# Tracking the Integration of Novel Words into the Mental Lexicon: Evidence from Mouse Movement Trajectories

Andrew P. Lucas[1], Richard J. O'Connor[1], Kevin J. Riggs[1], and Shane Lindsay[1]

[1]School of Psychology and Social Work, University of Hull

## Author Note

The authors have no conflict of interest to declare. This work was funded by an ESRC-matched PhD studentship awarded to Andrew Lucas at the University of Hull. We thank Anna Weighall for sharing the stimuli and training scripts from Weighall et al. (2017).

Correspondence concerning this article should be addressed to Shane Lindsay, School of Psychology and Social Work, University of Hull, Hull, East Riding of Yorkshire, United Kingdom. Email: s.lindsay@hull.ac.uk

**Abstract**

This study investigated lexical competition effects in novel word learning using mouse tracking methodology. We conducted two experiments examining how and when novel word representations become lexicalised, focusing on their ability to engage in competition with existing lexical entries. In Experiment 1, we used mouse tracking to conceptually replicate Weighall et al.'s (2017) eye tracking study, examining lexical competition between newly learnt word–referent pairs and familiar words both immediately after learning and following overnight consolidation. Mixed-effects analyses demonstrated immediate lexical competition effects for novel words. Competition did not differ significantly between words learnt on the previous day and those learnt on the day of testing. Experiment 2 demonstrated that displaying the referent associated with a newly learnt phonological competitor increased immediate lexical competition, even when targets in both conditions had trained phonological neighbours. The temporal dynamics of mouse trajectories revealed that familiar words produced earlier and larger peak competition effects compared to newly learnt words. These findings support models of word learning that allow for rapid lexical integration, while acknowledging that newly formed lexical representations may not immediately exhibit the same processing characteristics as well-established words.

*Keywords*: lexical competition, word learning, mouse tracking, consolidation

# Tracking the Integration of Novel Words into the Mental Lexicon: Evidence from Mouse Movement Trajectories

## Introduction

A central question in language learning is how unfamiliar sounds are transformed into fully functional words. This requires explaining when and how novel word representations achieve full lexical status. Lexicalisation, as conceptualised by Leach and Samuel (2007), comprises two distinct yet complementary processes: configuration and integration (referred to in their account as "engagement"). The configuration phase represents the initial, relatively rapid establishment of a new word form. At this stage, learners can establish multiple mappings, including phonological structure, episodic memory traces, and nascent semantic associations.

This early stage crucially involves mapping the novel form to corresponding referents. In contrast, integration reflects the subsequent, more gradual process through which this newly configured representation becomes integrated within the existing lexical network. Lexical competition serves as a key indicator of successful integration, whereby phonologically similar words compete for recognition during speech processing (Gaskell & Marslen-Wilson, 1997; McClelland & Elman, 1986). For instance, Gaskell and Dumay (2003) demonstrated that after learning the novel word "cathedruke", participants exhibited delayed processing of the familiar base word "cathedral", reflecting interference from shared phonological structures (inhibitory competition between lexical entries).

The complementary learning systems (CLS; McClelland et al., 1995) account of word learning (Davis & Gaskell, 2009; Gaskell, 2024; Lindsay & Gaskell, 2010) provides a framework for understanding this two-part lexicalisation process. According to the CLS model, configuration is primarily hippocampus-dependent, reliant on its rapid encoding abilities.

Integration, however, relies more on gradual neocortical consolidation, creating stable, less context-dependent representations. Initially, form-based phonological lexical competition between novel and known words was interpreted as reflecting primarily the integration stage. Early research (Dumay & Gaskell, 2007) supported this view of lexicalisation as a slow, sleep-dependent consolidation process.

However, a growing body of evidence suggests that pre-sleep form-based lexical competition can occur under particular conditions. Lindsay and Gaskell (2013) demonstrated that novel words can engage in lexical competition without sleep-based consolidation when learning employs spaced exposure with interleaved presentations, suggesting that certain training structures may accelerate lexical integration. A particularly intriguing finding stems from Coutanche and Thompson-Schill (2014), who examined whether the type of learning influences the timeline of lexical integration. Specifically, they tested whether fast mapping, a learning process in which new words are inferred through exclusion learning, accelerates lexical integration compared to explicit encoding, in which learners receive direct instructions to learn new words. They found that while explicit encoding led to stronger declarative memories, only fast mapping resulted in pre-sleep lexical competition. However, this surprising finding has thus far not been replicated in their original orthographic domain (Gurunandan et al., 2023; Lindsay et al., 2027) nor shown to generalise to the phonological domain (Lindsay et al., 2027).

One critical factor in whether lexical competition appears pre- or post-sleep is the measurement method used. Studies showing delayed lexical competition typically rely on single-point measures such as pause detection (Dumay & Gaskell, 2007) or lexical decision (Lindsay & Gaskell, 2013), both of which reveal only the end point of a complex process. In contrast,

methods that capture the temporal dynamics of lexical competition, such as the visual world paradigm (VWP), have revealed evidence for earlier integration effects.

The VWP tracks eye movements to displayed objects during spoken word recognition, revealing the continuous activation and competition between multiple lexical candidates. In their seminal work, Allopenna et al. (1998) demonstrated that upon hearing "beaker," participants fixated on both onset-overlapping objects (e.g., "beetle") and rhyme-overlapping objects (e.g., "speaker"), providing direct evidence of real-time competition between phonologically similar words. This time-sensitive approach has yielded findings that challenge traditional views about lexical integration. Kapnoula et al. (2015; see also Kapnoula & McMurray, 2016) taught participants novel competitors to existing monosyllabic words (e.g., "jod" for "job") and used the VWP to assess competition during recognition of the existing word. They observed robust competition effects immediately after learning, without a consolidation period. Furthermore, the time course of this novel word competition closely resembled that of existing lexical competitors (e.g., "jog").

Additional evidence for immediate lexical competition effects using the VWP comes from Weighall et al. (2017), who explicitly trained participants (both adults and children) on novel object-word pairings. At test, participants viewed four objects on-screen and were instructed to click on one of the familiar objects (e.g., "click on the alien") in the presence of potential competitors. Analyses revealed that participants fixated more on the novel competitor when its word form overlapped with the familiar target (e.g., when "aliet" was on-screen and the target was "alien") than when it did not (e.g., when "baboop" was on-screen and the target was "guitar"). Importantly, these competition effects emerged immediately after learning, with no evidence for enhanced competition following overnight consolidation. They also reported that

competition effects for novel words were weaker than those for familiar words, suggesting that although lexical integration begins rapidly, it may nonetheless continue to develop over time. Their explicit memory data revealed a dissociation between different aspects of word learning, with picture naming accuracy remaining poor (less than 20%) across both days while stem completion performance improved significantly for words learnt the previous day. This pattern suggests that different components of lexical knowledge may follow distinct consolidation trajectories, with some aspects benefiting from sleep while others do not.

Given that the observed time course of word-learning effects depends on methodology (for example, eye tracking compared with pause detection) and on the aspect of lexical knowledge being measured (for example, competition indices of integration versus explicit recall indices of declarative knowledge), lexicalisation is unlikely to be a single, uniform process. Different components of lexical knowledge appear to follow partially independent trajectories that are differentially detectable across methods. Resolving these issues likely requires converging evidence across methodologies and careful interpretation of what each measure reveals about the underlying processes.

**The Present Research**

In response to these questions, we employed mouse tracking as a complementary methodology to replicate the Weighall et al. (2017) eye tracking findings. Methodological considerations are important given the theoretical significance of determining whether lexicalisation requires sleep-based consolidation or can occur rapidly under certain learning conditions. Recent failures to replicate fast mapping effects in lexical integration (Gurunandan et al., 2023; Lindsay et al., 2027) highlight the challenges in establishing robust evidence. As Zwaan et al. (2018) argue, replication attempts are essential for elevating findings from single

observations to scientific evidence, particularly when those findings inform theoretical frameworks. By testing whether Weighall et al.'s findings generalise to mouse tracking, we contribute to a more robust empirical foundation for theory development.

Our approach was twofold. First, we tested whether the immediate lexical competition effect reported by Weighall et al. would replicate using mouse tracking (Experiment 1). Second, we modified their design to disentangle competing explanations of their lexical competition effect.

