Published in final edited form as:

Cognition. 2015 January ; 134: 85–99. doi:10.1016/j.cognition.2014.09.007.

Immediate lexical integration of novel word forms

Efthymia C. Kapnoula and

Stephanie Packard, Prahlad Gupta, Dept. of Psychology, University of Iowa

Bob McMurray

Dept. of Psychology, Dept. of Communication Sciences and Disorders, University of Iowa

Abstract

It is well known that familiar words inhibit each other during spoken word recognition. However, we do not know how and under what circumstances newly learned words become integrated with the lexicon in order to engage in this competition. Previous work on word learning has highlighted the importance of offline consolidation (Gaskell & Dumay, 2003) and meaning (Leach & Samuel, 2007) to establish this integration. In two experiments we test the necessity of these factors by examining the inhibition between newly learned items and familiar words immediately after learning.

Participants learned a set of nonwords without meanings in active (Exp 1) or passive (Exp 2) exposure paradigms. After training, participants performed a visual world paradigm task to assess inhibition from these newly learned items. An analysis of participants' fixations suggested that the newly learned words were able to engage in competition with known words without any consolidation.

Keywords

word learning; lexical integration; lexical engagement; inter-lexical inhibition; spoken word recognition; eye-tracking; visual world paradigm

Introduction

A critical component of learning a new word is binding the elements of its sound pattern (i.e. the phonological representation, or *word form*) together into something more abstract that can ultimately be associated with a meaning. This is often intuitively thought of as acquiring knowledge about the sound pattern of a word. However, decades of work in spoken word recognition (Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Dahan, Magnuson, & Tanenhaus, 2001; Luce & Pisoni, 1998; McClelland & Elman, 1986) suggest

© 2014 Elsevier B.V. All rights reserved.

Corresponding Author: Efthymia C. Kapnoula, E11 Seashore Hall, Iowa City, IA, 52242-1409, 319-383-9022, ekapnoula@iowa.uiowa.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

that this knowledge is embedded in complex ways in the lexical processing system. As a consequence, lexical representations interact with each other and with sublexical phonological representations during spoken word recognition. In this light, learning a word requires not only the acquisition of knowledge about the word's form and meaning, but also embedding this information in multiple components of the system to enable these complex interactions during word recognition. The goal of this study is to investigate the conditions under which such embedding occurs as words are learned, and in particular to determine whether this embedding can occur within the same set of experiences by which subjects learn a word form, or whether it requires additional knowledge (in particular, the meaning of the word) or additional processes (such as consolidation or interleaved exposure).

Lexical properties and their acquisition

Figure 1 offers a loosely connectionist framing of spoken word recognition. The lower layer comprises sublexical phonological representations and the upper layer lexical-phonological representations. Each level of representation contains many elements and, within each level, representations could be virtually anything: completely localist, completely distributed, somewhere in between, or a combination (though we represent them as localist here for ease of exposition). Independently of how the individual elements are represented, learning a novel word requires the adjustment of connections between these sublexical and lexical representations, depicted in terms of bottom-up connections (upward arrows in Figure 1). Thus, what we think of as “knowledge” of a word form consists of the entire system of representational levels, but, especially, the weighted connections between them that allow a listener to access the word form when the sound pattern is heard. This system of knowledge is clearly a crucial property of a word and must be acquired during the learning process.

However, these are not the only connections involved in recognizing a word. There is evidence for competition or inhibition *between* word forms (instantiated by the connections within the word form layer in Figure 1), such that active words suppress activation of less-active competitors (Dahan, Magnuson, & Tanenhaus, 2001; Luce & Pisoni, 1998). There is also evidence for feedback between word forms and sublexical representations (via the top-down connections in Figure 1), by which information can travel from higher to lower levels of processing. This top-down flow of information can influence perceptual processing over the long term as a form of learning (Norris, McQueen, & Cutler, 2003), and may also influence word recognition in real time (Magnuson, McMurray, Tanenhaus, & Aslin, 2003; McClelland, Mirman, & Holt, 2006; but see Norris, McQueen, & Cutler, 2000). In spoken word recognition frameworks, these inhibitory and feedback interactions are usually conceptualized within a localist scheme (McClelland & Elman, 1986; Norris, 1994). For convenience, we adopt that localist terminology here, although we acknowledge that competition and feedback effects can arise within a variety of representational systems. Thus, here the terms *feedback* and *competition* are only meant to represent the general lexical properties of feedback and competition, and not specific mechanisms for implementing them. However they are implemented, the abilities to engage in such interactions are additional *properties* of a word, over and above the property of the knowledge of its phonological word form.

