was the case. A second point is that the time-course of competition effects between words in an artificial lexicon tends to be slower and extended (Creel et al., 2008), perhaps by as much as 500 ms. Magnuson et al., (2003, p. 223) argued that "artificial lexicons may be considered functionally isolated from the native lexicon when the items conform to native phonotactics and have been presented recently and frequently". Considering these properties in the light of the complementary systems account of word learning described earlier, it may be that initial competition effects found by these artificial language studies are reliant on the hippocampally mediated initial representations of the novel words, which Davis & Gaskell (2009) suggested may be relatively slow by comparison with direct neocortical mappings that well-established words can use.

A more recent study examined whether novel words trained without a referent might engage in lexical competition as assessed using the VWP. Kapnoula et al (2015-a) trained a set of 10 novel competitors using similar methods to Gaskell and colleagues, but used a specific variant of the VWP to measure the time course of lexical competition (Dahan, Magnuson, Tanenhaus et al., 2001). For this task, looks to referents of target words (e.g., net) were examined in a variety of maindifferent circumstances. In one condition, the auditory stimulus was the final stop consonant of the target word (e.g., /t/) spliced onto a separate recording of the initial consonant and vowel of the same word. This condition provided a baseline for assessing the profile of fixations to the target referent in response to the auditory stimulus. In two mismatching conditions, the initial portion of the auditory stimulus was taken from a different syllable, which could either be a word (e.g., neck) or a pseudoword (e.g., nep). In the word case, as previously established (Dahan et al. 2001), the information in

the vowel relating to the place of articulation of the final consonant (e.g., /k/) tends to strengthen the activation of the competitor word (e.g., neck) and hence reduce fixations to the target. This competition effect was diminished when the initial portion came from a pseudoword, presumably because there was no alternative lexical item to favour over the target. Most crucially, when participants were familiarised with the pseudoword (without any referent) just prior to test the competition effect was enhanced, and indistinguishable from the competition effect of a well-established word. The authors argued that this result provides strong evidence for immediate inhibition between newly learned and familiar words. The same immediate effects have been observed even when the stimuli are presented in a different voice at test, perhaps suggesting these lexical competition effects are not purely the result of episodic, as opposed to long-term, memory (Kapnoula & McMurray, 2015-b). We return to the implications of these results in the Discussion.

Several studies have demonstrated that the VWP can be successfully adapted to investigate spoken language comprehension in young children and even infants (Swingley & Aslin, 2007). The paradigm is sensitive to cohort- (i.e., onset-matching) and rhyme-(i.e., offset-matching) competitor effects of phonologically similar words (e.g., candy/candle) in typically developing English-speaking children aged around 9 years old (Desroches et al., 2006) and in children as young as 5 and 6 years old in Russian (Sekerina & Brooks, 2007). In Sekerina and Brooks (2005), participants viewed scenes containing pictures of four familiar objects and used a mouse to click on a target specified in a spoken sentence. In the cohort condition, two objects names shared three-phoneme onsets; in the noncohort condition, all object names had unique onsets. Although cohort competition persisted for approximately 1000 ms longer in children than in adults, perhaps reflecting inefficiencies in competition suppression due to immature lexical processing, both children and adults showed a similar time course of eye movements. Therefore, this paradigm enables us to investigate lexical competition for newly learned words alongside existing lexical items for which we would certainly expect to observe lexical competition, and to compare adults and children directly. Since the majority of research on spoken word recognition and word learning has focussed on adults, a systematic comparison of developmental differences in a key marker of spoken word recognition could have important theoretical implications for language development.

1.1 The present study

The present study investigated the ability of a newly learned novel word-object pairs to interact with the items already in the lexicon in both adults and children as assessed via the VWP. We taught participants novel words (e.g., biscal) with pictured referents and examined the extent to which the trained novel word-object pairs would act as cohort competitors to real words (e.g. biscuit) using fixation patterns as the dependent variable. In the light of the VWP studies using artificial lexicons, we considered it crucial to examine competition effects for recently learned words in comparison with the competition effects for known items (e.g., candy/candle). This not only enabled us to seek replication of previous work in this regard and ensure the method is sensitive to competition for existing items, but it also enabled the comparison of the magnitude of any observed effects when novel competitors have been added to the lexicon. It is conceivable that where competition effects emerge the magnitude may be smaller and/or later than for well-established lexical items. In this manner the VWP was utilised to make direct comparisons between novel and existing competitors, and evaluate the extent to which newly learned pairings enter into the competitive process in the same way that well established lexical entries do. Explicit memory for the newly learned word-forms was also examined with a cued recall task.

Based upon the previous VWP studies outlined above we expected to see an increase in looks to trained competitors relative to untrained control objects soon after learning. If overnight consolidation facilitates stronger competition then we would expect significantly more or earlier fixations to the pairings that were learned on Day 1 compared to those learned on Day 2. It was also expected that adults would show superior recall of newly learned words (e.g., biscal) trained the previous day relative to those learned immediately prior to testing (similar to previous findings, e.g., Dumay & Gaskell, 2007). Finally, in line with systems models of sleep-associated memory consolidation, any advantages observed for pairings learned the previous day should correlate with sleep parameters shown to be associated with vocabulary consolidation in previous studies (namely, sleep spindles and slow-wave sleep; Tamminen et al., 2010).

The present study also investigates the extent to which the pattern of lexical competition is similar for children aged 7-9 years and adults. Although steps have been taken towards testing the hypotheses of the complementary systems model of vocabulary acquisition from a developmental perspective, studies to date have exclusively relied on the pause detection or lexical decision paradigms (Brown et al., 2012; Henderson et al., 2012, 2013, 2015). In order to facilitate comparison with our previous studies children aged 7 – 9

years were compared with adults. This age range is of theoretical interest (Henderson et al., 2013) as the lexicon does not yet appear to be fully adult-like: At 7 years old children only show lexical competition effects for highly familiar words (e.g. Metsala, Stavrinos & Walley, 2009; Munson, Swenson & Manthei, 2005), show enhanced lexical competition effects in the pause detection task relative to adults (Henderson et al., 2013) and it is not until roughly 12 years of age that the lexicon begins to appear more adult-like, characterised by swift and automatic word recognition (Ojima, Matsuba-Kurita, Nakamura & Hagiwara, 2011). Furthermore, studies based on similar age ranges have demonstrated developmental differences in sleep associated memory consolidation, including enhanced levels of slow wave activity (Wilhlem et al., 2013, 2014).

Despite the fact that previous studies have found that children of this age demonstrate non-adult like lexical processing in some tasks, we anticipated a comparable overall time course of effects for adults and children, if our own previous findings (e.g., Henderson et al., 2012) extend to competition as assessed by the VWP. Similarly, it was expected that both adults and children would show advantages in recall for newly learned words trained the previous day (similar to previous findings e.g., Brown et al., 2012; Henderson et al., 2012; Henderson et al., 2013). However, it was also anticipated that competition effects may be larger for children than adults (Henderson et al., 2013) and that overnight changes in explicit memory and lexical competition may be more pronounced for children if sleep-dependent consolidation is enhanced in childhood (Wilhelm et al., 2013).

## 2. Method and materials

### 2.1 Participants

A total of 83 participants took part in this study: 42 adults (15 males, 27 females; mean age 24.05 years, SD=6.29 years, range 18-38 years) and 41 children (22 males, 19 females; mean age 7.9 years, SD=.6 years, range 7-9 years). Adult participants were students at Sheffield Hallam University (n= 21) and the University of York (n= 21) who completed the experiment in laboratories on their respective campus. In addition, the adult participants from the University of York slept in the sleep lab on the night after the first training session to allow for sleep EEG recordings. Children were recruited from two mainstream Primary Schools in North Yorkshire, situated in areas representing a range of socioeconomic backgrounds. Adults provided informed written consent; parents provided informed written consent and children provided verbal consent. All adult participants and parents of child

participants confirmed an absence of diagnosed learning or neurological disabilities, that they had normal or corrected to normal vision and hearing, and were native English monolingual speakers.

Participants were assessed on standardised tests of nonverbal and verbal ability to ascertain whether they represented samples with normal distributions of ability. Namely, the Matrix Reasoning and Vocabulary subtests from the Wechsler Abbreviated Scales of Intelligence (Wechsler, 1999). The Vocabulary subtest measured expressive vocabulary knowledge and required participants to produce definitions of words that increased in difficulty. Mean scores for children and adults fell within the average range (children's mean T score 43.71, SD=9.35, range 33-61; adult's mean T score 57.81, SD=9.85, range 39-74). The Matrix Reasoning subtest measured nonverbal ability and required participants to complete visually presented puzzles/sequences by selecting a missing element from an array. Again, mean standard scores for both age groups fell within the average range (children's mean T score 47.27, SD=10.04, range 31-69; adults mean T score 55.91, SD=6.73, range 41-68).

2.2 Stimuli

Novel competitors. The critical stimuli were 72 word/nonword pairs, comprising a familiar "base word" (e.g., "donkey") and a fictitious novel word competitor (e.g., "donkop") (see Appendix A). All base words were high frequency nouns, selected to be highly familiar to children aged 7-9 years old (on the basis of surveying three teachers of children in this age group), with an age-of-acquisition rating of 7.5 or less (as reported in Brown et al., 2012). The novel words were all phonemically identical to base words until the point at which the word becomes unique according to CELEX (M=4 phonemes) and we created by changing the final few phonemes of the base word after the uniqueness point (see Appendix A). Ten adults (who did not participate in the experiment) were asked to name a large pool of pictures and pictures were selected to represent base words if naming agreement was >80%. When pictures had naming agreement of <80% the pictures were changed and were re-named by the same adults.

The same 10 adults were asked to name a large pool of novel objects. Novel objects were included if they were not given a specific name by >80% of the raters and yet identified by >80% of raters as belonging to one of four categories (animal, musical instrument, plant, tool). Seventy-two novel objects were paired with novel words using the following criteria:

(1) There was no semantic overlap between the base word and the category of the novel object (e.g., the novel word "donkop" was not paired with a novel object from the 'animal' category), and (2) there was no perceptual overlap between shape or colour properties of the base word picture and the properties of the novel word object. All pictures (for base words and novel words) were sized at 200 x 200 pixels (see Supplementary Materials).

The 72 stimulus pairs were divided into three matched lists of 24 novel word-object pairs (see Appendix A). The base words in each of the three lists were matched on CELEX frequency (M=8.11, SD=8.93), n syllables, (M=2.38, SD=.49) n phonemes (M=6.35, SD=1.08) uniqueness point (obtained from CELEX expressed as number of phonemes from onset; M=4.22; SD=.95), and n phonemes after the uniqueness point (M=2.16; SD=.83). Novel objects and base word pictures in each list were also matched for visual complexity (including number of object features (parts) and number of colours) to ensure that the novel competitor objects were not more or less salient than the base word objects. Participants were trained on one list on Day 1 (creating an overnight lag between training and test) and another on Day 2 (creating a shorter same day lag), leaving the third list 'untrained'; lists were rotated around participants and conditions. Counterbalancing ensured that each novel item was used as distracter in all conditions and that each target and competitors and distracters appeared with equal frequency across all four quadrants of the screen. Further details can be found in Appendix A.