**Experiment 1: Conceptual Replication Using Mouse Tracking**

Mouse tracking is a well-established method for capturing dynamic cognitive processes that offers insights into decision-making and motor responses (Freeman & Ambady, 2010). This approach records the continuous trajectory of mouse movements as participants make choices, revealing the temporal unfolding of cognitive processes, including the motor planning and execution components of decisions (Koop & Johnson, 2013).

The method has proven particularly effective for studying the temporal dynamics of language processing, including spoken word recognition and lexical competition (e.g., Bartolotti & Marian, 2012; Lin et al., 2023). Following Spivey et al. (2005), we asked participants to select one of two visual referents upon hearing a spoken word, allowing us to examine competition from newly learnt words through their mouse movements. Mouse tracking offers a practical alternative to eye tracking while retaining key strengths of the visual world paradigm, and converging evidence from different approaches can offer more robust support for theoretical claims than relying on a single technique (Denzin, 1978).

**Experiment 2: Disentangling Mechanisms of Lexical Competition**

We addressed a key interpretive ambiguity in Weighall et al.'s original design. This ambiguity concerns the specific mechanism underlying lexical competition effects. Two distinct types of competition can be differentiated:

(1) Purely phonological competition between words with overlapping sound forms. For example, Gaskell and Dumay (2003) found that learning the novel word "cathedruke" slowed recognition of "cathedral" (see also Dumay & Gaskell, 2007). This competition results solely from the phonological overlap between the target familiar words and their novel competitors; it does not rely upon participants activating the meanings of the novel competitors. Indeed, in these studies participants did not learn any possible meaning or referent for the novel words.

(2) Referent-mediated competition that arises when the visual referent of a word that is phonologically related to the target word is present during processing. This type of competition is demonstrated by Spivey et al. (2005), in which the on-screen presence of a picture of a candle interferes with participants' mouse trajectories when instructed to "click on the candy". Such competition requires some form of semantic processing. Specifically, recognising the picture of the candle can activate its meaning and associated lexical label, allowing that label to interfere with processing of the target word "candy". While both mechanisms are theoretically important to understanding lexical processing, they involve different cognitive processes and have distinct implications for models of lexical integration. Crucially, deciding between them speaks to whether semantic information – necessary for referent-mediated competition, but not for purely

phonological competition – influences lexical competition at the very earliest stages of word learning.

The design of Weighall et al. does not allow one to distinguish between these mechanisms. In Competing trials, participants heard a familiar target word (e.g., “baboon”) while seeing both the target and a novel competitor whose label was phonologically related (e.g., “baboop”). In Non-competing trials, they heard a familiar target word (e.g., “athlete”) while seeing the target and an unrelated novel competitor (e.g., “walrick”). Crucially, in Weighall’s design participants had not learnt any novel competitors whose labels were phonologically related to the target words on Non-competing trials. These targets were “untrained” base words in the sense that the novel word derived from each of them (e.g., “athlove” from “athlete”) was never taught.

Competing trials therefore differed from Non-competing trials in two ways. First, the targets themselves differed: participants had learnt novel competitors with labels phonologically related to Competing-trial targets, but not Non-competing-trial targets. Second, the target and novel competitor were phonologically related on Competing trials but not on Non-competing trials. If participants showed greater interference in Competing trials compared to Non-competing trials, this could reflect either phonological competition from having learnt novel words similar to familiar targets, or the specific influence of seeing a referent associated with a phonologically similar word form.

Experiment 2 disentangled these explanations through a critical modification to the target words in Non-competing trials. In contrast to Experiment 1, in which targets on Non-competing trials were untrained base words, Experiment 2 used familiar targets whose corresponding novel

phonological neighbour had been trained. Consequently, participants learnt a novel competitor whose label overlapped phonologically with each target, exactly as in Competing trials.

Hearing the target could therefore activate its learnt neighbour on both trial types, even when that neighbour's referent was absent. On a Non-competing trial, for example, hearing "baboon" could activate "baboop" although the display showed only the baboon and an unrelated learnt object, much as the novel words in Kapnoula et al. (2015) competed with familiar words despite having no referent on-screen. If participants showed greater interference in Competing than Non-competing trials in Experiment 2, this additional interference could not be explained by phonological competition alone. Such competition is predicted to arise equally in both Competing and Non-competing trials. Greater interference in Competing trials would instead indicate referent-mediated (semantic) competition, caused by the presence of an object with a label phonologically similar to the target in Competing trials. Conversely, a reduction in the competition effect from Experiment 1 to Experiment 2 would suggest that phonological competition contributed to the interference effect in Experiment 1. This manipulation therefore tests whether the observed lexical competition effect includes a referent-mediated contribution and whether it involves access to the semantic representation of the newly learnt word.

**Research Hypotheses**

The following hypotheses guided our investigation of lexical competition effects across both experiments. In all cases, lexical competition is revealed by differences in mouse movements between Competing and Non-competing trials.

***Familiar Word Competition Test***

We examined whether mouse tracking would reveal the well-established lexical competition effects between familiar words (e.g., "candy/candle"). This test served as our

baseline measure and methodological validation, building on research using both eye tracking (Allopenna et al., 1998) and mouse tracking paradigms (Spivey et al., 2005) that consistently demonstrate competition between phonologically similar words. We also tested whether the magnitude of this effect was larger than the corresponding effect for novel words.

***Immediate Novel Word Competition Test***

We investigated whether newly learnt novel words would show lexical competition effects immediately after learning, without overnight consolidation. By examining whether participants' mouse movements were drawn towards pictures of novel competitors (e.g., "biscal") when they heard a familiar target word (e.g., "biscuit"), we could determine whether Weighall et al.'s finding of immediate competition effects would replicate using mouse tracking methodology.

***Consolidation Effect Test***

We explored whether overnight consolidation enhanced lexical competition effects for novel words. Weighall et al. found no significant increase in competition strength after sleep for adults, but other studies suggest that consolidation bolsters lexical integration (Dumay & Gaskell, 2007; Henderson et al., 2012). By comparing mouse tracking data for words learnt the previous day (Day 1) with those learnt just before testing (Day 2), we assessed whether this method might reveal consolidation benefits not apparent in prior eye tracking studies. Additionally, we examined explicit memory using stem completion (e.g., completing "bis–" as "biscal") and picture naming tasks (recalling novel object names). Weighall et al. reported a dissociation in which stem completion improved after sleep, suggesting phonological consolidation, while picture naming remained poor across both days. Replicating this pattern

would support the idea that different facets of word learning follow distinct consolidation trajectories.

### *Temporal Dynamics Test*

We also examined the temporal unfolding of lexical competition for both familiar and novel words because it could reveal differences between established and newly learnt words that might not be captured by overall competition magnitude measures. Using growth curve analysis to model the dynamics of competition, we tested whether competition from novel words developed more slowly than from familiar words, as observed by Weighall et al.

### *Mapping Specificity Test (Experiment 2)*

Finally, we asked whether specific form–referent mappings contribute to immediate competition beyond general phonological similarity. By modifying the non-competing trials and addressing the confounds noted above, we tested whether the presence of the learnt referent contributes to competition beyond phonological overlap alone.

## Experiment 1

### Methods

### *Participants*

We recruited 59 participants (mean age = 22.5 years, *SD* = 6.54; 13 were male) who volunteered their time in exchange for a £16 Amazon voucher or course credit. For both experiments, sample size was determined by participant availability within fixed testing windows; each experiment exceeded the adult sample in Weighall et al. (2017; *N* = 42). All were fluent English speakers. While 48 participants were right-handed, all participants normally used a mouse with their right hand. No participants had confounding disorders of sleep, learning, language processing, or memory, and all had normal (or corrected-to-normal) vision and hearing.

They were tested in a quiet laboratory environment in accordance with procedures approved by the Faculty of Health Sciences Ethics Committee at the University of Hull.

***Design, Procedure, and Stimuli***

The experiment took place over two days, with participants trained on one set of novel words and objects on the first day, and a second set trained on the second day. Training on the second day was immediately followed by testing: the mouse tracking task and the two recall tasks (stem completion and picture naming). An overview of the design and trial structure in Experiment 1 is shown in Figure 1.

For the training phase, we used the same 72 matched pairings of audio recordings of novel words and novel objects as Weighall et al. (2017; see their Methods for details), comprising three lists of 24 word-object pairs. Each list contained known base words (e.g., "angel") and novel competitors derived from them (e.g., "angesh") by altering the base word at its uniqueness point. Participants were trained on the word-object pairs from two of the three lists, one list on each day. The third list remained untrained. All pictures were standardised to 300 × 300 pixels.