Given this framing, *learning* a word consists of not only acquiring information about its phonological word form (and in particular, encoding it in the bottom-up connections), but also the development of interactive properties, such as the capacity for feedback to sublexical representations, and inhibition among fellow word forms. This raises a fundamental question: What must happen for a word to *acquire* these interactive properties?

In addressing this question, some terminology is in order. Leach and Samuel (2007) proposed a dichotomy between lexical *configuration* and *engagement*. They used the term *lexical configuration* to refer to knowledge about the word itself. In the present context, this could be viewed as the bare minimum informational content required to “know” a word form, which specifies the sound pattern of the word and allows listeners to recognize it (by the bottom up connections in Figure 1). In contrast, the term *lexical engagement* refers to the manner in which a word affects the processing of other representations (e.g. other known words or phonemes), instantiated by the lateral and feedback connections in Figure 1. What is not clear in Leach and Samuel’s formulation is whether the dichotomy between configuration and engagement applies only to the properties themselves, or also to the mechanisms by which these properties are formed.

This is a crucial distinction, because the existence of distinct properties does not necessarily imply the existence of different *learning mechanisms*. The *property* of lexical configuration for word forms is based on the feedforward connections between a representation of the sound input and some abstract representation of the word, whereas the *acquisition* of this property requires the formation of these connections. Similarly, the *property* of lexical engagement is based on inhibitory connections among words, and feedback connections to lower levels of processing, but the *acquisition* of this property again requires the formation of these connections. This formulation helps clarify the constructs of configuration and engagement by operationalizing the distinction between these properties: they depend on different sets of connections. It also speaks to the question of whether this distinction necessitates distinct mechanisms of acquisition. Specifically, it suggests that the dichotomy between properties of word forms may not necessitate distinct mechanisms for their acquisition; the different connections subserving the different properties might nevertheless all develop via similar experience-driven learning mechanisms, as commonly occurs in neural networks.

This frames our central question more precisely. When a new word is learned, what conditions are required for lexical engagement (the capacity for feedback and/or inhibition) to be acquired? Does this require learning experiences or circumstances over and above those needed for the acquisition of a word form’s configuration? Addressing these questions has important implications for the issue of whether there are separate mechanisms by which a word acquires its configuration and its engagement with other words or the phonology; if engagement can emerge concurrently with configuration, then a parsimonious account would be that the same learning mechanism may be operating across both types of connections. In the present study we address these questions focusing on the acquisition of the *engagement via inhibition* property (where the bulk of the prior research has focused), and will have relatively little to say about the *engagement via feedback* property (which has only been examined in one study, Leach & Samuel, 2007).

Engagement of novel word forms

Gaskell and Dumay (2003) investigated the conditions under which *engagement via inhibition* (i.e., between words) arises. Across trials of a phoneme monitoring task, participants were exposed to nonwords, such as *cathedruke*, which closely overlap with real words, like *cathedral*. After 12 exposures (within the same day) knowledge of the novel word forms was sufficient to enable accurate forced-choice recognition, suggesting rapid acquisition of lexical configuration. However, the results for inhibition (engagement) were quite different. Gaskell and Dumay measured inhibition with both a lexical decision task and a pause detection task. In the lexical decision task, the logic is that the presence of a newly learned word, such as *cathedruke*, should slow listeners' recognition of a known word, such as *cathedral*. Similarly, in the pause detection task prior work suggests that listeners are slower to detect a pause after a real word than after a nonword due to the overall larger amount of lexical activity (Mattys & Clark, 2002). Gaskell and Dumay predicted that participants' pause detection latencies for pauses embedded into *cathedral* would be longer after the participants had learned the novel word *cathedruke*. Results suggested that lexical decision for the base words was slowed by the new words, but only after the fourth day of training; similarly, pause detection was affected only in the retest, seven days after the first exposure to the novel words. In a later study, such results were obtained whether or not meaning was provided for the word forms (Dumay, Gaskell, & Feng, 2004). These findings suggest that the establishment of inhibitory connections (i.e. engagement via inhibition) may require substantially more experience than just establishing a word's configuration, and may also require some form of consolidation.