Existing competitors. A further set of 20 existing "cohort competitor" trials were included to assess the extent of cohort effects in existing words, using frequency-matched cohort competitors such as candy/candle (cf. Sekerina & Brooks, 2007). These trials were included to provide a yardstick against which to compare any novel competitor effects. Stimuli comprised 20 cohort word pairs that were closely matched on verbal and written frequency, concreteness, familiarity and imageability (MRC Psycholinguistic Database, Wilson, 1988). Pictures were selected to represent these words from www.clipart.com and all had >80% naming agreement (according to our 10 adult raters). Participants viewed one of each cohort pair (e.g., candy or candle) in either the Displayed (click on candy when candle is present) or Absent Cohort condition (click on candy when candle is not present). The design was fully counterbalanced using a Latin square rotation to ensure that each member of the cohort pair appeared as a target, and that each one appeared in each condition across the experiment, but that no participant experienced any pair more than once. The target picture was always present. In the Displayed Cohort condition one of the distracters was the phonological

competitor one was a phonologically unrelated familiar object and one was a novel object. In the Absent Cohort condition, two of the distracters were phonologically unrelated (one of which was consistently treated as the 'distracter' for the purpose of the analysis across all trials) and one was a novel object. The position of the objects was fully counterbalanced across the four quadrants of the screen.

2.3 Design

Training tasks. The training tasks combined phonics based methods used in previous studies (i.e., novel word repetition and phoneme isolation; Brown et al., 2012; Henderson et al., 2012; 2013a, 2013b) with a two-alternative forced choice (2AFC) procedure used in previous word learning/eye tracking studies (e.g., Magnuson et al., 2003). Participants were exposed to each novel word-object pair nine times during training.

In Block 1 of the training, each novel word was presented via headphones and simultaneously the corresponding novel object was displayed in the centre of a laptop screen. Participants were asked to repeat each novel word aloud. Each novel word-object pair was presented twice in a randomised order. Mean accuracy scores were near ceiling (>96%) for both groups on Days 1 and 2 (children, Day 1 mean % correct 95.88%, SD=4.5%, Day 2 97.82%, SD=2.18%; adults, Day 1 97.52%, SD=4.87%, Day 2 97.97%, SD=4.28%). Repetition accuracy was higher on Day 2 than Day 1 (t(1,82)=2.66, p<.01), but the children and adults did not differ in terms of accuracy (p>.05) and there was no significant difference between Lists (p>.05).

Following this, each novel word-object pair was presented and participants were asked to segment the initial (Block 2) and final (Block 3) sounds, as a way to draw attention to the phonological forms of the words and encourage learning. Novel word-object pairs were presented twice for initial and twice for final segmentation, in a randomised order. For initial segmentation, children and adults did not significantly differ on Day 1 (adults mean 96.98%, SD=6.04%; children's mean 95.43%, SD=5.04%, F(1,81)=1.60,p>.05) but adults outperformed children on Day 2 (adults mean 98.66%, SD=2.29%; children's mean 96.54%, SD=3.80%, F(1,81)=9.48,p<.01). For final segmentation, adults outperformed children on Day 1 (adults mean 94.10%, SD=8.17%; children's mean 85.26%, SD=1.63%, F(1,81)=9.74,p<.01) and Day 2 (adults mean 96.73%, SD=5.25%; children's mean 89.28%, SD=11.08%, F(1,81)=15.43,p<.001). Initial segmentation accuracy was significantly lower on Day 1 (mean 96.21%, SD=5.59%) than Day 2 (97.52%, SD=3.29%; t(1,82)=2.50,p<.05);

similarly, final segmentation accuracy was lower on Day 1 (mean 89.73%, SD=13.56%) than Day 2 (93.05%, SD=9.36%; t(1,82)=2.91,p<.01). There were no significant differences between Lists on either Day 1 or 2 (all F < 1; p >.05).

In the final part of training (Block 4), participants were presented with two trained novel objects and simultaneously they heard the name of one of the objects (e.g., "donkop") via headphones. Participants were asked to select the correct picture by pressing one of two buttons on the keyboard. The correct object remained on screen for a further 500ms (regardless of the participants' response) and the name of the object was repeated. This 2AFC procedure was repeated 3 times for each item. Item order was randomised for each participant via E Prime 1.2 (Schneider, Eschman, & Zuccolotto, 2002). Adults (mean % correct Day 1 95.33%, SD=4.75%, Day 2 95.46%, SD= 4.85%) performed significantly better than children (mean % correct Day 1 85.97%, SD=9.46%, Day 2 84.48%, SD=7.64%), F(1,79)=54.50,p<.001; but there were no significant differences between Days or Lists (Fs<1).

Visual World Eye Tracking Task. Lexical competition between novel words (e.g., "biscal") and existing words (e.g., "biscuit") for each of the Day1/Day2 and Untrained conditions was assessed using a visual world eye tracking task (based on Allopenna et al., 1998; Sekerina & Brooks, 2007). Eye movements were recorded with a Tobii T120 Eye Tracker, sampling at 60 Hz. Stimuli were presented via the Tobii monitor in a random order using E-Prime 1.2 (Schneider, et al, 2002). Participants fixated on a central cross to initiate a trial; the onset of each trial was gaze dependent to ensure that participants began each trial attending to the central fixation point.

Participants were then presented with a quadrant of four pictures for 1000 ms prior to the automatic onset of the pre-recorded verbal instruction (e.g., "Click on the ___") and these carrier phrases were variable in length as we used naturalistic speech. The arrangement was fully counterbalanced so that the target and competitor appeared in different locations for different items. Once the participant made their response by clicking on one of the four pictures with a standard mouse, all pictures disappeared from the screen. Participants were encouraged to respond as accurately and quickly as possible. Experimental, novel filler and existing cohort competitor trials were presented in a randomised order for each participant:

(I)     Experimental trials

Participants completed a total of 48 experimental trials from the three novel word conditions (12 trained on Day 1, 12 trained on Day 2, 24 Untrained trials). In each trial, participants were instructed to click on a target picture and selected one of four pictures arranged in quadrants on a laptop computer screen: (i) The target picture (e.g., biscuit), (ii) a novel competitor (e.g., biscal), (iii) an untrained novel object, (iv) a familiar distracter (e.g., newspaper). See Figure 1 for an example. For the trained conditions the novel competitors had been learned on Day 1 and Day 2, respectively. For the Untrained condition, 12 of novel competitors were taken from the each of training lists (24 in total). If items were used as novel competitors in the Untrained condition, then they were not presented as novel competitors in the trained conditions, hence why there are 12 of each Day1/Day2 trials, but 24 Untrained trials (so for each trained item there is a corresponding untrained item). All items were fully counterbalanced across conditions and participants.

(II)   Novel filler trials

Since the novel word experimental trials always required participants to select an already familiar item (e.g., "Click on the biscuit."), a set of 24 untrained 'novel filler trials' were included to discourage participants from ignoring the unfamiliar novel objects in the array. These novel words were taken from the Graded Nonword Reading Test (Snowling, Stothard et al., 2001) and the Blending Nonwords subtest of the Comprehensive Test of Phonological Processing (Wagner, Torgesen, & Rashotte, 1999). The four pictures in each trial comprised three existing words and a novel object, to allow children to infer the novel object via mutual exclusivity.

(III)   Existing cohort (real word) trials

Twenty trials to identify the profile of competition effects between existing cohort competitors were interleaved with the novel competitor trials. Participants were required to respond in exactly the same way as described above (click on the target object) and in this case a picture of phonological competitor was either displayed or absent. So each trial featured (i) a target (e.g., candle), (ii) either a phonological competitor (e.g., candy) or a phonologically unrelated distracter (e.g., stamp, (iii) an existing phonologically unrelated distracter (e.g., lorry); (iv) a novel distracter object. Again, targets and competitors were rotated across

participants, giving four versions of the task to counterbalance, since participants also receive only half of the stimuli in the displayed/absent condition.
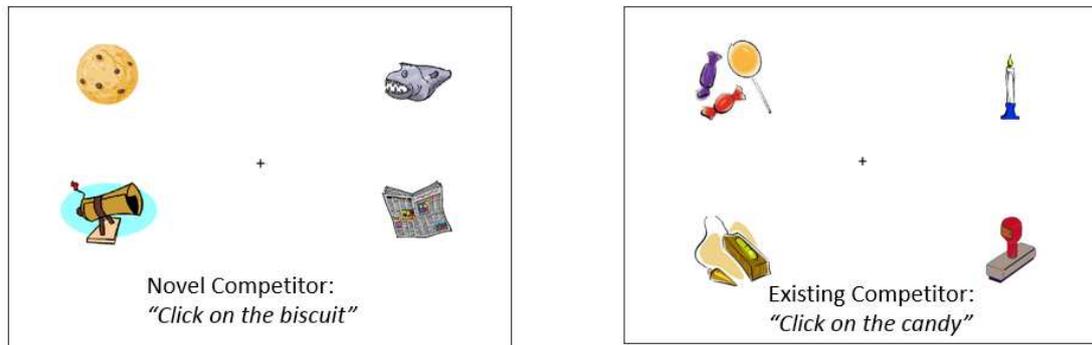


Figure 1: Examples of visual display used in the novel and existing competitor trials. Following a 1000 ms preview participants heard a pre-recorded verbal instruction instructed to "click on the [target object]".

Measures of explicit knowledge. Participants' explicit phonological knowledge of the novel words was measured using a stem completion cued recall task presented on DMDX (Forster & Forster, 2003). Participants heard the first syllable (e.g., "bis-") of the 48 novel words trained on Day 1 and Day 2 and were asked to complete the cue using one of the new words. Accuracy was recorded. Adults' explicit knowledge of novel word-object pairs was measured with an object naming task (administered via DMDX; Forster & Forster, 2003). Naming accuracy and RT were measured. Pilot data revealed that children's performance on the object naming task was at floor; hence, their explicit knowledge of novel word-object pairs was measured with a true/false recognition task (administered via E Prime 1.2; Schneider et al., 2002). On each trial, children heard one of the novel words and simultaneously saw one of the trained novel objects, presented centrally on the laptop screen. They were instructed to press one button (labelled "true") if they thought the novel word picture "went together as they did during the training" or a different button (labelled "false") if they did not. A total of 48 trials were presented (24 Day 1 and 24 Day 2); each novel word was presented only once with either the correct or incorrect novel object to avoid repeating items across "true" and "false" conditions. For each of these tasks the order of the stimuli was randomised for each participant. No feedback was provided for any task.