The mouse tracking task followed the Spivey et al. (2005) methodology, with both Competing trials (where objects with phonologically overlapping labels appeared together) and Non-competing trials (where objects had no phonological overlap). Two objects were presented on each trial across three categories:

**Familiar word trials** (20 trials) used 20 familiar words as targets, with each appearing on-screen with a distractor object. These familiar words were matched on verbal and written frequency, concreteness, familiarity, and imageability according to the MRC Psycholinguistic Database (Wilson, 1988). The corresponding pictures (sourced from clipart.com) achieved over

80% dominant-name agreement based on ratings from 10 adults (Weighall et al., 2017). In Competing trials (10), participants heard a target word (e.g., "Click on the candle") while viewing both the target and an object whose label was a phonologically similar competitor (e.g., a picture of "candy"). In Non-competing trials (10), participants viewed the target alongside an object whose label was not phonologically related to the target word (e.g., "stamp").

**Novel word trials** (48 trials) examined competition effects between the base words and novel words. These trials were divided into:

- **Day 1 competing trials** (12): Participants heard a known target base word (e.g., "Click on the alien") and saw the target object alongside a novel competitor object whose label ("aliet") was phonologically derived from the target word and was learnt on Day 1.
- **Day 2 competing trials** (12): Participants heard a known target base word (e.g., "Click on the angel") and saw the target object alongside a novel competitor object whose label ("angesh") was phonologically derived from the target word and was learnt on Day 2.
- **Non-competing trials** (24): Participants heard a known target base word (e.g., "Click on the athlete"), taken from the untrained list, and saw the target object alongside a novel object from a trained list whose label had no phonological overlap (e.g., "lantobe") with the target word. Critically, participants had never learnt the novel competitor derived from the target (e.g., "athlove") or seen its referent, so these trials contained no newly learnt phonological competitor of the target.

**Filler trials** (20 trials) used 20 phonotactically legal nonwords in British English, selected from the Graded Nonword Reading Test (Snowling et al., 1996) and the Blending Nonwords subtest of the Comprehensive Test of Phonological Processing (Wagner et al., 1999).

Participants heard these untrained novel words (e.g., “Click on the balras”) while viewing a “super-novel” object (not seen before) alongside a familiar object (e.g., a kettle). These trials were included to discourage participants from developing a strategy of always ignoring novel objects, since the experimental trials always required selecting familiar objects. Although participants had not learnt associations between these novel objects and labels, the principle of mutual exclusivity would lead them to select the novel object when hearing an unfamiliar word.

The 72 base words were counterbalanced across three stimulus lists. In each list, 24 items served as targets in competing trials (12 learnt on Day 1 and 12 on Day 2). A second set of 24 items did not appear as targets; instead, their novel referents acted as competitors in non-competing trials. The final 24 items remained untrained and were later used as targets in non-competing trials, paired with novel objects whose labels shared no phonological overlap. Across the three lists, each set of 24 base words served once as the untrained set.

***Training***

Participants completed a structured training task designed to establish novel word-object associations. The design followed the training protocol used in Weighall et al. (2017), which was based on methodology established in previous studies (Henderson et al., 2012, 2013), combining phonics-based methods with a two-alternative forced-choice (2-AFC) procedure.

The training was administered via E-Prime 2.0 (Schneider et al., 2002). Blocks 1–3 were each repeated twice, with items presented in random order. Block 4 comprised three repetitions of the 24-item two-alternative forced-choice task. Training occurred on two consecutive days, with participants learning a new list of words on the second day.

Blocks 1–3 featured novel objects displayed on-screen for 2 seconds with a 500 ms inter-trial interval. In Block 1 (whole word listen and repeat), participants heard a novel word (e.g.,

"baboop") whilst viewing its corresponding novel object and were instructed to repeat the entire word aloud. Block 2 (initial segmentation) required participants to hear the whole word but repeat only the first syllable (e.g., "ba-"). In Block 3 (final segmentation), participants heard the entire word but repeated only the final syllable (e.g., "boop"). These blocks were designed to draw attention to the phonological structure of the novel words.

Block 4 consisted of a two-alternative forced-choice (2-AFC) task with feedback. On each trial, participants viewed two novel objects from the training list for 2 seconds whilst hearing one of their labels. Participants indicated whether the word matched the left or right object by pressing '1' or '9', respectively. After responding, the correct object appeared alone for 2 seconds with the novel word replayed as feedback. This block comprised 72 trials (24 novel words × 3 repetitions). The training-accuracy check reported below uses the final repetition of this block (trials 49–72).

Through this protocol, participants were exposed to each novel word-object pair 12 times: twice in each of Blocks 1–3, plus six additional exposures in Block 4 (once during each trial and once during feedback, with each trial repeated three times). The entire training session lasted approximately 30 minutes. After completing Day 1 training, participants returned approximately 24 hours later to complete identical training procedures with a new list of novel words.

***Mouse Tracking Task***

Following training on the second day, participants immediately performed the mouse tracking task, which was run using the MouseTracker program (Freeman & Ambady, 2010). Our mouse tracking design was based on Spivey et al. (2005).

For each trial, object images appeared simultaneously in the top-left and top-right corners of the screen. Participants heard a word over headphones via a carrier phrase (e.g., “Click on the bacon”) and were instructed to move the mouse and click on the corresponding picture. Left- and right-side targets were counterbalanced, and stimuli were presented in a random order for each participant. Participants were told to begin moving as quickly as possible from the start of a trial and not to wait until they had heard the target word. We enforced a 450 ms initiation-time threshold. If movement had not begun by 450 ms, an on-screen prompt instructed participants to start moving faster. The experimental trials were preceded by 16 practice trials using different stimuli and words.

### *Explicit Memory Tasks*

After the mouse tracking task, participants completed two explicit memory tasks adapted from Weighall et al. (2017). In the stem completion task, participants heard the initial syllable of each trained novel word (e.g., “bis-”) and were asked to complete the cue aloud with one of the novel words learnt during training. In the picture naming task, participants saw each trained novel object and were asked to name it aloud using its trained novel label. Day 1 and Day 2 items were randomly intermixed.

### *Analysis Methods and Data Processing*

All analyses were performed in R (R Core Team, 2021). The *tidyverse* (Wickham et al., 2019) and *mousetrap* (Kieslich & Henninger, 2017; Wulff et al., 2025) packages were used for data processing and visualisation. Mixed-effects models were fitted with *lme4* (Bates et al., 2015) and *lmerTest* (Kuznetsova et al., 2017), contrasts were estimated with *emmeans* (Lenth, 2022), and *performance* was used to summarise model fit (Lüdecke et al., 2021).

For mixed-effects models, we report effect estimates, test statistics, *p* values, and confidence intervals for planned contrasts. Linear-model effects are reported as unstandardised estimates with standard errors, and binomial-model effects are reported as odds ratios. For Day contrasts in binomial models, odds ratios compare Day 1 with Day 2; values greater than 1 indicate higher model-estimated accuracy for Day 1 items, and values below 1 indicate higher model-estimated accuracy for Day 2 items.

For random-effects selection, we compared small, design-driven candidate sets that differed only in random-effects structure. Among candidates that converged and were non-singular, we selected the model with the lowest AIC (Matuschek et al., 2017).

Binary accuracy outcomes were analysed using binomial mixed-effects models. Training models were fitted separately for each experiment, and explicit memory models were fitted separately for each task and experiment, with Day as the fixed effect in each model. The candidate set comprised crossed participant and item Day slopes, participant Day slopes with item intercepts, participant intercepts with item Day slopes, and crossed participant and item intercepts. The model selection process led to participant and item intercepts for the Experiment 1 training model, and participant Day slopes with item intercepts for the Experiment 2 training model. For the explicit memory models, the model selection process led to participant and item Day slopes, except for Experiment 1 picture naming, where it led to participant Day slopes and item intercepts.

**Preprocessing.** Raw mouse tracking data were resampled at 20 ms to standardise temporal resolution, spatially aligned so starts were at (0, 0) and responses ended at (8, 9), and left-target trials were mirrored so the target lay on the right. Trajectories were time-locked to target-word onset within the carrier phrase (“Click on the X”); onset timestamps were rounded to

the nearest 20 ms and pre-onset samples were discarded. The resulting trajectories were time-normalised to 101 steps. These preprocessing steps apply to all analyses below.