More recent studies have targeted the role of sleep in these hypothesized consolidation processes. Dumay and Gaskell (2007; 2012) used the pause detection paradigm as a measure of inter-lexical inhibition and found an effect of training (i.e. longer pause detection latencies after training on a set of novel words), but only after a period of sleep had elapsed. Dumay and Gaskell (2007) also showed that the emergence of this effect was due to the participants having slept rather than the elapse of time alone. In addition, Tamminen, Payne, Stickgold, Wamsley and Gaskell (2010) used a design based on the Gaskell and Dumay (2003) study (i.e. *cathedral*-like stimuli and a lexical decision task to evaluate lexical engagement) and found a significant correlation between spindle activity (measured polysomnographically) and overnight lexical integration.

Finally, Dumay and Gaskell (2012) used a word-spotting task to disentangle whether novel words are consolidated in the lexicon or are merely stored in some episodic trace that then competes for resources during testing tasks. Subjects had to detect a target word (e.g. *muck*) that was embedded within a nonword (e.g. *lirmuckt*), which was either previously trained or completely novel. Immediately after nonword training, target words were detected faster when embedded within a trained item, suggesting that *lirmuckt* was not yet inhibiting *muck*, and in fact, that episodic familiarity was helping participants notice the embedded word. However, 24 hours later this pattern *reversed*, with now slower detection of *muck* in a trained item. This seems to offer strong evidence that something special happens during sleep that allows newly learned words to engage in inhibition of other existing words (but see Lindsay & Gaskell, 2009). Taken together, the several studies by Gaskell and Dumay

(2003), Dumay and Gaskell (2007, 2012) suggest that simply learning the phonological form of a novel word (configuration) is not sufficient for lexical engagement via inhibition; something additional is needed (see also Leach & Samuel, 2007 for a similar finding for engagement via feedback). In all of these studies, sleep-based consolidation was either directly manipulated (Dumay, Gaskell, and colleagues), or available (Leach and Samuel).

However, sleep may not be uniquely necessary for engagement. Lindsay and Gaskell (2013) used a set of training and testing tasks administered in different sessions 2.5 hours apart within the same day. Phoneme monitoring and stem completion were used to train participants on the novel word forms, while lexical decision was used to evaluate the lexical engagement of the novel words (by looking for interference from similar known words). The arrangement of the different tasks was manipulated across three experiments. While there was no evidence for immediate engagement, there was evidence for its eventual (but not immediate) emergence, without sleep, in two of the three experiments. Crucially, in those two experiments, participants were exposed to both the novel words and their real-word phonological neighbors within the same session. This finding suggests that, even though sleep-based consolidation might not be as crucial as previously thought, interleaved, parallel exposure to the competing word forms is a requirement for lexical integration, along with some form of non-sleep-based consolidation.

The broad pattern that emerges from this body of work is that some set of fairly specialized processes (but not exclusively sleep-based consolidation) must occur in order to establish lexical engagement for newly learned words. This would seem to require distinct learning mechanisms for configuration and engagement. However, before reaching this conclusion it is important to take a closer look at the measures used to assess lexical engagement. The studies cited above have typically used lexical decision (Dumay Gaskell, & Feng, 2004; Gaskell & Dumay, 2003; Lindsay & Gaskell, 2013), word-spotting (Dumay & Gaskell, 2012), and pause detection (Dumay & Gaskell, 2007, 2012; Gaskell & Dumay, 2003) as measures of lexical inhibition. However, these tasks may not uniquely assess lexical inhibition and may be sensitive to other factors. The most important point is that to accurately assess inhibitory engagement, we must look for the ability of a newly learned word to inhibit a specific known word. However, as we discuss below, all three of these tasks measure activation of the target word indirectly, showing slower responses to make various judgments (word/nonword, presence of any word, presence of a pause). Consequently, the *response* to the task may be slowed (inhibited) or not, and this may be somewhat independently of whether the specific activation of the target word is slowed.

First, *lexical decision* RTs may not necessarily depend only on the activation of the target word, but also on the summed lexical activation across multiple words (Grainger & Jacobs, 1996). One possible consequence of this is that extra activity from a novel word speeds lexical decision. That is, listeners may respond “word” faster on the basis of a partially activated competitor (*cathedruke*) and/or the target (*cathedral*). Another possible consequence is that if participants make lexical decision responses only after having *uniquely* recognized the target word, then whenever multiple words are co-active it may be *slower* to make a response, *even if these words are not inhibiting each other*, because unique identification is delayed (e.g., a race model). In this case, RT to *cathedral* may be slowed