2.4 Procedure

The experimental procedure is illustrated in Figure 2.
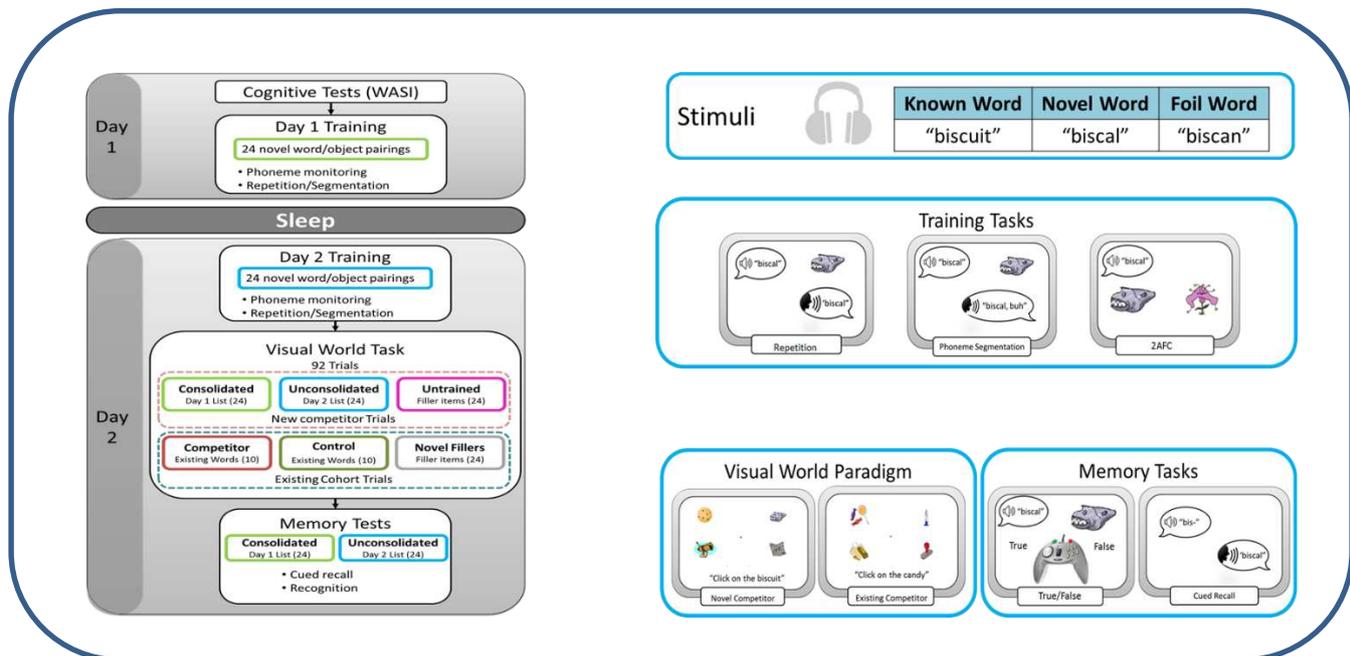


Figure 2: A summary of the procedure used in this study.

Word learning. Day 1 and Day 2 sessions were administered at varying points throughout the day between 08:00 – 18:30 (time of Day 1 session = children's mean 11:09, SD=01:32 hours; adults' mean 13:21, SD=02:57 hours, p>.05; time of the Day 2 session children's mean 11:43, SD=01:43 hours; adults' mean 13:36, SD=03:03 hours, p>.05). The mean time elapsing between Day 1 and Day 2 sessions was 24.08 hours for adults (SD=1.15hours) and 24.54 hours for children (SD=1.45 hours) (p>.05). All participants reported that they had experienced typical sleep on the nights running up to the experiment. There were no significant group differences for participants ratings of their night of sleep between Day 1 and Day 2 tests (children's mean rating 7.12, SD=2.45, adult's mean 6.78, SD=1.85; on a scale of 0-10 with 10 being a very good night of sleep) (p>.05). Children self-reported significantly more hours of sleep on the night between Day 1 and Day 2 sessions than adults (children's mean hours 10.94 hours, SD=1.03 hours, adult's mean hours 7.57 hours, SD=1.22 hours) (p<.001).

Sleep EEG recording. In order to establish when any changes observed for pairings learned the previous day correlated with relevant sleep parameters, a subset of adults slept in the lab overnight between Day 1 and Day 2 (N= 20 sets of useable data were successfully recorded). An Embla N7000 system and Remlogic software were used to record sleep EEG data. Six scalp electrodes were positioned according to the international 10-20 system (F3,

F4, C3, C4, O1, O2) referenced to contralateral mastoids. Two electro-oculographic (EOG) channels monitored eye movements, and two chin electromyographic (EMG) channels monitored muscle tone. All technical and digital specifications, including impedance levels, sampling rates and filter settings were set according to the recommended specifications in the standardised American Academy of Sleep Science Manual (AASM; Iber, Ancoli-Israel Quan, 2007). Data were scored manually in 30-second epochs according to the AASM sleep staging criteria. To confirm primary sleep staging, all data were scored by a second independent scorer who was blind to the original scoring. Remlogic generated inter-scorer reliability report revealed an average of 88% (SD = 2.65%) overall score agreement between the two independent scorers across all stages (wake, NREM, REM) with agreement for Stage 2 sleep at 95% (SD = 3%) and SWS at 94% (SD = 5%) .

Sleep Spindle Analysis.

Polysomnographic (PSG) epochs scored as either Stage 2 sleep or SWS were extracted from central (C3, C4), frontal (F3, F4) and occipital (O1, O2) EEG channels for spindle analysis. Artefacts were rejected from the data using EEGLAB version 10.0 and excessively noisy channels were excluded. A linear finite impulse response filter, also in EEGLAB version 10.0, was then used to bandpass filter each channel at 13.5 – 15 Hz (fast spindles) and 12 – 13.5 Hz (slow spindles). An automated detection algorithm (Ferrarelli, Huber et al. 2007) counted discrete spindle events as amplitude fluctuations within the filtered time series that exceeded a threshold of eight times the mean channel amplitude. Fast and slow spindle density was calculated on all remaining EEG channels for each participant. (total sleep spindles / total minutes of sleep; averaged across frontal and central channels) Several studies have used this method to probe the role of spindles in sleep-dependent memory consolidation. (Tamminen, Payne et al. 2010, Tamminen, Lambon Ralph et al. 2013, Cairney, Durrant et al. 2014; Cairney, Lindsay, Sobczak, Paller, & Gaskell, 2016; Mölle, Bergmann, Marshall, Born, (2011).

## 3. Results

### 3.1.Behavioural data

A summary or memory performance for adults and children can be seen in Table 1 for the cued recall and picture recognition tasks.

Table 1. Descriptive statistics for measures of explicit novel word knowledge and mean (and SD) at day 2 test for words trained on day 1 and day 2, for adults and children.

Note: Picture recognition is picture naming for adults and picture matching for children

| | | Children | | Adults | |
|---|---|---|---|---|---|
| | | Mean (SD) | Range | Mean (SD) | Range |
| Cued Recall (% correct; max 24) | Trained Day 1 | 10.61 (4.48) | 3- 21 | 11.38 (5.12) | 2-20 |
| | Trained Day 2 | 1.90(2.29) | 0 - 12 | 5.74 (3.82) | 0-15 |
| Overnight recall advantage<br>Difference between day 2 and day 1 words; %increase in brackets | | 8.71 (82%) | | 5.64 (50%) | |
| Picture Recognition (% correct) | Trained Day 1 | 69.34 (12.3) | 45.83-95.83 | 19.04 (16.18) | 0-58.33 |
| | Trained Day 2 | 69.12 (13.91) | 37.5 – 96.7 | 17.56 (16.98) | 0 – 79.17 |

### 3.1.1 Cued recall

Cued recall memory performance at test was clearly superior for words learned the previous day (Trained on Day 1) than those that had been recently learned (Trained on Day 2) for adults and children can be seen in Figure 3, and the percentage difference between recall for each word list can also be seen in Table 1. Both groups recalled a similar number of words learned the previous day despite children's poorer recall for recently learned items.
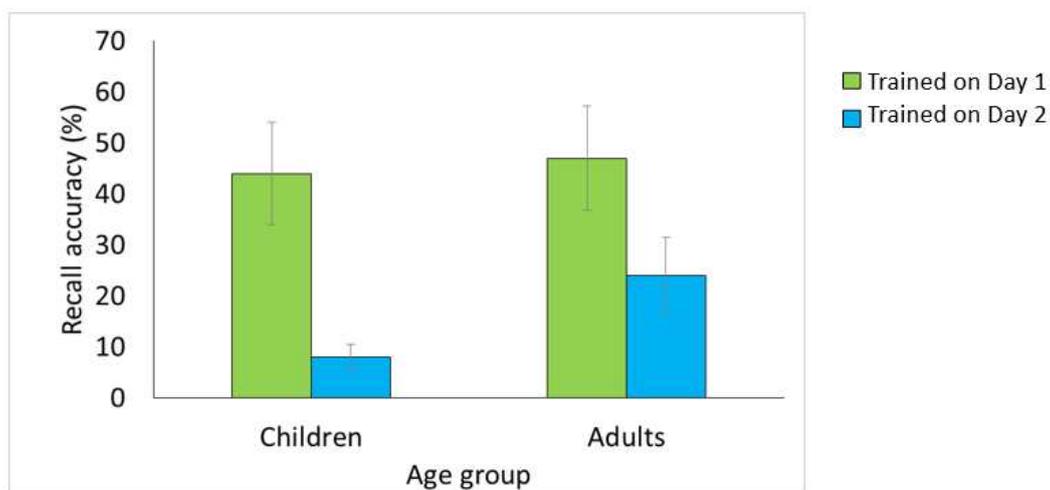


Figure 1: The mean percentage of correctly recalled novel words at test on day 2 when presented with the initial segment (e.g., *bis…)* trained on Day 1 or Day 2 for adults and children (standard error bars are shown). Significantly better recall was observed for words learned the day prior to test for both groups, but children recalled fewer recently learned (same day, day 2) words than adults.

A 2x2 mixed design ANOVA with the within subject factor Day of Training (Day 1/2) and between subjects factor Age (adult/child) confirmed this pattern. Children recalled fewer words than adults ($F_1(1,81)=9.41$, p<.01, $\eta_p^2=.10$; $F_2(1,142)=4.567$, p=.034, $\eta_p^2=.031$) and participants recalled more words learned on Day 1 than Day 2 ($F_1(1,81)=219.29$, p<.0001, $\eta_p^2=.73$; $F_2(1,142)=363.97$, p<.001, $\eta_p^2=.719$). There was also a Day x Age interaction ($F_1(1,81)=10.00$, p < .01, $\eta_p^2=.11$; $F_2(1,142)=30.19$, p<.001, $\eta_p^2=.175$)): Both groups had better recall for the set that were trained on the previous day (Day 1) than the set trained on the day of testing (Day 2). However, children demonstrated a greater Day 1 advantage (MD=36.28%, SD=16.33, 95% CI=31.13 – 41.44%, $t_1(40)=14.23$, p<.001;

$t_2(71)=16.05, p<.001$) than adults (MD=23.51, SD=20.20, 95% CI=17.22 – 29.81%, $t_1$ (41)=7.54, p < .001; $t_2(71)=10.56, p<.001$). Whilst children and adults did not significantly differ for words learned on Day 1 ($t_1<1$, p>.05; $t_2(135.14)=1.10, p=.274$), children recalled significantly fewer novel words learned on Day 2 than adults ($t_1(67.37) = 5.56$, p < .001; $t_2(119)=6.39, p<.001$). This result suggests poorer initial learning of words for children compared with adults, but that enhanced overnight consolidation neutralises this disadvantage.