***Maximum Deviation Analysis***

We used the signed maximum absolute deviation (MAD), as returned by *mousetrap*'s *mt_measures*, as the primary scalar outcome. This measure identifies the point of maximum absolute deviation from the direct path while retaining its sign. After mirroring left-target trials so that the target always lay on the right, positive values indicate peak deviation towards the distractor side of the display and negative values indicate peak deviation towards the target side. On Competing trials, the distractor side is the phonological-competitor side. Thus, MAD corresponds to the signed perpendicular deviation from the ideal start-target line at the sample with the largest absolute deviation, measured in normalised screen units.

For each experiment, we fitted linear mixed-effects models to signed MAD with fixed effects of Word type, Competition, and their interaction. The candidate set comprised crossed participant and item Competition slopes, participant Competition slopes with item intercepts, participant intercepts with item Competition slopes, and crossed participant and item intercepts. The model selection process led to participant intercepts and item Competition slopes for Experiment 1, participant and item Competition slopes for Experiment 2, and participant and item Competition slopes for the between-experiment novel word model. Planned contrasts from estimated marginal means tested Familiar word competition, Day 2 novel word competition, the Familiar versus average novel word competition difference, the Day × Competition contrast within novel words, and the Experiment × Competition contrast across experiments. Positive Competing − Non-competing estimates indicate greater attraction toward the phonological competitor.

### *Growth Curve Analysis*

To examine temporal dynamics, we conducted a growth curve analysis (GCA; Mirman, 2016) on the full time-series of signed trajectory deviation at each sample, computed with *mousetrap::mt_deviations()* after the preprocessing described above. Competition was summarised as the moment-by-moment Competing − Non-competing difference in signed deviation, with positive values indicating greater lateral attraction toward the competitor.

We fitted linear mixed-effects growth-curve models at the trial level. Time was represented by an orthogonal polynomial basis of order 3 over the 101 time-normalised samples, yielding linear, quadratic, and cubic components. Fixed effects were Word type, Competition, the three time terms, and their interactions. Categorical predictors used treatment contrasts with Familiar words as the reference level for Word type and Competing trials as the reference level for Competition. In figures and interpretive summaries, we report competition as the Competing − Non-competing difference.

The retained model included the linear, quadratic, and cubic time terms, Word type, Competition, and their interactions, with participant and item intercepts:

Trajectory deviation ~ (linear time + quadratic time + cubic time) × Word type × Competition + (1 | participant) + (1 | item)

For GCA, the candidate set compared a participant-only random-effects structure with a structure that included participant and item intercepts, and the model selection process led to the structure with participant and item intercepts. For Figure 4, we plot observed competition difference curves from participant-mean time courses alongside fixed-effect model predictions by Word type over time.

Model fit was summarised using Nakagawa's marginal and conditional $R^2$ (Nakagawa & Schielzeth, 2013), computed with *performance::r2_nakagawa()* (Lüdecke et al., 2021). *p* values for fixed effects were obtained with Satterthwaite's approximation as implemented in *lmerTest* (Luke, 2017).

## Results

### *Training*

Training accuracy was summarised from the final repetition of the 2-AFC block (trials 49 –72 of the 72-trial block). Participants identified the correct target 96.3% of the time on Day 1 (SD = 5.1%) and 97.2% of the time on Day 2 (*SD* = 3.7%). There was no significant difference between days in the binomial mixed-effects model, OR = 0.73, 95% CI [0.48, 1.12], $z = -1.45$, $p = .146$.

### *Explicit Memory Tests*

For stem completion, we found better recall for words learnt on Day 1 ($M = .37$, $SD = .18$) compared to words learnt on Day 2 ($M = .25$, $SD = .16$). This difference was statistically significant in the binomial mixed-effects model, OR = 2.01, 95% CI [1.56, 2.57], $z = 5.50$, $p < .001$.

Performance was lower for the picture naming task, with poor performance for both Day 1 items ($M = .20$, $SD = .18$) and Day 2 items ($M = .21$, $SD = .18$), with no significant difference in the binomial mixed-effects model, OR = 0.88, 95% CI [0.64, 1.22], $z = -0.76$, $p = .450$.

These results are shown in Figure 2.

### *Mouse Tracking Task*

We applied a filtering pipeline that avoids trimming on the dependent measure. For Experiment 1, error trials were removed (2.1%). Within each participant × word type, initiation-

time outliers were removed using robust z-scores (|z| > 3), eliminating 69 trials. Two participants were excluded for atypical starting positions (mean |x|-at-start z > 3). Finally, an absolute trial-level starting-position filter (|x| > 10 px at movement onset) removed 9 trials (0.24%). A quality check for excessive horizontal reversals (x-flips ≥ 10) removed no additional trials.

**Maximum Deviation Analysis.** Descriptive statistics for this variable are shown in Table 1.

The mixed-effects model of signed MAD revealed significant main effects of Competition, $F(1, 91.7) = 26.98$, $p < .001$, and Word type, $F(2, 162.9) = 13.32$, $p < .001$. A planned contrast showed that competition for Familiar words was significantly higher than the average competition estimate for novel words, $b = 0.379$, $SE = 0.182$, 95% CI [0.018, 0.740], $t(120.4) = 2.08$, $p = .040$. Because MAD preserves the direction of peak deflection, the Competing > Non-competing contrast reflects stronger distractor-directed attraction in competing displays. Below, we examine how these results address our key theoretical questions using planned contrasts.

***Familiar Word Competition Test.*** A planned contrast for familiar words confirmed robust competition, $b = 0.663$, $SE = 0.160$, 95% CI [0.346, 0.980], $t(115.7) = 4.15$, $p < .001$. This strong effect validates our methodology and replicates the well-established finding that phonologically similar familiar words compete for recognition, consistent with previous mouse tracking research (Spivey et al., 2005).

***Immediate Novel Word Competition Test.*** A planned contrast comparing Competing versus Non-competing conditions for Day 2 novel words, learnt on the same day as testing, showed significant competition, $b = 0.245$, $SE = 0.121$, 95% CI [0.005, 0.484], $t(114.0) = 2.02$, $p$

= .045. This indicates that competition occurred without overnight consolidation, replicating the immediate competition effects reported by Weighall et al. (2017) using a different methodology.

***Consolidation Effect Test.*** To examine whether overnight consolidation enhanced competition for novel words, we used the planned Day × Competition contrast within novel words. This analysis showed that the Day 2 − Day 1 difference in competition was small and not significant, $b$ = -0.078, $SE$ = 0.166, 95% CI [-0.405, 0.248], $t$(347.4) = -0.47, $p$ = .637. Thus, competition was not significantly greater for words learnt on the previous day than for words learnt on the day of testing, consistent with the findings of Weighall et al. (2017) in adults.

### *Growth Curve Analysis*

Whereas MAD summarises the single largest signed deviation on a trial, the growth-curve analysis models the full signed deviation series across the trajectory. The two analyses therefore address complementary questions about competition magnitude and time course.

The primary GCA for Experiment 1 showed significant linear ($\beta$ = −109.737, $SE$ = 3.010, $t$ = −36.46, $p$ < .001), quadratic ($\beta$ = −455.885, $SE$ = 3.010, $t$ = −151.48, $p$ < .001), and cubic ($\beta$ = 60.099, $SE$ = 3.010, $t$ = 19.97, $p$ < .001) time components. Familiar words showed a clear competition effect (Non-competing relative to Competing, $\beta$ = −0.212, $SE$ = 0.064, $t$ = −3.30, $p$ = .001). Interactions with Word type were expressed primarily in curvature, with significant interactions involving the quadratic and cubic terms. The model-estimated competition differences were ≈ 0.21 for Familiar, 0.12 for Day 1, and 0.08 for Day 2. The model explained ≈ 18% (marginal $R^2$) and ≈ 29% (conditional $R^2$) of the variance.

To identify peak competition effects in the raw trajectories, we located the point of maximum difference between Competing and Non-competing conditions for each word type. In Experiment 1, the largest deviation difference for Familiar words occurred at 53% of the

movement trajectory (magnitude = 0.67), while Day 1 and Day 2 words showed later peaks at 63% (magnitude = 0.33) and 69% (magnitude = 0.28), respectively. Percentages refer to the 101-point time-normalised trajectory following target-word onset. Competition from familiar words peaked earlier in the mouse movement and reached a larger peak than competition from newly learnt words.