simply because *cathedruke* and *cathedral* are co-active, leading to overall greater lexical activation that must be allowed to subside before a lexical decision can be made (as Mattys & Clark, 2002, have suggested for pause detection), even if *cathedruke* and *cathedral* are not actually inhibiting each other. That is, acquiring *cathedruke* may push back the uniqueness point on *cathedral*. Thus according to Grainger and Jacobs' (1996) model of lexical decision, the net observed effect on RT in lexical decision is some uncertain combination of 1) speeding up because of increased overall parallel lexical activation, 2) slowing down because of a race induced by this increased parallel activation, and 3) slowing down because of any true inhibition. Therefore, it is very difficult to interpret the results of a lexical decision task as indicating either the presence or absence of inhibition. Further adding to this difficulty is the fact that when newly learned words are involved, lexical decision RTs may reflect the listener's uncertainty about the lexical status of the newly learned word.

Similarly, *pause detection* is thought to reflect the overall level of lexical activity at the time when the pause is presented; when lexical activation levels are high, listeners have to wait for further evidence to make sure that this is the end of a short word and not the middle of a longer word (Mattys & Clark, 2002). By inserting new words in the set of candidates (i.e. by learning new words) the overall lexical activity is raised, thus delaying pause detection RTs, but again, this decision could be slowed even if the overall set of co-active words is not inhibiting each other.

Finally, *word-spotting* has the same limitations as lexical decision in its susceptibility to the overall level of lexical activation; the task is to spot the presence of *any* word not one specific word. In addition, the results that have been obtained using the word-spotting paradigm are somewhat contradictory. For instance, although as described earlier, Gaskell and Dumay (2003), Dumay and Gaskell (2007; 2012) reported an inhibitory effect of a trained embedded word in word-spotting, Vroomen van Zon, and de Gelder (1996) found a *facilitatory* effect of neighborhood size in word-spotting, which seems contrary to what one would expect by a lexical inhibition account of the word-spotting effect. Similarly, Norris, McQueen and Cutler (1995) found different patterns of effects for different stimuli (CVCCs versus CVCs).

Thus, while it is clear that each of these three paradigms is quite sensitive to the overall degree of lexical activation in the system, it is not clear that word-to-word inhibition is uniquely necessary for any of them to show inhibitory effects. Consequently, what is needed is a way to 1) more directly measure activation of a single and specific target word, and 2) look for the influence of a specific newly learned competitor on the activation of the specific target word.

In studies of inhibition between *well-known* words, this latter goal has been accomplished using splicing techniques. Marslen-Wilson and Warren (1994) attempted one of the earliest tests of between-word inhibition among familiar words capturing this logic. They developed stimuli which had been acoustically manipulated to elicit lexical competition between two specific words. Their stimuli were constructed by combining the final part of one word (e.g. *-b* from *job*) with either 1) the initial portion of another token of the same word (*jo-* from

job; henceforth the *jobb* condition), 2) the initial portion of a different word (*jo*-from *jog*; henceforth the *jogb* condition), or 3) the initial portion of a nonword (*jo*from *jod*; henceforth the *jodb* condition). As a result, the vowels in the constructed stimuli provided either accurate or misleading co-articulatory information about the upcoming consonant. If real words inhibit each other, then when the co-articulatory mismatch misleadingly predicted a real word (*jog*) that was not the target (*job*), this active competitor should inhibit the target word, slowing listeners' ability to recognize it. In contrast, when misleading information came from a nonword (*jodb*) there should not be any interference with the activation of the target word. In this paradigm the competitor word, *jog*, is temporarily over-activated, which should enable it to inhibit *job* (if there is indeed lexical inhibition). Thus, the splicing paradigm allows the experimenter to look at inhibition *from* a specific word. The results showed that lexical decision RTs in the *jogb* and *jodb* conditions did not differ from each other, seeming to suggest the absence of inhibition between words. However, crucially, this was measured using lexical decision which, as we argued, may not measure the specific activation of *job*, but may rather reflect activation across any number of words.