### 3.1.2    Picture naming / true/false recognition

Adults' explicit knowledge of the novel word-object mapping on the picture naming task was found to be very low (<20% for words learned on both days), and a one-way ANOVA revealed that there was no effect of Day on picture naming accuracy ($F(1,41) = .41$, p>.05, $\eta_p^2 =.01$). Similarly, children's explicit knowledge of the mappings on the true/false recognition task did not differ for Day 1 (69%) and Day 2 (69%) items ($F(1,38) <1$; p >.05). It is notable that recognition is equivalent for Day 1 and Day 2 items, despite robust differences in cued recall for items learned on Day 1 and Day 2.

## 3.2. Eye tracking data

### 3.2.1 Overview

In order to assess the extent to which trained novel word-object pairings would engage in competition with existing base words, fixations to competitor objects were examined.  If the novel word-object pairs enter into competition with the existing lexical item we would expect to see an increase in looks to the trained competitor relative to the untrained control.  Furthermore, if overnight consolidation facilitates stronger competition then we would expect significantly more fixations to the pairings that were learned on Day 1 compared to those learned on Day 2. The extent to which this pattern for is similar for adults and children was examined.  Finally, comparison was made between novel and existing competitors to examine whether newly learned pairings enter into the competition process in a similar way to well established lexical entries.

### 3.2.1 Statistical approach

To assess the timing of various effects, we used a cluster randomization approach originally developed in the neuroimaging literature (Bullmore et al., 1999; Maris & Oostenveld, 2007; for application to visual world data, see Barr, Jackson, and Phillips, 2014).

Cluster randomization uses a "cluster mass statistic" whose null-hypothesis distribution is determined by permutations of the data. This makes it possible to perform multiple testing over many timepoints while keeping the familywise error rate fixed at $\alpha = .05$, and is less conservative than standard Bonferroni correction (Bullmore et al., 1999).

We time-aligned the data for each competitor pair at the lexical disambiguation point (the point in the sound file at which the new or existing competitor deviated from the target). For novel competitor pairs, the mean deviation point was 403 ms from word onset, SD = 106, whereas for existing pairs it was 286 ms, SD = 81. We analysed a window from -500 to 1500 ms (i.e., 500 ms before the deviation point to 1500 ms after). We fitted logistic regression models at each time point, calculated cluster mass statistics for each effect, and derived a null-hypothesis distribution for the effects in the model using synchronized permutation tests (Pesarin, 2001; Salmaso, 2003), a type of permutation test that is suitable for obtaining orthogonal tests of main effects and interactions in a factorial design. We performed separate by-subjects and by-items analyses; the p-values are reported as $p_1$ and $p_2$ respectively. All analyses were performed in R version 3.1.2 (R Core Team, 2014). Further implementation details and results can be found in Appendix C.
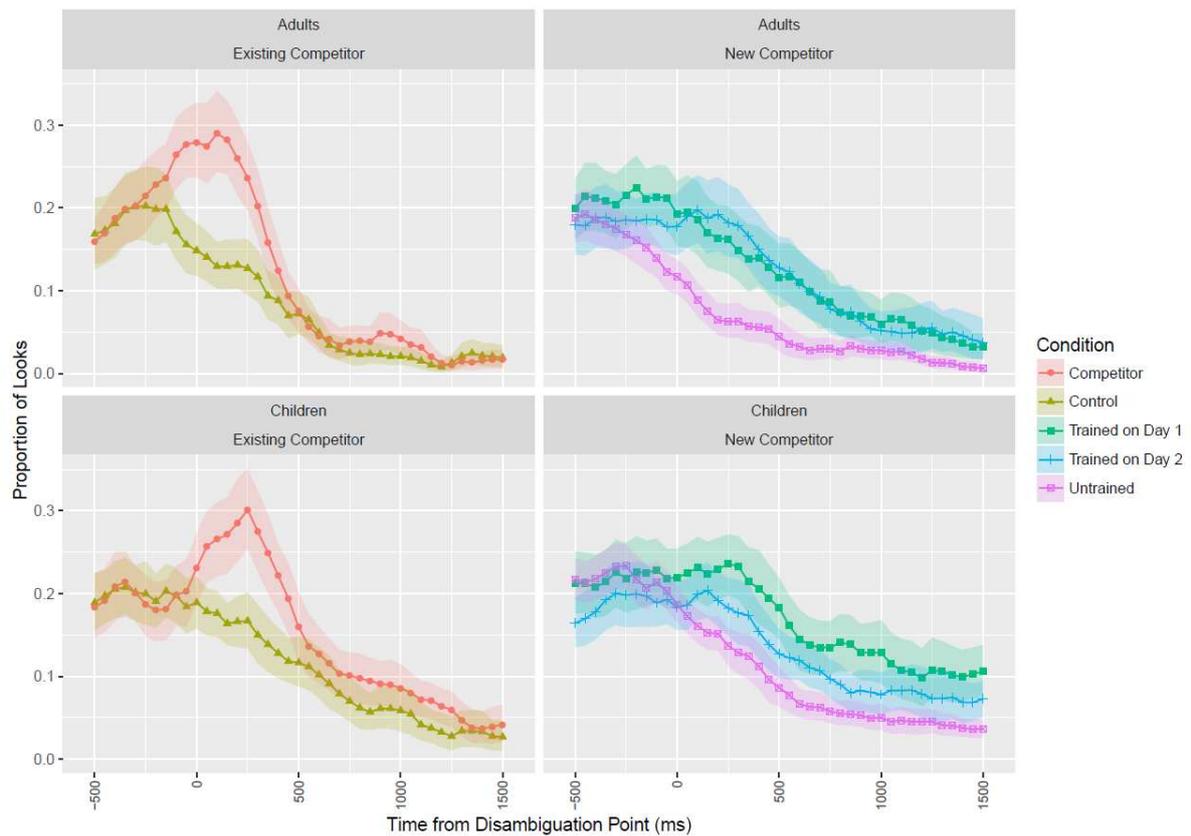
Figure 4. Probability of fixating competitor / control object by Age Group and Competitor Type, time aligned at the lexical disambiguation point (0 ms). The shaded regions represent the 95% confidence interval for each time-series, derived by bootstrapping subject data. The left-hand panels show the expected lexical competition effect for existing words for adults (top) and children (bottom). The right hand panel reveals more subtle, slower, competition effects for novel words. Adults (top panel) fixate items trained on both day 1 and day 2 more than untrained controls; children also show this pattern but competition is boosted for items learned the previous day (day 1).
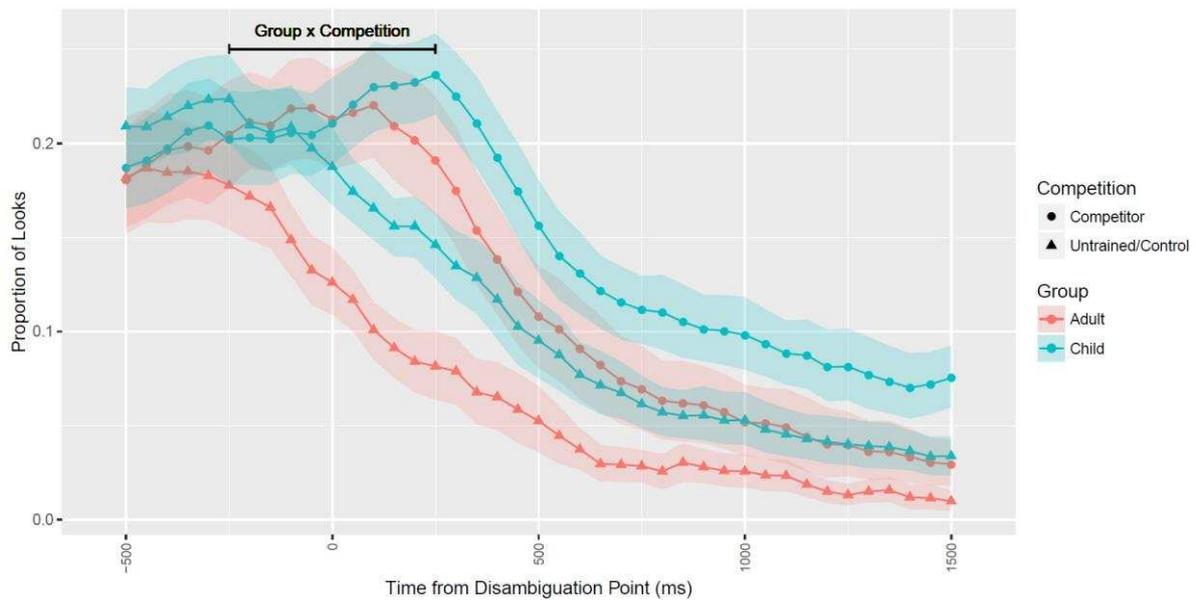
Figure 5: Group by Competition interaction effect for newly learned novel word-novel object associations. The figure shows the probability of fixating novel trained competitors versus untrained control objects by Age Group (collapsed across day of training), time aligned at the lexical disambiguation point (0 ms). The shaded regions represent the 95% confidence interval for each time-series, derived by bootstrapping subject data. The horizontal black bars show the intervals where corresponding effects were statistically reliable. Note that the difference between competitor versus untrained/control pairings rises faster for adults than for children.
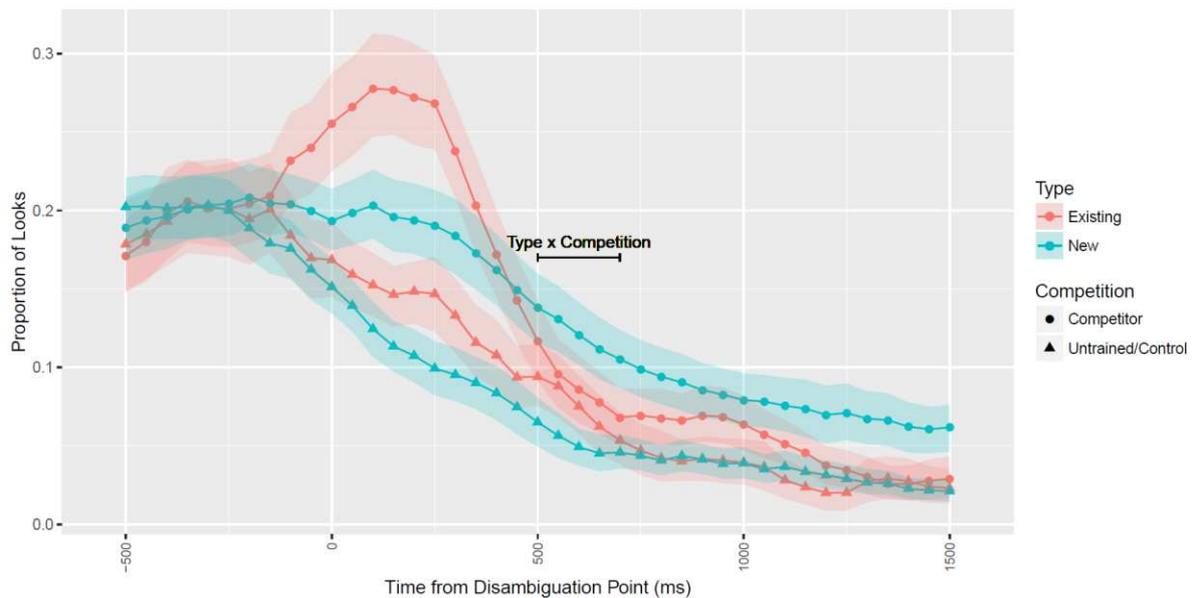


Figure 6: Type by Competition effect. The figure shows the probability of fixating novel trained (collapsed across day of training) or existing competitors versus untrained/control objects (collapsed across age group), time aligned at the lexical disambiguation point (0 ms).