## Experiment 2

### Methods

#### *Participants*

Sixty-two participants (mean age = 28 years, *SD* = 10.7; 24 were male), who did not take part in Experiment 1, volunteered their time in exchange for a £16 Amazon voucher or course credit. All were fluent English speakers. All participants reported using a mouse with their right hand. No participants reported having disorders of sleep, learning, language processing, or memory, and all had normal or corrected-to-normal vision and hearing.

#### *Design, Procedure, and Stimuli*

The overall design matched Experiment 1 (see Figure 1), except for a key modification to the Non-competing trials in the mouse tracking task. In Experiment 2, Non-competing trials used familiar targets for which the corresponding novel phonological neighbour had been trained, as in Competing trials. The visible distractor was an unrelated trained novel object with a label derived from a different familiar base word. For example, when the target word was “baboon”, participants had previously learnt the phonologically overlapping novel label “baboop” for one novel object. The display contained the baboon and an unrelated distractor, such as the referent for “pumpkige” (derived from “pumpkin”). This meant that Competing and Non-competing trials were matched for whether the target had a newly learnt phonological neighbour but differed in

whether the display contained that neighbour's referent. This modification was designed to differentiate between the two types of competition that may explain the effects seen in Experiment 1 and Weighall et al. (2017). These were a purely phonological effect caused by learning a new phonologically similar word, or a specific form–referent mapping (semantic) effect caused by the presence of a referent associated with a phonologically related word. Competing trials, Familiar word trials, and super-novel filler trials remained identical to Experiment 1.

## Results

### *Training*

Participants showed similar performance to Experiment 1, with 94.1% accuracy on Day 1 (*SD* = 8.3%) and 94.4% on Day 2 (*SD* = 7.0%). The binomial mixed-effects model showed no significant Day effect, OR = 1.25, 95% CI [0.72, 2.17], $z = 0.80$, $p = .425$. The high accuracy rates across both experiments indicate that participants encoded the novel word–referent pairings during the training sessions.

### *Explicit Memory Tests*

For stem completion, we again found significantly better recall for words learnt on Day 1 (*M* = .45, *SD* = .23) compared to words learnt on Day 2 (*M* = .30, *SD* = .18), OR = 2.23, 95% CI [1.63, 3.06], $z = 5.02$, $p < .001$. This replicates the Day 1 recall advantage observed in Experiment 1, as shown in Figure 2.

Picture naming results mirrored those of Experiment 1, with consistently low performance for both Day 1 items (*M* = .19, *SD* = .15) and Day 2 items (*M* = .20, *SD* = .18). No significant difference emerged in the binomial mixed-effects model, OR = 0.77, 95% CI [0.54, 1.11], $z = -1.41$, $p = .159$.

### *Mouse Tracking Task*

We used the same filtering pipeline as Experiment 1, with the following Experiment 2 outcomes: we removed 153 error trials (3.6%), then applied robust z-score filtering ($|z| > 3$) to initiation times within each participant × word type, removing 14 additional trials (0.3%). One participant was excluded for atypical starting positions (mean |x|-at-start $z > 3$), leaving 61 participants for mouse tracking analyses. An absolute trial-level starting-position filter ($|x| > 10$ px at movement onset) removed 8 trials (0.20%). A quality check on horizontal reversals (x-flips ≥ 10) removed 2 additional trials.

**Maximum Deviation Analysis.** Similar to Experiment 1, we analysed maximum deviation to assess lexical competition effects in Experiment 2. The mixed-effects model revealed significant main effects of Competition, $F(1, 87.1) = 14.20$, $p < .001$, and Word type, $F(2, 166.9) = 6.02$, $p = .003$. The estimated competition effect was numerically larger for Familiar words than for novel words, but this contrast was not significant, $b = 0.238$, $SE = 0.230$, 95% CI [-0.220, 0.696], $t(86.3) = 1.03$, $p = .304$. The results for our contrasts of interest are shown below.

***Familiar Word Competition Test.*** A planned contrast comparing Competing versus Non-competing conditions for Familiar words confirmed robust competition, $b = 0.606$, $SE = 0.193$, 95% CI [0.223, 0.989], $t(98.9) = 3.14$, $p = .002$. This replicates the familiar word effect from Experiment 1 and provides a baseline lexical competition effect for the Experiment 2 task.

***Immediate Novel Word Competition Test.*** A planned contrast comparing Competing versus Non-competing conditions for Day 2 novel words, learnt on the same day as testing, revealed significant competition, $b = 0.473$, $SE = 0.159$, 95% CI [0.159, 0.787], $t(135.2) = 2.98$, $p = .003$.

This finding is particularly important in Experiment 2, where Non-competing trials were modified so that their targets also had trained phonological neighbours. The persistence of immediate competition effects in this more stringent design demonstrates that immediate lexical competition has a referent-dependent component that cannot be explained by learning new phonological neighbours alone.

***Consolidation Effect Test.*** To examine whether overnight consolidation enhanced competition for novel words, we used the planned Day × Competition contrast within novel words. This analysis showed that the Day 2 − Day 1 difference in competition was small and not significant, $b = 0.210$, $SE = 0.142$, 95% CI [-0.068, 0.488], $t(3762.5) = 1.48$, $p = .139$. Thus, competition was not significantly greater for words learnt on the previous day than for words learnt on the day of testing, consistent with the findings of Weighall et al. (2017) in adults.

***Growth Curve Analysis***

The primary GCA for Experiment 2 revealed the same qualitative temporal pattern as Experiment 1, with significant linear ($\beta = -107.733$, $SE = 2.964$, $t = -36.34$, $p < .001$), quadratic ($\beta = -507.541$, $SE = 2.964$, $t = -171.22$, $p < .001$), and cubic ($\beta = 60.613$, $SE = 2.964$, $t = 20.45$, $p < .001$) components. Familiar words again showed a clear competition effect (Non-competing relative to Competing, $\beta = -0.220$, $SE = 0.075$, $t = -2.94$, $p = .004$). Interactions with Word type were again expressed in both intercept-level competition estimates and time course shape. The model-estimated competition differences were ≈ 0.22 for Familiar, 0.11 for Day 1, and 0.21 for Day 2. The model explained ≈ 24% (marginal $R^2$) and ≈ 34% (conditional $R^2$) of the variance.

To identify peak competition effects in the raw trajectories for Experiment 2, we again located the point of maximum difference between Competing and Non-competing conditions for each word type. This analysis revealed that the largest deviation difference for Familiar words

occurred at 56% of the movement trajectory (magnitude = 0.55), while Day 1 and Day 2 words showed peaks at 65% (magnitude = 0.30) and 61% (magnitude = 0.48), respectively. As in Experiment 1, competition from Familiar words peaked earlier in the mouse movement and reached a larger peak than competition from novel words.

### *Mapping Specificity Test: Comparing Experiment 1 and Experiment 2 Effects*

To directly test whether the competition effects observed in Experiment 2 differed from those in Experiment 1, we conducted a between-experiments comparison. This analysis complements the within-experiment test of our Mapping Specificity hypothesis by assessing whether competition was reduced when targets in both conditions had trained phonological neighbours. If competition effects are primarily driven by specific form–referent mappings rather than phonological similarity, we would expect similar competition effects across experiments despite the design change to Non-competing trials.

We conducted a mixed-effects analysis on maximum deviation, focusing on novel words only, with Experiment, Day, Competition, and their interactions as fixed effects. This analysis revealed a significant main effect of Experiment, $F(1, 119.9) = 7.33$, $p = .008$, with higher overall deviation for novel words in Experiment 2 than in Experiment 1. There was also a significant main effect of Competition, $F(1, 70.7) = 12.50$, $p < .001$, and no main effect of Day, $F(1, 2635.6) = 0.52$, $p = .470$. The planned Experiment × Competition contrast was not significant, $b = 0.004$, $SE = 0.127$, 95% CI [-0.248, 0.255], $t(176.6) = 0.03$, $p = .978$. All other interactions were nonsignificant.

## General Discussion

Our study replicates and extends Weighall et al. (2017) by showing immediate competition between newly learnt words and phonologically similar familiar words. Across two

mouse tracking experiments, competition emerged shortly after learning. The Experiment 2 pattern is consistent with a referent-mediated interpretation of this interference. In both Competing and Non-competing trials, the target had a newly learnt phonological neighbour; the only difference was whether the competitor's referent appeared on-screen. Greater interference occurred when that referent was visible, consistent with specific form–referent mappings contributing to the effect beyond phonological overlap. The persistence of competition suggests that learners rapidly establish lexical representations with both phonological and referential content.