The shaded regions represent the 95% confidence interval for each time-series, derived by bootstrapping subject data. The horizontal black bars show the intervals where corresponding effects were statistically reliable.

Note: the difference between competitors (existing or novel trained on day1/day2) and noncompetitors (control/untrained) lingers longer for the trained associations than the existing associations.

3.2.2 Competition effects for novel and existing competitors

The first set of analyses compared existing competitors to new competitors, collapsing over the Day 1/Day 2 distinction. This analysis had three factors: Group (Adults, Children), competitor Type (Existing, New), and Competition (Competitor, Untrained/Control). As Figure 4 shows, for children as well as for adults, trained competitors and existing competitors were fixated at higher rates than control or untrained objects, main effect of Competition over three intervals in the by-subject analysis (-100 to 50, 350 to 1100, and 1300 to 1350 ms, ps =.005, .001, .001, respectively; and over a single -100 to 1000 ms window in the by-item analysis, $p < .001$). However, adults showed greater overall competition effects (Competitor - Untrained/Control) than children from -250 to 200 ms (-150 to 0 by items; note that these differences in timing between the by subject and by item analyses are a by chance consequence of the separate analyses required by the cluster mass statistic technique), Group x Competition interaction, which is illustrated in Figure 5, $p_1 = .002$, $p_2 = .042$. In this interval, adults were about 1.84 times more likely (in log odds terms) to fixate a competitor than a noncompetitor, whereas children were only about 1.21 times more likely to do so. It was also observed that from 500-700 ms after the DP (450-750 by item), new competitors continued to show competition effects (with listeners about 1.67 times more likely to look at referents trained on Day 1 or Day 2 than an untrained object) whereas competition effects for existing competitors had largely dissipated (1.06 times greater looks to competitor than control), Type x Competition interaction, $p_1 = .013$, $p_2 = .014$, shown in Figure 6. There was no evidence that these differences between existing and novel competitors varied across age group, Group x Type x Competition interaction ps>.31.

Taken together these analyses suggested that existing competitors were activated more strongly and suppressed more efficiently than novel competitors for both adults and children.

3.2.3 The time-course of the emergence of lexical competition

To test effects of day of training on lexical competition, we performed another cluster randomization analysis on new word-object pairings only, including the factors Group (Adults, Children) and Training (trained on Day 1, trained on Day 2). Figure 4 suggests a numerical advantage for Day 1 relative to Day 2 competitors among children. However, the statistical evidence for such a Group-by-Training interaction was unclear, with the effect significant from 200-500 only by items, $p_1 = .341$, $p_2 = .037$.

The cluster randomization analysis is a good way to identify time-points at which particular effects are significant while protecting against Type I errors. However, putting the issue of timing aside we can ask a more basic question of whether across the whole time window competition effects differ for the items trained on Day 1 and Day 2. In order to test this, we repeated the analysis using synchronized permutation tests (as above), except the analysis was performed over the entire window instead of at individual time points. In order to test this, we repeated the analysis collapsing over time, using synchronized permutation tests to get p-values from a logistic regression model. For the original as well as for each permuted version of the data set, we fit a logistic regression model to the data. The analysis yielded evidence for a Group-by-Training interaction, $p_1 = .018$, $p_2 = .013$, with a larger training effect for children than for adults; further analysis of the simple effects of Training for each group showed a reliable effect for children, such that children were about 1.32 times more likely to gaze at competitors trained on Day 1 versus Day 2, $p_1 = .003$, $p_2 = .001$; in contrast, adults showed no such advantage (odds ratio 1.03), $p_1 = .606$, $p_2 = .604$.

Taken together these analyses suggest that novel items enter into competition with their corresponding base words immediately for adults and children, but that there are qualitative differences in the timing and magnitude of these effects compared to existing competitors. Furthermore, children but not adults show an enhancement of this competition effect for items learned the previous day, consistent with a consolidation effect. The extent to which this competition is lexical will be considered in the General Discussion.

3.3.4 Sleep stage analysis

The main sleep parameters of the 20 adult participants for whom sleep EEG data were obtained are displayed in Table 2. The most common measures reflecting spindle activity are depicted in Table 2, comprising the count of all NREM (stage 2 and SWS) spindles detected and the spindle density (mean number of spindles per minute in stage 2 and SWS combined).

Table 2: Observed sleep parameters during the night between Day 1 and Day 2 (N=20).

| Sleep Parameter | Mean time in minutes | Time as a % of total sleep time |
|---|---|---|
| Total sleep time | 489.85 (75.75) | |
| Wake time after sleep onset | 31.60 (17.34) | 0.7 (0.47) |
| Stage 1 | 20.43 (12.03) | 4.09 (2.22) |
| Stage 2 | 273.93 (44.24) | 56.34 (5.77) |
| SWS | 81.9 (27.31) | 16.97 (5.45) |
| REM | 112.75 (33.77) | 22.61 (4.61) |

Note. SWS = slow wave sleep; REM = Rapid eye movement sleep

We calculated correlations (Pearson's R) between time spent in Stage 2 sleep, SWS, rapid eye movement sleep, and the difference in performance for Day 1 and Day 2 items in cued recall and the difference in lexical competition for items trained on Day 1 versus items trained on Day 2. We also evaluated correlations between sleep stages and performance on Day 1 and Day 2. For each task, p values were corrected for multiple comparisons (Bonferroni). No significant relationships between these sleep components and cued recall or gaze behaviour were revealed.

Table 3: Spindles detected and spindle density (N=20)

| | All | F3 | F4 | C3 | C4 | 01 | 02 |
|---|---|---|---|---|---|---|---|
| Fast spindle density | 1.09 (.41) | 1.49 (.61) | 1.45 (.58) | 1.29 (.52) | 1.32 (.58) | 0.52 (.33) | 0.49 (.31) |
| Slow spindle density | 1.02 (.30) | 1.59 (.54) | 1.53 (.50) | 1.09 (.42) | 1.07 (.41) | 0.44 (.23) | 0.40 (0.20) |

Note: Spindle Density = mean spindle count per minute

We then examined the role of fast (13.5 – 15Hz) and slow (12 – 13.5 Hz) sleep spindles in the observed sleep-dependent changes in memory. Correlation coefficients were calculated for sleep spindle density (total sleep spindles / total minutes of sleep; averaged across frontal and central channels) and magnitude of the cued recall advantage (trained on Day 1 – trained on Day 2). Spindle density for Frontal (F) and Central (C) channels can be

seen in Table 3. Occipital (O) channels are also included for reference but were not included in the analyses. There was a highly significant positive relationship between the magnitude of the sleep advantage for cued recall and fast spindle density (r = .567, p =.009) explained by a significant correlation between spindle density and cued recall for items learned on the previous day (r=.471, p =.036) but not items on the same day (r=-.143, p= .547). The same difference score analysis for slow spindles revealed a nonsignificant trend in the same direction (r = .40, p = .08).

The density of fast (r = .103, p = .667) or slow (r = .009, p = .970) spindles was not found to be related to the magnitude of the consolidation effect in the VWP task. Furthermore, there were no correlations between cued recall or fixation data and SWS duration. This pattern further supports the interpretation that the cued recall advantage for items learned the previous day may in part depend upon sleep-associated memory consolidation; and the absence of differences in the competition effect reflect a process that is independent of the passing of time and not associated with sleep in adults.

## 4. General Discussion

Competition between lexical candidates has been thought of as a product of automaticity in spoken word recognition. The present study made use of the VWP to assess the time course with which novel referents (e.g., biscal) compete for recognition when children and adults are presented with familiar spoken competitors (e.g., biscuit). Crucially, the observation of competition effects for known items (e.g., candy/candle) enabled a direct comparison between well integrated lexical items and newly learned words.

Consistent with previous studies (e.g., Brown et al., 2012; Henderson et al., 2012; Henderson et al., 2013), explicit memory for the novel words (measured via a cued recall task) was superior for words that had a 24-hour consolidation opportunity than words learned on the day of the test for both age groups, with this consolidation effect being stronger for children. Furthermore, for adults, cued recall performance for the consolidated novel words (but not unconsolidated words) correlated with sleep spindle density (fast spindles), suggesting an active role for sleep in the strengthening of new lexical representations (e.g., Tamminen et al., 2010).

For adults, novel words exhibited similar competition effects regardless of whether they were learned immediately prior to test or on the previous day, suggesting that these

effects emerged swiftly. Children also showed immediate competition effects, but in contrast to adults the competition effect was enhanced for words learned the previous day. Thus, whilst an opportunity for consolidation leads to dramatic improvements in cued recall in both age groups (adults doubled the number of words they recalled, whereas for children the improvement was more than six-fold), whatever changes in representation that underpin this improvement have little or no effect on the ability of the novel words to act as competitors in the VWP for adults. The evidence of immediate competition effects marks a clear contrast with findings obtained using the pause detection task (e.g., Henderson et al, 2012; Dumay & Gaskell, 2007) but is similar to the pattern observed in another recent study using the VWP (Kapnoula et al, 2015a).