These mappings were strong enough to produce measurable competition despite picture naming accuracy of around 20%. This suggests that partial form–meaning links can shape online processing even when overt recall is poor, perhaps because visible referents preactivate candidate lexical representations (Altmann & Kamide, 1999; Huettig & McQueen, 2007). Across the two experiments, the findings support a referent-mediated account of competition in which the presence of the specific learnt referent associated with a phonologically overlapping label contributes to competition beyond phonological overlap alone.

The immediacy of these competition effects is consistent with visual world paradigm findings (e.g., Kapnoula et al., 2015) but contrasts with pause-detection and lexical-decision tasks, which typically reveal competition only post-sleep (Dumay & Gaskell, 2007; Henderson et al., 2012). This methodological discrepancy likely reflects different facets of lexical integration. Mouse tracking and eye tracking paradigms present visual referents before auditory input, potentially pre-activating word representations (Altmann, 2004), and provide continuous measures sensitive to subtle dynamics often missed by discrete reaction time tasks. McMurray et al. (2016) outline how different paradigms may tap distinct processing levels, with more

automatic effects emerging post-consolidation. Our findings show that referent-mediated competition can emerge before overnight consolidation, consistent with the view that lexicalisation develops in stages. Early competition appears to be facilitated by the visible referent, whereas later consolidation may reduce reliance on that context, enabling competition effects to be observed post-sleep in tasks that do not present visible referents, such as pause-detection and lexical-decision tasks.

While demonstrating rapid integration, our results also reveal differences between newly formed and established lexical representations. The observed competition curves peaked earlier for familiar words (at 53–56% of the time-normalised trajectory) than for novel words (at 61–69%), with larger peak differences for familiar words. This temporal signature indicates that, while novel words engage in referent-mediated competition, their processing characteristics differ. Studies of artificial lexicons also show that the dynamics of competition change with learning. Magnuson et al. (2003) found that additional training produced a more familiar-like pattern, with competition increasingly reflecting the unfolding speech signal. Creel et al. (2008) observed relatively slow and extended competition among newly learnt words, although differences in the materials limited comparison with familiar words.

Our mouse tracking data extend this picture, showing robust competition from novel words but later and smaller peaks than those observed for familiar words, a pattern consistent with partially integrated lexical entries. These relative timing differences suggest that newly learnt words engage in competition less automatically than established words (McMurray et al., 2016; Shtyrov, 2012; Tham et al., 2015). The temporal shift in peak competition for novel words supports this view, and the consistency of this pattern across both experiments adds to its reliability.

Regarding overnight consolidation, neither experiment detected a reliable difference in competition between words trained on Day 1 and Day 2. In contrast, stem-completion accuracy was higher for Day 1 items, replicating Weighall et al. (2017). This pattern suggests that different facets of word learning may benefit unequally from consolidation, although the comparison between learning days combines the effects of sleep and elapsed time. Prior work links sleep-spindle activity to the overnight integration of novel word forms (Tamminen et al., 2010) and shows that targeted memory reactivation during slow-wave sleep can enhance vocabulary retention (Schreiner & Rasch, 2017). The lack of a detectable overnight boost to competition here may reflect the still-incomplete status of new form-meaning links or the need for a longer consolidation interval, as has been reported for cross-modal lexicalisation after 24 hours or more (Bakker et al., 2014).

A crucial question arising from our findings is whether increased training would alter the observed patterns of lexical competition and consolidation. The present dissociation between robust referent-mediated competition and poor explicit recall suggests that while the training provided was sufficient to influence online processing, it may have been insufficient for strong declarative memories. Future research could systematically manipulate training intensity to determine whether more robust declarative representations lead to enhanced competition effects, potentially eliminating the observed differences between novel and familiar words. This approach could help determine whether the observed timing differences stem from fundamental characteristics of new lexical entries (such as differences in neural representation or processing routes) or simply insufficient exposure.

Our results are compatible with the Complementary Learning Systems model of word learning (Davis & Gaskell, 2009; Gaskell, 2024; Lindsay & Gaskell, 2010), which posits initial

hippocampally dependent representation of novel words through pattern-separated, context-sensitive encoding followed by gradual neocortical consolidation. Gaskell's (2024) updated framework emphasises lexicalisation as graded rather than all-or-nothing. This account fits both the pre-sleep referent-mediated competition observed here and evidence that different aspects of word learning follow distinct temporal trajectories (McMurray et al., 2016). This perspective, emphasising continuous integration, aligns with connectionist models that highlight gradual strengthening of connections between phonological, semantic, and orthographic representations (Gaskell & Marslen-Wilson, 2002; Smith et al., 2021). Novel word competition emerged immediately but was weaker and peaked later than familiar word competition, consistent with gradual entrenchment.

In conclusion, newly learnt word–referent mappings can influence recognition of spoken words shortly after training, although competition from familiar words peaked earlier in the mouse movement and reached larger peaks. The convergence with Weighall et al.'s eye tracking findings suggests that this immediate effect is not specific to one visual world measure and illustrates the utility of mouse tracking for capturing graded competition dynamics. Experiment 2 demonstrated greater competition when the displayed distractor was the referent of the newly learnt phonological competitor than when it was an unrelated trained referent, even though targets in both conditions had trained phonological neighbours. One possibility is that mouse tracking and the visual world paradigm reveal early competition because visible referents support activation of newly learnt mappings, whereas discrete tasks such as pause-detection become sensitive only as those mappings consolidate. This account remains to be tested directly. The immediate online competition and better stem completion for previous-day items suggest that these measures capture partly distinct aspects of learning. Future work should test whether

additional training strengthens novel-word representations sufficiently for lexical competition to emerge without the support of visible referents.

## Open Practices Statement

The data and analysis code for both experiments are available at https://osf.io/dbyz9/.

## 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 from sound to meaning. *Journal of Memory and Language, 38*(4), 419–439.

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

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*(3), 247–264.

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.

Bartolotti, J., & Marian, V. (2012). Language learning and control in monolinguals and bilinguals. *Cognitive Science, 36*(6), 1129–1147.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Coutanche, M. N., & Thompson-Schill, S. L. (2014). Fast mapping rapidly integrates information into existing memory networks. *Journal of Experimental Psychology: General, 143*(6), 2296–2303.

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

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

Denzin, N. K. (1978). The research act: A theoretical introduction to sociological methods (2nd ed.). McGraw-Hill.

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

Freeman, J. B., & Ambady, N. (2010). MouseTracker: Software for studying real-time mental processing using a computer mouse-tracking method. *Behavior Research Methods, 42*(1), 226–241.

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

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

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

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

Gurunandan, K., Cooper, E., Tibon, R., Henson, R. N., & Greve, A. (2023). No evidence of fast mapping in healthy adults using an implicit memory measure: Failures to replicate the

lexical competition results of Coutanche and Thompson-Schill (2014). *Memory, 31*(10), 1320–1339.

Henderson, L. M., Weighall, A. R., Brown, H., & Gaskell, M. G. (2012). Consolidation of vocabulary is associated with sleep in children. *Developmental Science, 15*(5), 674–687.

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

Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language, 57*(4), 460–482.

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

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

Kieslich, P. J., & Henninger, F. (2017). mousetrap: An integrated, open-source mouse-tracking package. *Behavior Research Methods, 49*(5), 1652–1667.

Koop, G. J., & Johnson, J. G. (2013). The response dynamics of preferential choice. *Cognitive Psychology, 67*(4), 151–185.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26.

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

Lenth, R. V. (2022). emmeans: Estimated marginal means, aka least-squares means. R package. https://CRAN.R-project.org/package=emmeans

Lin, Y.-C., Lin, P.-Y., & Yeh, L.-H. (2023). Syllable or phoneme? A mouse-tracking investigation of phonological units in Mandarin Chinese and English spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 49*(1), 130–176.

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

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

Lindsay, S., Lucas, A., O'Connor, R. J., Riggs, K. J., & Gaskell, M. G. (2027). Does fast mapping accelerate lexical integration? Evidence from spoken and written novel words. *Journal of Memory and Language, 152, Article 104815.*

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods, 49*(4), 1494–1502.

Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software, 6*(60), 3139.

Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The time course of spoken word learning and recognition: Studies with artificial lexicons. *Journal of Experimental Psychology: General, 132*(2), 202–227.

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

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

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.

McMurray, B., Kapnoula, E. C., & Gaskell, M. G. (2016). Learning and integration of new word-forms: Consolidation, pruning, and the emergence of automaticity. In M. G. Gaskell & J. Mirković (Eds.), *Speech perception and spoken word recognition* (pp. 116–142). Psychology Press/Routledge.

Mirman, D. (2016). Growth curve analysis and visualization using R. Chapman and Hall/CRC.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining $R^2$ from generalized linear mixed-effects models. *Methods in Ecology and Evolution, 4*(2), 133–142.

R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime: User's guide*. Psychology Software Tools, Inc.

Schreiner, T., & Rasch, B. (2017). The beneficial role of memory reactivation for language learning during sleep: A review. *Brain and Language, 167*, 94–105.

Shtyrov, Y. (2012). Neural bases of rapid word learning. *The Neuroscientist, 18*(4), 312–319.

Smith, A. C., Monaghan, P., & Huettig, F. (2021). The effect of orthographic systems on the developing reading system: Typological and computational analyses. *Psychological Review, 128*(1), 125–159.

Snowling, M. J., Stothard, J., & McLean, J. F. (1996). *Graded Nonword Reading Test*. Thames Valley Test.

Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences, 102*(29), 10393–10398.

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.

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

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive Test of Phonological Processing (CTOPP)*. PRO-ED.

Weighall, A. R., Henderson, L. M., Barr, D. J., Cairney, S. A., & Gaskell, M. G. (2017). Eye-tracking the time-course of novel word learning and lexical competition in adults and children. *Brain and Language, 167*, 13–27.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., & Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software, 4*(43), 1686.

Wilson, M. (1988). MRC psycholinguistic database: Machine-usable dictionary, version 2.00. *Behavior Research Methods, Instruments, & Computers, 20*(1), 6–10.

Wulff, D. U., Kieslich, P. J., Henninger, F., Haslbeck, J. M. B., & Schulte-Mecklenbeck, M. (2025). Movement tracking of psychological processes: A tutorial using mousetrap. *Behavior Research Methods, 57*, Article 307. https://doi.org/10.3758/s13428-025-02695-2

Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream. *Behavioral and Brain Sciences, 41*, e120.

**Table 1**

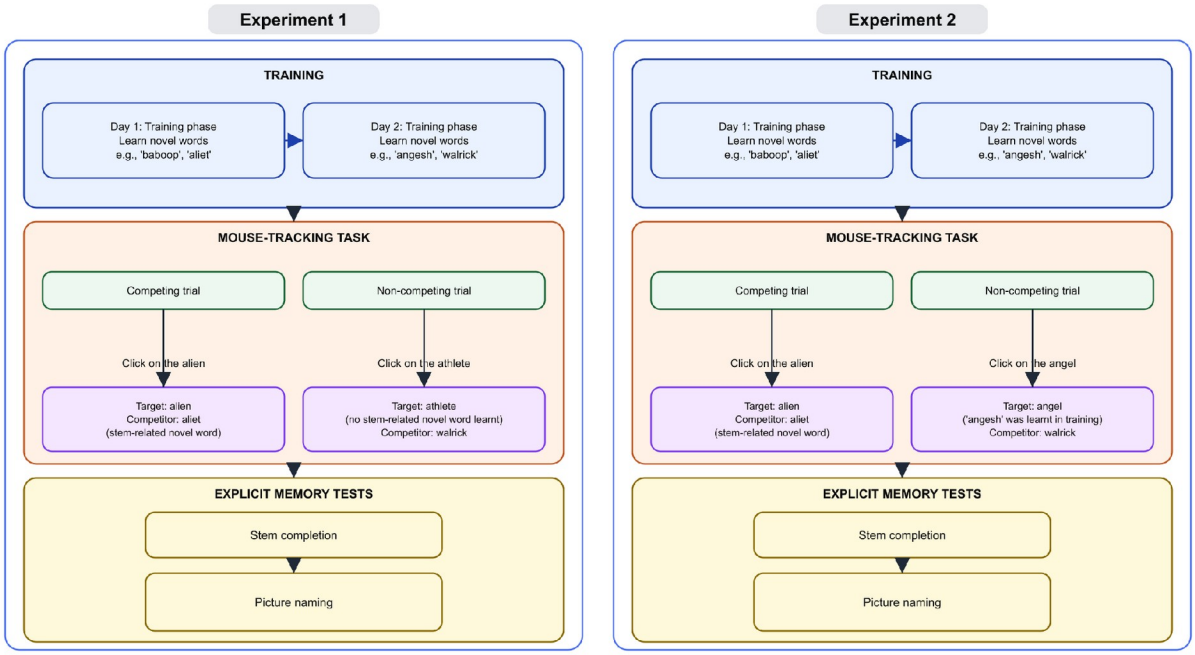
*Signed Maximum Absolute Deviation by Word Type and Competition Condition*

| Word type | Competing | Non-competing | Cohen's dz |
|---|---|---|---|
| **Experiment 1** | | | |
| Familiar | 2.61 (1.31) | 1.94 (1.17) | 0.60 |
| Day 1 | 1.97 (1.07) | 1.64 (0.84) | 0.43 |
| Day 2 | 1.88 (1.02) | 1.64 (0.88) | 0.42 |
| **Experiment 2** | | | |
| Familiar | 2.87 (0.96) | 2.28 (1.08) | 0.64 |
| Day 1 | 2.26 (0.95) | 2.02 (1.03) | 0.27 |
| Day 2 | 2.47 (1.16) | 2.06 (0.89) | 0.36 |

*Note. Means and SDs are computed at the subject level (mean of trials per subject per condition). Cohen's dz is computed from paired subject-level differences (Competing − Non-competing).*

**Figure 1**

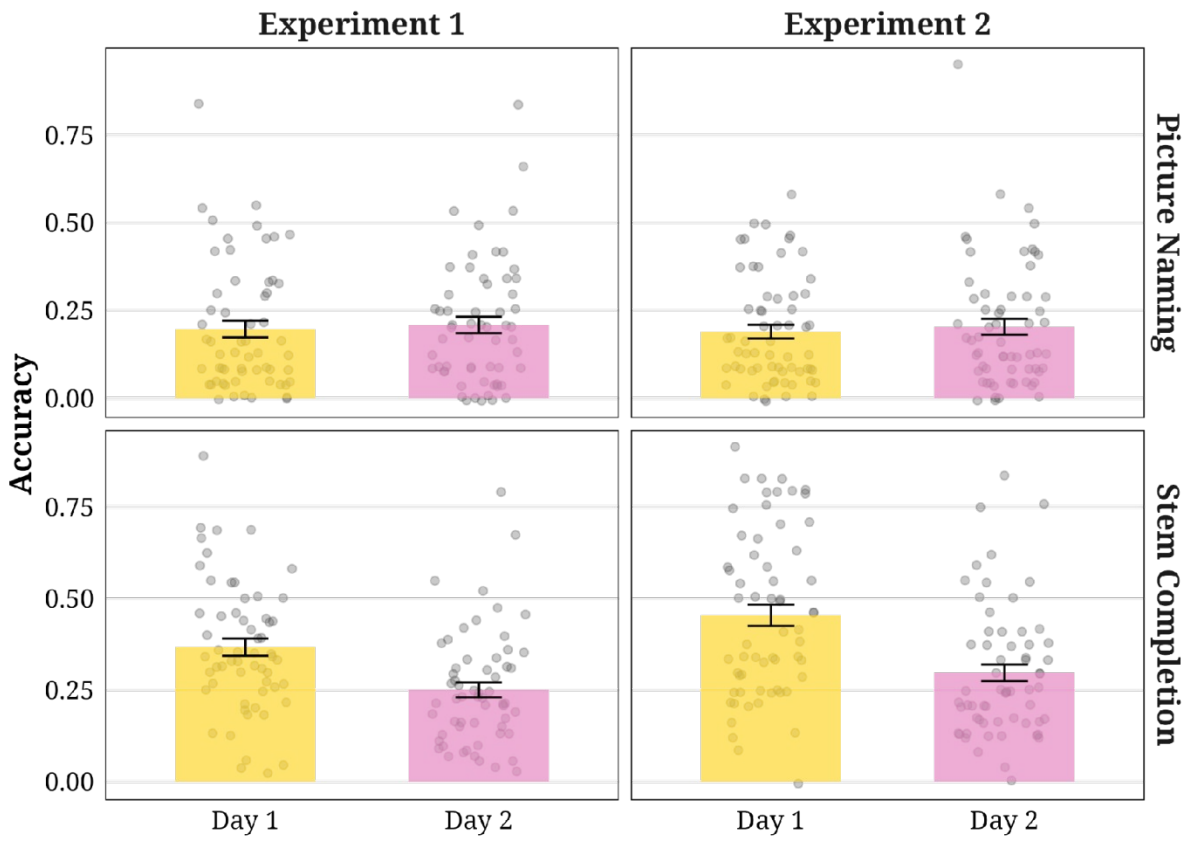
*Schematic of experimental design and trial structure (Experiments 1 and 2).*




*Note*. Training occurred across two days followed by mouse tracking and explicit memory tasks. Experiment 2 differs only in the Non-competing trials of the mouse tracking task as detailed in text.