Nevertheless, a clear difference in the time course of competition between the existing and the novel words suggests that neither the items learned just before test nor the items learned the previous day were word-like in their time-course. Specifically, competition persisted for longer for the novel competitors, in comparison to existing competitors, which this difference restricted to the duration of the competition period rather than the onset. This suggests that whilst newly learned words may enter in competition in the VWP soon after learning (consistent with Kapnoula et al, 2015-a) they may not behave liked fully-fledged lexical representations. This resonates with findings from other studies that have revealed competition effects between novel items in artificial lexicons immediately after learning that tend to have a slower time-course than existing words (Creel et al., 2008; Magnuson et al., 2002; Kapnoula et al., 2015-a; Pirog-Revill et al., 2008). However, recent studies by Kapnoula and colleagues (Kapnoula et al, 2015-a; 2015-b) have found competition effects for novel words that are indistinguishable from established word competitors immediately after learning. A key difference that separates their studies from the other VWP studies discussed here is that the novel words were trained without an association to a referent. Therefore, new word forms may become wordlike as assessed by competition in the VWP quite quickly (as demonstrated by Kapnoula and colleagues), but establishing an automatic link to stored knowledge of the meaning of the word may be a slower process lasting days or weeks (cf. Tamminen & Gaskell, 2012). A second important difference is that because of the way the materials in Kapnoula and colleagues' experiments are cross-spliced, participants hear vowel cues that for a brief period of time (before encountering the final burst) favour a different consonant, and this may particularly encourage participants to relate the incoming material to auditory sequences from their recent experience. We return to this issue in the following

section addressing the theoretical interpretation of our results and their integration with previous findings.

4.1 Theoretical implications

The present pattern of results can be explained at least to some extent by reference to the complementary learning systems account (CLS) of word learning (Davis & Gaskell, 2009; Lindsay & Gaskell, 2010), which assumes that different facets of word learning may follow different trajectories. On this account, the diverse set of mappings involved in linking a word's phonological, orthographic and semantic representations can be thought of as processing pathways that must develop and strengthen over time and with repeated exposure. New words are initially acquired using the hippocampus to provide mediating links between relevant areas of the cortex (e.g., representing lexical form and meaning). These mediating links allow many aspects of lexical processing to be fulfilled (e.g., retrieving the meaning of a newly learned word), but indirect links between the areas of the cortex via the hippocampus are likely to be relatively slow or to have a lower priority compared with the extant direct cortical links for words that are long-established in the lexicon. The CLS model predicts that competitor fixations to the novel word referent would be initially supported by these slow hippocampal links, allowing competition to be observed, but with a relatively slow and extended time-course compared to well-established competitors.

What is less clear is why the emergence of competition effects has such different apparent time courses depending on the test applied. Swift (pre-sleep) competition effects are consistently found in VWP studies, whereas the emergence of these effects is typically post-consolidation period in response time paradigms such as pause detection (e.g., Dumay & Gaskell, 2007; Henderson et al., 2012). We think that both these apparently conflicting observations are correct, but they are based on competition paradigms that highlight different components of the same process. Here, we examine the key differences between these types of test in an attempt to explain how the same underlying system could support both patterns of data.

First, an important distinction between pause detection and the standard VWP as used here is that in the VWP the novel competitor object is present on screen (along with the other three distractors) for 1000 ms prior to the onset of the critical target sentence (e.g., "click on the biscuit"). Hence, the novel object is effectively cued in the standard VWP, but not in the pause detection paradigm, which may facilitate the parallel activation of both representations,

even if the novel word representation is reliant on slower hippocampal retrieval. If the competitor was similarly cued prior to the presentation of the base words in the pause detection task, it is plausible that immediate competition effects would ensue.

Second, there is an important difference in terms of the continuity of the dependent variable. RT measures rely on a single speeded judgement, and it may well be that slower hippocampally mediated links available immediately after learning are not efficient enough to contribute to the formation of this kind of response. In comparison, fixation data provide a more continuous and extended measure of competition, and so may be better able to incorporate information arriving relatively slowly via recently learned hippocampal links. In other words, the speeded response time tests may be purer tests of the availability of direct links within the cortical network that consolidation provides.

Thirdly, there may well be significant differences in the types of memory system that are recruited for a VWP experiment as compared with pause detection or lexical decision tasks. As participants view a visual scene, they clearly develop a mental model of the objects they are shown. Such a mental model allows for linguistically mediated fixations even when the visual display has been removed (e.g., Altmann, 2004). When listeners encounter incoming speech and relate it to their mental model, they are undoubtedly informed by the cortically stored lexical knowledge of the words that they hear. But nonetheless, speculatively, the generation of the mental model might depend much more heavily on the hippocampal episodic memory (e.g., Horner et al., 2015) of the novel words, alongside the short-term memory of the visual scene. Thus the discrepant results that we see between the VWP and response time tasks may also be a product of different weightings applied to hippocampal and neocortically mediated representations of newly learned words as a consequence of the memory demands of the task.

All three of these differences are relevant to the notion of automaticity and the possibility that VWP and response time tasks rely to differing extents on automatic perceptual processes. Automaticity is a much-debated notion in cognition, and many contrasting definitions have been proposed. Nonetheless, as a graded and multidimensional concept (Moors and De Houwer, 2006), incorporating components such as intentionality, speed, efficiency and controllability it has proved useful for distinguishing between different types of perceptual process. Recent studies have shown that consolidation of new words leads to enhanced automaticity, as measured by semantic decision (Tham, Lindsay & Gaskell, 2015) and Stroop tasks (Geukes et al., 2015). For example, Tham, Lindsay & Gaskell (2015)

used two tasks that have been conceptualised as tests of automaticity in word learning: semantic distance and the semantic congruity. They observed that whilst the semantic distance effect emerged very quickly after learning, effects of semantic congruity—deemed a stronger test of automaticity—only emerged after a delay and were associated with slow wave sleep and spindle activity. The mapping necessary to elicit competition on the VWP is set up "on the fly" and so likely to be non-automatic by most definitions, and therefore available early in the consolidation process. On the other hand the lexical decision and pause detection tasks require fast and efficient evaluation of the relationship between new words and their existing neighbours, and so may require a greater level of automaticity. Furthermore, it is possible that swift and efficient links between spoken input and fixations to their referents may require an even greater level of automaticity than is available after a night or two of sleep: word-like competition effects in the VWP, which we did not observe in the current experiment, may depend on automatic processing across the entire chain of processes that link the perception of a sequence of sounds to goal direct fixations to a familiar object.

In line with the view that competition effects measured in the VWP and the pause detection paradigm demand different levels of automaticity, McMurray, Kapnoula and Gaskell (in press) found that lexical effects emerge at substantially different rates depending on the nature of the effect and the way in which it is tested. Many aspects of lexical processing are observed immediately after the learning has taken place (e.g., we can make judgements about the way a word sounds, and use it appropriately in conversation). Other aspects, however, emerge or become more evident over time, with sleep often being important (e.g., Dumay & Gaskell, 2007; Tamminen et al., 2010). For some types of test (e.g., masked semantic priming) evidence of lexical status may not emerge without a period of consolidation lasting several days or even weeks.

Intriguingly, a greater reliance on hippocampus-dependent representations in the VWP might also explain why the competitor effects we find here are extended in time, suggesting inefficient suppression of the novel word activations. Normal lexical competition will rely on operational bidirectional inhibitory links between new and existing words within the mental lexicon. But if the VWP paradigm makes use of a mental model that combines information from cortical representations of existing words with hippocampal representations of novel words, then there may be no way to inhibit the hippocampal representations properly, leading to competition over a longer period of time. In other words, the mechanism that underpins linguistically mediated fixations in the VWP may be able to recruit both

hippocampally dependent and cortically dependent word representations and combine the information from them to guide eye movements, but without being able to recreate the competitive links that would emerge in the longer term as the novel word representation becomes fully embedded in the neocortex.

An alternative explanation for the lack of a word-like profile of competition effects observed for novel compared with existing words might be that the novel words are, to some extent, functioning as typical word competitors but with a relatively low frequency. However, it is important to note that the profile of competition comparing new and existing items is very different to that observed when simply comparing low and high frequency words. Dahan et al., (2001) found that the strength of the competitor effect was sensitive to the frequency of the competitor, with more fixations to high than low frequency competitors. Nonetheless, the time course of these competitor effects was highly similar, with no evidence that either competitor type led to a more extended competition. This is in direct contrast to our results which reveal novel words being suppressed relatively late compared to existing words. Another point worth bearing in mind is that the comparison between depicted and undepicted existing word competitors is not entirely equivalent to the comparison within novel words between learned (and depicted) items and untrained (and therefore not depicted) controls. For the estimate of competition in the case of existing words, the state of the lexicon does not change between test and control, only the presence or absence of the pictured referent changes. On the other hand, for the novel words the state of the lexicon may change and also the familiarity of the pictured referent. Further research will be needed to confirm that the differences observed here are robust across a range of visual scene manipulations.

4.2 Comparison of competition effects for adults and children

With respect to changes over ontogenetic time, in many ways we observed strikingly similar performance in adults and children. A similar pattern of lexical competition for existing lexical items (e.g., more looks to phonological competitors [e.g., candy] compared to unrelated objects [e.g., pencil], when presented with a spoken target noun, [e.g., candle]) was observed for both adults and children, suggesting that adult-like lexical competition mechanisms are in place for familiar words earlier in development. However, the time course of this lexical competition was a little delayed in children compared to adults, perhaps reflecting their less automatic and slower spoken word recognition processes (as suggested by Ojima, Matsuba-Kurita, Nakamura & Hagiwara, 2011). Despite these striking similarities, there were some crucial differences. For children the VWP competition effect was boosted

for items that had been learned prior to overnight sleep, consistent with previous data (e.g., Henderson et al., 2015).

Children also showed a significantly larger benefit of consolidation in the cued recall data, suggesting that explicit memory was boosted off-line to a greater extent for children than adults. Retention was much poorer for children than adults for words learned that day, but this disadvantage was completely eliminated after a 24 hour consolidation opportunity. This developmental difference is made even more striking by the finding that children performed ~10% more poorly on average than adults in a 2AFC training task that assessed knowledge of the same items that were recalled at adult-like levels on the day after training. Together, these data are consistent with findings that sleep enhances memory to a greater degree in childhood than in adulthood, possibly as a consequence of greater SWS in childhood (e.g., Wilhelm et al., 2013, 2014). It is plausible that sleep underpins a boost in hippocampally-mediated competition (as measured by the VWP) for children. The observation of a sleep associated-boost in memory consolidation for children, but not adults, is particularly intriguing in the light of the recent finding that enhanced slow wave activity (defined as EEG 1 – 4.5 Hz) after visuo-motor learning may reflect heightened experience-dependent plasticity in children (aged 9 – 11 years old) compared to adults (Willhelm et al., 2015). Further research including comparative polysomnographic data for children and adults is required to further elucidate the extent to which behavioural differences are a manifestation of sleep-associated neural change across development.

5. Conclusions

We found that explicit memory was superior for words learned the day before testing; this effect correlated with sleep-spindle density and suggests that this aspect of memory is actively boosted during sleep. In contrast, data from the VWP revealed that recently acquired words can compete for recognition, and for adults no further boost was evident after a delay including overnight sleep. Despite this, newly learned words were not recognised with the same time course as existing lexical items. In children, however, a further increase in strength of competition was observed, and children also demonstrated greater improvements explicit memory for words learned the previous day. Together, the results suggest that different facets of new word learning follow different time courses: visual world competition effects can emerge swiftly, but are qualitatively different from those observed with established words. Furthermore, the findings fit with the view that word learning earlier in development is boosted by sleep to a greater degree. This rich pattern of engagement of novel

words in lexical processes can be interpreted in terms of a complementary systems account of word learning, for which learning and consolidation of novel words involves an extended process of enhancement of automaticity in recognition.

## References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. Journal of memory and language, 38(4), 419-439.

Altmann, G. T. (2004). Language-mediated eye movements in the absence of a visual world: The 'blank screen paradigm'. Cognition, 93(2), B79-B87.

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

Barr, D. J., Jackson, L., & Phillips, I. (2014). Using a voice to put a name to a face: the psycholinguistics of proper name comprehension. Journal of Experimental Psychology: General, 143, 404-413.

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

Brown, H., Weighall, A., Henderson, L.M., & Gaskell, M.G. (2012). Enhanced recognition and recall of new words in 7- and 12-year old children following a period of offline consolidation. Journal of Experimental Child Psychology, 112, 56-72.

Bellander, M., Berggren, R., Mårtensson, J., Brehmer, Y., Wenger, E., Li, T. Q., ... & Lövdén, M. (2015). Behavioral correlates of changes in hippocampal gray matter structure during acquisition of foreign vocabulary. NeuroImage.

Bloom, P., & Markson, L. (1998). Capacities underlying word learning. Trends in Cognitive Science, 2, 67-73.

Bouwmeester, S., & Verkoeijen, P.P.J.L. (2011). Why do some children benefit more from testing than others? Gist trace processing to explain the testing effect. Journal of Memory and Language, 65, 32-41.

Bullmore, E. T., Suckling, J., Overmeyer, S., Rabe-Hesketh, S., Taylor, E., & Brammer, M. J. (1999). Global, voxel, and cluster tests, by theory and permutation, for a difference between two groups of structural MR images of the brain. IEEE Transactions on Medical Imagine, 18, 32-42.

Cairney, S. A., Lindsay, S., Sobczak, J. M., Paller, K. A., & Gaskell, M. G. (2016). The Benefits of Targeted Memory Reactivation for Consolidation in Sleep are Contingent on Memory Accuracy and Direct Cue-Memory Associations. Sleep.

Cairney, S. A., S. J. Durrant, J. Hulleman and P. A. Lewis (2014). Targeted memory reactivation during slow wave sleep facilitates emotional memory consolidation. Sleep **37**: 701-707.

Cairney, S. A., S. J. Durrant, R. Jackson and P. A. Lewis (2014). Sleep spindles provide indirect support to the consolidation of emotional encoding contexts. Neuropsychologia **63**(0): 285-292.

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. Child Language Development, 15, 17-29.

Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2008). Heeding the voice experience:The role of talker variation in lexical access. Cognition, 106(2), 633-664.

Dahan, D., Magnuson, J.S., & Tanenhaus, M.K. (2001). Time course of frequency effects in
spoken-word recognition: Evidence from eye movements. Cognitive Psychology, 42, 317 367.

Dahan, D., Magnuson, J.S., Tanenhaus, M.K., & Hogan, E. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. Language and Cognitive Processes, 16(5/6), 507-534.

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

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

Diekelmann, S., & Born, J. (2010). The memory function of sleep. Nat Rev Neurosci, 11(2),

114-126.

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

Dumay, N., Gaskell, M. G. & Feng, X. (2004). A day in the life of a spoken word. Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society. (pp. 339-344). Mahwah, NJ: Lawrence Erlbaum Associates.

Ferrarelli, F., R. Huber, M. Peterson, M. Massimini, M. Murphy, B. Riedner, A. Watson, P. Bria and G. Tononi (2007). Reduced sleep spindle activity in schizophrenia patients.  The American Journal of Psychiatry, **164**: 483-492.


Fernandes, T., Kolinsky, R., & Ventura, P. (2009). The metamorphosis of the statistical
segmentation output: Lexicalization during artificial language learning. Cognition, 112(3), 349-366.

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

French, R. M. (1999). Catastrophic forgetting in connectionist networks. Trends in Cognitive Sciences, 3(4), 128-135.

Gaskell, M. G., & Marslen-Wilson, W. D. (2002). Representation and competition in the perception of spoken words. Cognitive psychology, 45(2), 220-266.

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

Geukes, S., Gaskell, M.G., & Zwitserlood, P. (2015). Stroop effects from newly learned color words: Effects of memory consolidation and episodic context. Frontiers in Psychology, 6, 1-16.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. Ear and hearing, 19(1), 1.

Massimini, Marcello, et al. "Triggering sleep slow waves by transcranial magnetic stimulation." Proceedings of the National Academy of Sciences 104.20 (2007): 8496-8501.

Mölle, M., & Bergmann, T. O; Marshall l; Born J. (2011) Fast and slow spindles during the sleep slow oscillation: disparate coalescence and engage-ment in memory processing. Sleep, 34(10), 1411-1421.

Nishida, M., & Walker, M. P. (2007). Daytime naps, motor memory consolidation and regionally specific sleep spindles. PloS one, 2(4), e341.

Henderson, L., 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.

Henderson, L., Weighall, A., Brown, H., & Gaskell, G. (2013). Online lexical competition during spoken word recognition and word learning in children and adults. Child development, 84(5), 1668-1685.

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

Horner, A. J., Bisby, J. A., Bush, D., Lin, W. J., & Burgess, N. (2015). Evidence for holistic episodic recollection via hippocampal pattern completion. Nature communications, 6.

Horváth, K., Myers, K., Foster, R., & Plunkett, K. (2015). Napping facilitates word learning in early lexical development. Journal of sleep research.

Leach, L., & Samuel, A. (2007). Lexical configuration and lexical engagement: When adults learn new words. Cognitive Psychology, 55, 306-353.

Magnuson, J. S., McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2003). Lexical effects on compensation for coarticulation: a tale of two systems?. Cognitive Science, 27(5), 801-805.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. Journal of Neuroscience Methods, 164, 177-190.

Marslen-Wilson, W., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. Psychological Review, 101(4), 653-675.

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

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. Cognitive Psychology, 10, 1-86.

Nishida, M., & Walker, M. P. (2007). Daytime naps, motor memory consolidation and regionally specific sleep spindles. PloS one, 2(4), e341.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. Cognition, 52(3), 189-234.

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

Kapnoula, E., Gupta, P., Packard, S., & McMurray, B. (2015). Immediate lexical integration of novel word forms. Cognition, 134(1), 85-99.

Kapnoula, E. C., & McMurray, B. (2015-b). Newly learned word forms are abstract and integrated immediately after acquisition. Psychonomic bulletin & review, 1-9.

Luce, P.A., & Pisoni, D.B. (1998). Recognizing spoken words: The neighborhood activation model. Ear and Hearing, 19(1), 1-36.

McMurray, B., Samelson, V. M., Lee, S. H., & Tomblin, J. B. (2010). Individual differences in online spoken word recognition: Implications for SLI. Cognitive psychology, 60(1), 1-39.

McMurray B, Horst J, Samuelson L. 2012 Word learning emerges from the interaction of online referent selection and slow associative learning. Psychological Review. 119, 831–877.

McMurray, B. Kapnoula, E. C., & Gaskell, M.G. (2015, in press). Learning and integration of new word forms: Consolidation, pruning and the emergence of automaticity

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

Mölle, M., & Bergmann, T. O. (2011). Marshall l; Born J. Fast and slow spindles during the sleep slow oscillation: disparate coalescence and engage-ment in memory processing. Sleep, 34(10), 1411-1421.

Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. Psychological Review, 110(4), 611-646.

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

Pesarin, F. (2001). Multivariate permutation tests: with applications in biostatistics. Chichester: Wiley.

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rasch, B., & Born, J. (2013). About sleep's role in memory. Physiological reviews, 93(2), 681-766.

Salmaso, L. (2003). Synchronized permutation tests in $2^k$ factorial designs. Communications in Statistics-Theory and Methods, 32, 1419-1437.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-*prime user's guide*. Pittsburgh: Psychology Software Tools.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. Science, 268(5217), 1632-1634.

Schreiner, T., & Rasch, B. (2016). The beneficial role of memory reactivation for language learning during sleep: A review. Brain and Language.

Speigel, C. & Halberda, J. (2011). Rapid fast-mapping abilities in 2-year-olds. Journal of Experimental Child Psychology, 109, 132 – 140.

Szmalec, A., Page, M., & Duyck, W. (2012). The development of long-term lexical representations through Hebb repetition learning. Journal of Memory and Language, 67(3), 342-354.

Tamminen, J. & Gaskell, M.G. (2008). Newly learned spoken words sholong-term lexical competition effects. The Quarterly Journal of Experimental Psychology, 61, 361-371.

Tamminen, J., & Gaskell, M.G. (2012). Novel Word Integration in the Mental Lexicon: Evidence from Unmasked and Masked Semantic Priming. Quarterly Journal of Experimental Psychology. iFirst, 1-25. DOI: 10.1080/17479218.2012.724694

Tamminen, J., Payne, J. D., Stickgold, R., Wamsley, E. J., & Gaskell, M. G. (2010). Sleep Spindle Activity is Associated with the Integration of New Memories and Existing Knowledge. Journal of Neuroscience, 30(43), 14356-14360.

Tamminen, J., M. A. Lambon Ralph and P. A. Lewis (2013). The Role of Sleep Spindles and Slow-Wave Activity in Integrating New Information in Semantic Memory. The Journal of Neuroscience, **33**(39): 15376-15381.

Tham, E. K. H., Lindsay, S., & Gaskell, M. G. (2015). Markers of automaticity in sleep-associated consolidation of novel words. Neuropsychologia, 71, 146-157.

Wechsler, D. (1999).  Wechsler Abbreviated Scale of Intelligence (WASI).  San Antonio, TX:  Harcourt Assessment.

Wilhelm, I., Rose, M., Imhof, K.I., Rasch, B., Buchel, C., & Born, J. (2013). The sleeping child outplays the adult's capacity to convert implicit into explicit knowledge. Nature Neuroscience, 16, 391-393.

Wilson, M. (1988). MRC Psycholinguistic Database – machine-usable dictionary, version 2.00. Behaviour Research Methods Instruments & Computers, 20, 6 – 10.

Wilson, M. A., & McNaughton, B. L. (1994). Reactivation of hippocampal ensemble memories during sleep. Science, 265, 676-679.

Rasch, B., & Born, J. (2013). About sleep's role in memory. Physiological reviews, 93(2), 681-766.