**Figure 2**

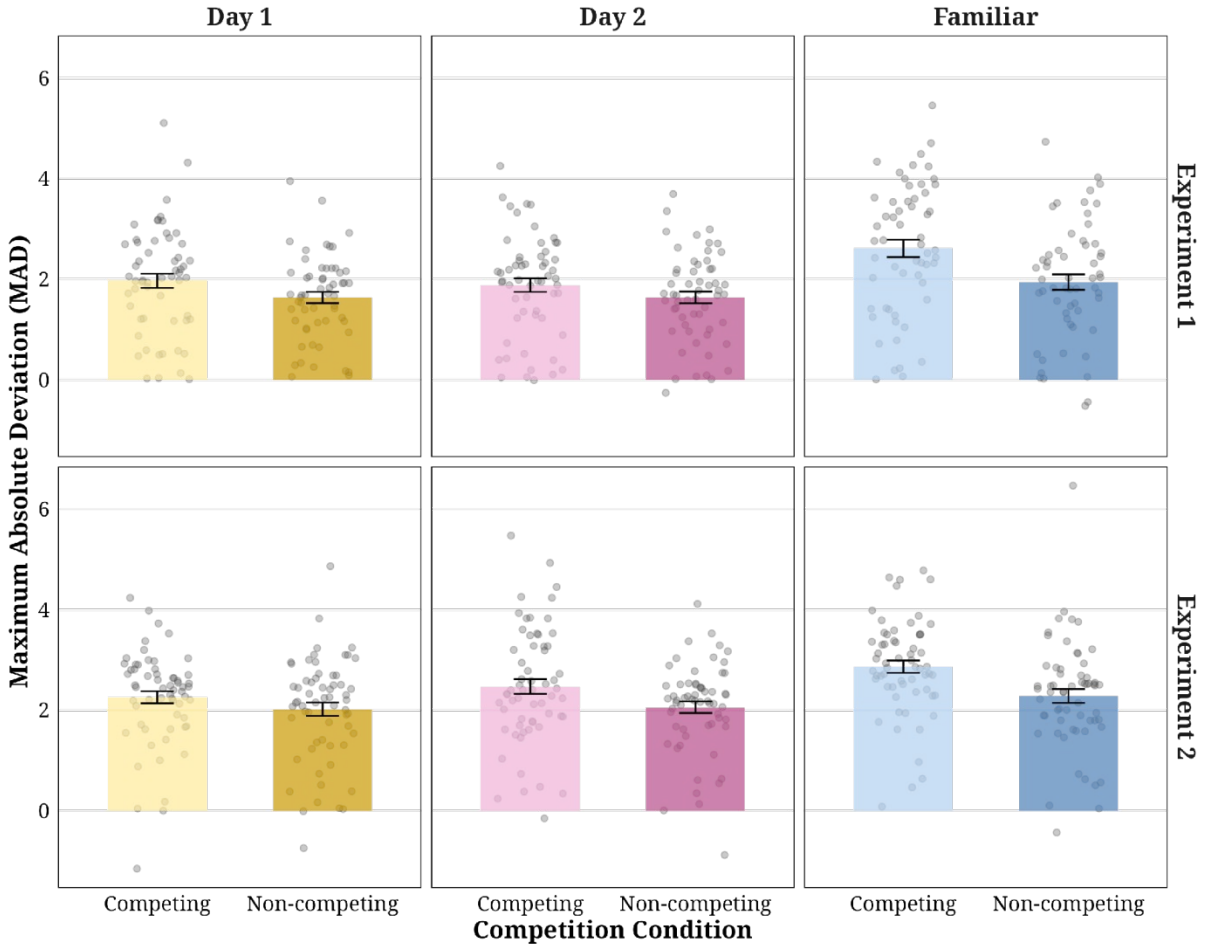
*Mean Accuracy for Picture Naming and Stem Completion Tasks Across Both Experiments*




*Note*. Columns show Experiment 1 (left) and Experiment 2 (right). Rows show picture naming (top) and stem completion (bottom). Bars represent means with 95% confidence intervals, and dots show individual participants.

**Figure 3**

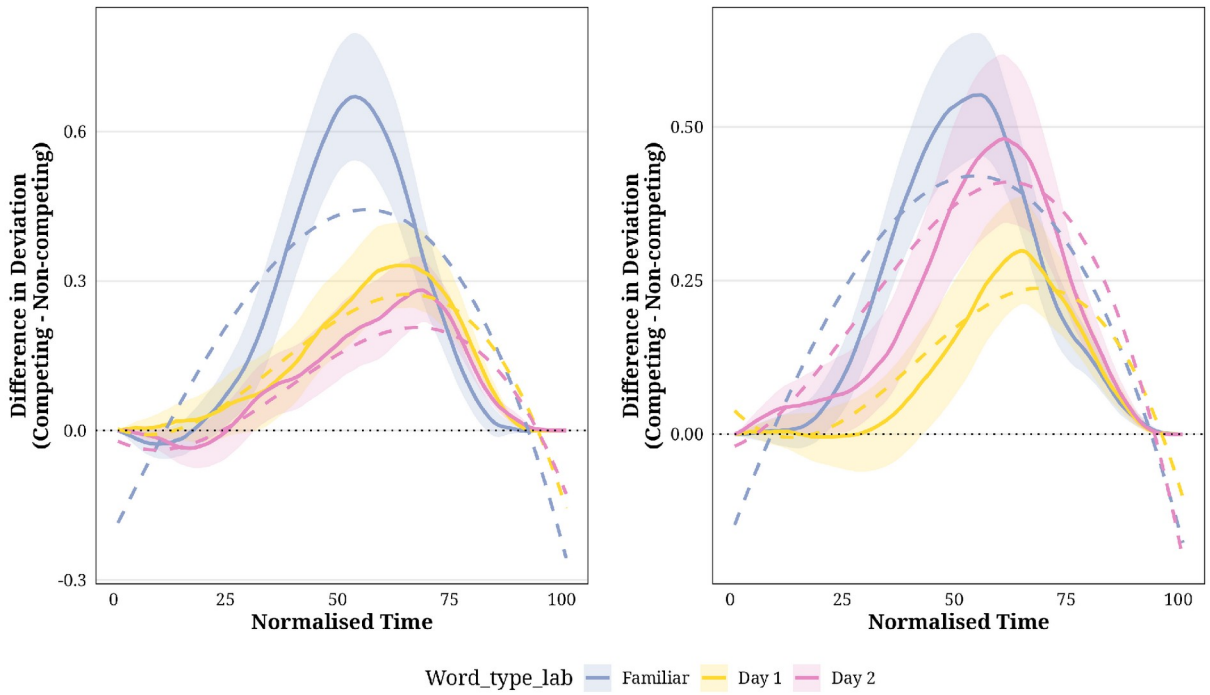
*Signed Maximum Absolute Deviation (MAD) Across Both Experiments*




*Note*. Error bars represent 95% confidence intervals. Individual data points are shown as jittered dots.

**Figure 4**

*Observed and Model-Predicted Competition Time-Courses by Experiment*




*Note*. Solid lines show participant-mean competition effects (Competing − Non-competing) summarised at each time point for Familiar (blue), Day 1 (yellow), and Day 2 (pink); dashed lines show growth-curve model predictions (fixed effects) for the fitted Comp−Noncomp difference by Word type. Shaded ribbons indicate standard errors around the observed summaries. Panels correspond to Experiment 1 (left) and Experiment 2 (right).