# Appendices

## Appendix A: Novel words and corresponding basewords

| List 1 | | List 2 | | List 3 | |
|---|---|---|---|---|---|
| **Base word** | **Novel Word** | **Base word** | **Novel word** | **Base word** | **Novel word** |
| apricot | apricam | angel | angesh | athlete | athlove |
| alien | aliet | badminton | badminteef | balcony | balcozo |
| baboon | baboop | biscuit | biscal | blossom | blossail |
| bikini | bikinar | bramble | brambo | breakfast | breakfal |
| bracelet | bracelop | broccoli | broccaroo | buffalo | buffaluk |
| cactus | cactul | caravan | caravat | cardigan | cardigite |
| caramel | caramen | chocolate | chocolor | clarinet | clarinone |
| chimpanzee | chimpantu | costume | costuke | crocodile | crocodol |
| donkey | donkop | dinosaur | dinosut | dungeon | dungeoth |
| dolphin | dolphik | daffodil | daffodote | flamingo | flamingist |
| fountain | fountel | gadget | gadgel | guitar | guitas |
| graffiti | graffino | mermaid | mermiff | kangaroo | kangami |
| lantern | lantobe | nugget | nuggev | mushroom | mushrood |
| mayonnaise | mayonnote | onion | oniot | octopus | octopum |
| napkin | napkig | pelican | pelical | parachute | parasheff |
| ornament | ornameld | penguin | pengwove | parsnip | parsnin |
| parade | parafe | pyramid | pyramon | picnic | picnin |
| potato | potatuck | sergeant | sergeast | reptile | reptite |
| pumpkin | pumpkige | signature | signatik | siren | siredge |
| rugby | rugbock | somersault | somersaumf | spider | spidet |
| skeleton | skeledu | target | targil | tornado | tornadus |
| squirrel | squirrome | tattoo | tattefe | tulip | tulode |
| tissue | tissove | trombone | trombal | volcano | volcagi |
| walnut | walnog | walrus | walrick | yoghurt | yogem |

## Appendix B: Existing cohort competitor stimulus lists

| List 1 | List 2 |
|---|---|
| beaker | beetle |
| candle | candy |
| paddle | padlock |
| sandal | sandwich |
| bacon | baker |
| pencil | penny |
| monkey | money |

| | |
|---|---|
| lolly | lorry |
| camper | camel |
| window | winner |
| pasta | pasty |
| kitten | kitchen |
| cartoon | carton |
| butter | button |
| medal | metal |
| packet | package |
| letter | lettuce |
| caterpillar | catalogue |
| circle | circus |
| robin | robber |

**Appendix C**

**Cluster-Based Randomization of Visual-World Eyetracking Data**

The cluster randomization approach, originally developed for neuroimaging (Bullmore et al., 1999; Maris & Oostenveld, 2007) can also be applied to visual world eyetracking data (e.g., Barr, Jackson, and Phillips, 2014), or indeed to any situation involving multiple testing, and is advangeous to other approaches such as Bonferroni correction whenever testing points show autocorrelation over some dimension. As applied here, inferential tests were performed at two separate stages. At the first stage, uncorrected tests were performed on each sample of eye data over the analysis window. At the second, the results from the individual tests were combined into a cluster mass statistic, whose null-distribution was derived through data permutation. Let us consider each of these stages in more detail.

To speed up the computations, the data were binned into 50 ms bins (3 samples) over the entire 2000 ms window (-500 to 1500 ms), resulting in 41 bins. At the first stage, parameters were estimated at each of these 41 bins of data using logistic regression (the multinom function in package nnet of R). The estimation itself did not take into account sampling unit variaibility, as this was estimated instead for each parameter on 1000 bootstrap resamples of the data taken over the relevant sampling unit (subjects or items). The test statistic used was a Wald statistic (parameter estimate divided by bootstrapped standard errors). The p-value for each effect at each bin was calculated, and runs in the time-series were identified for a given effect where $p < .05$ and for which all estimates of the effect were of the same sign. For each run, a cluster mass statistic was calculated for all $p_i$ in the run using the formula $CMS = \Sigma\text{-}2 \log p_i$ (Barr, Jackson, and Philips, 2014).

Cluster mass statistics for each effect were calculated for the original data. At the second stage, NHDs for each of these statistics were obtained by creating 1,000 permutations of the dataset according to a synchronized permutation scheme (Pesarin, 2001; Salmaso, 2003) and calculating the maximum CMS for each effect in each permuted dataset. The synchronized permutation scheme was used because it yields orthogonal tests of main effects and interactions for factorial designs. The distribution of the CMS over the 1,000 datasets provides a NHD for the original CMSs.

**Supplementary materials**

S1: Base words, novel competitors and corresponding novel objects.

| Base word | Novel Competitor | Base word Object | Novel Competitor Object |
|-----------|------------------|------------------|-------------------------|
| alien | aliet | | |
| angel | angesh | | |
| apricot | apricam | | |
| athlete | athlove | | |
| baboon | baboop | | |
| badminton | badminteef | | |
| balcony | balcozo | | |
| bikini | bikinar | | |

| biscuit | biscal | | |
| blossom | blossail | | |
| bracelet | bracelop | | |
| bramble | brambo | | |
| breakfast | breakfal | | |
| broccoli | broccaroo | | |
| buffalo | buffaluk | | |
| cactus | cactul | | |
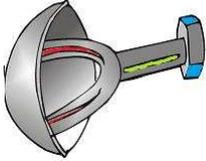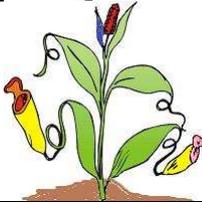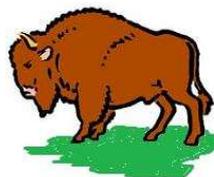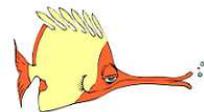
caramel    caramen
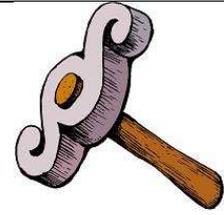
caravan    caravat

cardigan    cardigite
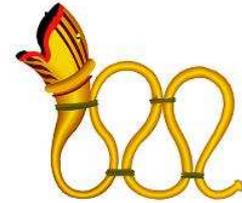
chimpanzee    chimpantu

chocolate    chocolor
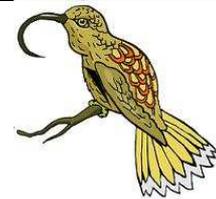
clarinet    clarinone
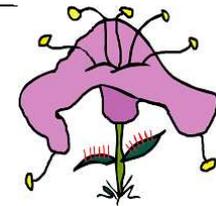
costume    costuke
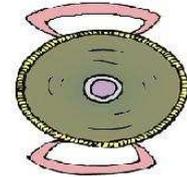
crocodile    crocodol
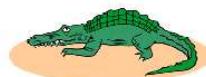
daffodil    daffodote
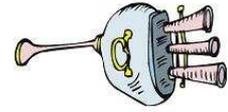
dinosaur    dinosut
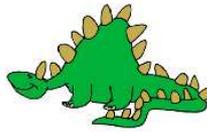
dolphin    dolphik

donkey    donkop

dungeon    dungeoth
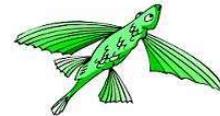
flamingo    flamingist

fountain    fountel
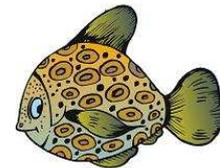
gadget    gadgel
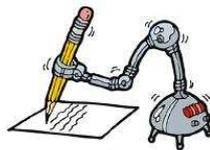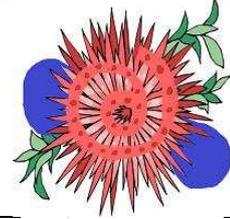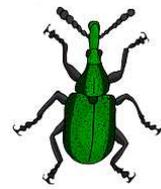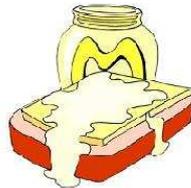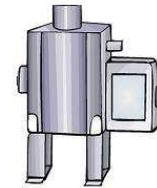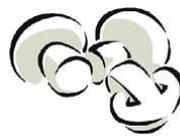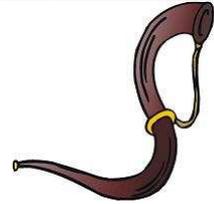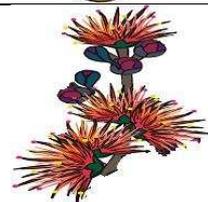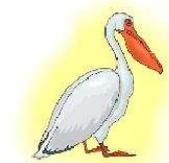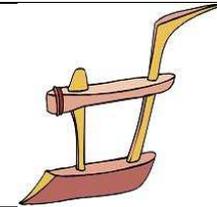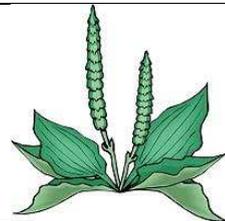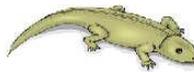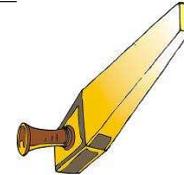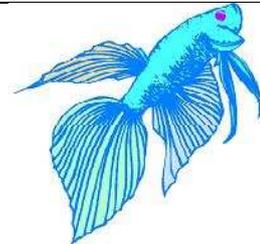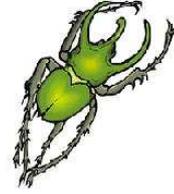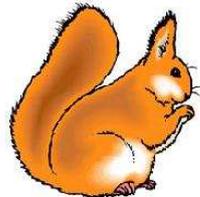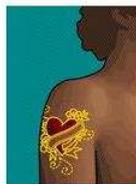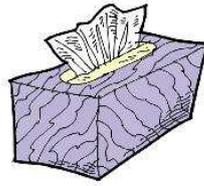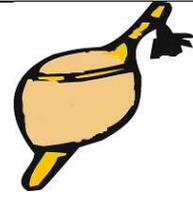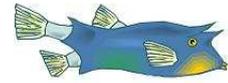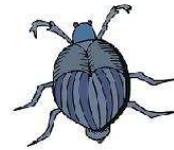
| | | | |
|---|---|---|---|
| graffiti | graffino | | |
| guitar | guitas | | |
| kangaroo | kangami | | |
| lantern | lantobe | | |
| mayonnaise | mayonnote | | |
| mermaid | mermiff | | |
| mushroom | mushrood | | |
| napkin | napkig | | |

| | | | |
|---|---|---|---|
| nugget | nuggev | | |
| octopus | octopum | | |
| onion | oniot | | |
| ornament | ornameld | | |
| parachute | parasheff | | |
| parade | parafe | | |
| parsnip | parsnin | | |
| pelican | pelical | | |

| | | |
|---|---|---|
| penguin  pengwove | | |
| picnic  picnin | | |
| potato  potatuck | | |
| pumpkin  pumpkige | | |
| pyramid  pyramon | | |
| reptile  reptite | | |
| rugby  rugbock | | |
| sergeant  sergeast | | |

| | | |
|---|---|---|
| signature    signatik | | |
| siren    siredge | | |
| skeleton    skeledu | | |
| somersault    somersaumf | | |
| spider    spidet | | |
| squirrel    squirrome | | |
| target    targil | | |
| tattoo    tattefe | | |

| | | | |
|---|---|---|---|
| tissue | tissove |  |  |
| tornado | tornadus |  |  |
| trombone | trombal |  |  |
| tulip | tulode |  |  |
| volcano | volcagi |  |  |
| walnut | walnog |  |  |
| walrus | walrick |  |  |
| yoghurt | yogem |  |  |