To alleviate this problem, Dahan, Magnuson, Tanenhaus, et al. (2001) adopted Marslen-Wilson's and Warren's splicing paradigm, but instead of using lexical decision they used the visual world paradigm (VWP; Allopenna, Magnuson, & Tanenhaus, 1998; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) to assess activation of the specific target word. In contrast to lexical decision this paradigm offers a measure that is much more directly tied to the specific activation of a single target word (since it measures looks to its specific referent)¹. Auditory stimuli from the three splicing conditions were played while the participants viewed the picture for the target word (e.g., *job*) accompanied by three other pictures, none of which corresponded to the competing word (e.g., *jog*). Participants' fixations to each picture were monitored as a measure of ongoing lexical activation. Participants made significantly less fixations to the target picture when the target word was spliced with a different word (*jo_gb*) than when it was spliced with itself (*jo_bb*), or a nonword (*jo_db*). It is crucial to point out that if no inhibition existed between the familiar words, then boosting the activation of the competitor word via spliced mismatch should not have any greater effect on the activation of the target word, since the competitor was not present on the screen. Thus, when we manipulate activation of the inhibiting word (using splicing), and measure only activation for the inhibited word, there is clear and specific evidence for lexical inhibition, among known words. By using the VWP instead of lexical decision Dahan, Magnuson, Tanenhaus, et al., (2001) found a lexical inhibition effect using stimuli that were identical to those used by Marslen-Wilson and Warren (1994).

The Dahan, Magnuson, Tanenhaus, et al. (2001) study coupled a stimulus set specifically designed to boost naturally occurring lexical inhibition, with a testing paradigm that is sensitive to lexical activation of a specific word (but less susceptible to effects of overall levels of lexical activity). The present study thus uses this method to assess inhibition

¹While looks in the VWP are sensitive to non-displayed competitors (e.g., Dahan, Magnuson, Tanenhaus, et al., (2001); Magnuson, Dixon, Tanenhaus, & Aslin, 2007), current models linking activation to fixations (Allopenna et al., 1998; Salverda, Brown, & Tanenhaus, 2011) suggest that fixations are largely a product of the visible objects and effects of non-displayed items therefore represent inhibition of the target word. This is borne out by the fact that Dahan, Magnuson, Tanenhaus, et al., (2001) were able to detect inhibitory effects using the VWP, but Marslen-Wilson and Warren (1994) were not.

between a recently learned nonword and a specific target word, even when the learning situation does not provide sleep, an explicit meaning, or spaced interleaved exposure to the real word neighbors. Further, the use of CVC words in the subphonetic mismatch paradigm effectively holds the uniqueness point constant. Moreover, while we do not claim that this particular formulation of the VWP is not susceptible to processes other than inhibition (which we discuss in the general discussion), its apparent ability to measure inhibition more directly offers an important complement to prior methods like lexical decision.

Twenty word-sets, similar to Dahan, Magnuson, Tanenhaus, et al. (2001), were constructed. Each consisted of two words and a nonword (*job, jog, jod*). Participants were trained on half (10) of the nonwords (e.g. *jod*) for a brief period of time, and tested immediately afterward in a VWP experiment similar to Dahan, Magnuson, Tanenhaus, et al. (2001). If evidence for inhibition from the newly learned word forms on the real word is found in this testing paradigm following the brief training, it would show that the kinds of connections necessary for inhibitory lexical engagement can be built from the earliest stages of word learning, and that neither sleep nor other consolidation nor an explicit referent nor contrastive exposure to competitors are necessary for novel word forms to start competing with other words. In other words, the property of inhibitory lexical engagement can be acquired immediately after brief exposure to a meaningless word form, concurrently with lexical configuration. This in turn would suggest that it may not be necessary to assume distinct learning mechanisms for the emergence of configuration and (inhibitory) engagement. Such results would leave open the question of what is changing with sleep, and we return to that in the general discussion.

We conducted two experiments. In Experiment 1 training had the form of a Listen-and-Repeat task paired with a Stem-Completion task (Karpicke & Roediger, 2008; Packard & Gupta, 2009; Zhao, Gupta & Packard, 2010; Packard, 2010), akin to the word-fragment completion task used by Schwartz and Hashtroudi (1991). Training using the word stem as a cue was employed in Experiment 1 because previous work in the Gupta lab has shown higher levels of word learning from stem completion compared to simple repetition training (Packard, 2010). In Experiment 2 we examined whether production and effortful learning of the novel word form are necessary for this effect to arise, using a different training procedure (phoneme monitoring) that did not entail production and did not require participants to memorize the word forms.

Experiment 1

Methods

Participants—Thirty-eight native speakers of English participated in this experiment. All were students at the University of Iowa and received course credit as compensation. Two of them were excluded from the analyses due to problems with the eye-tracking.

Design and materials—All tasks were based on a set of 20 nonwords. Participants were first trained on 10 of these nonwords using a combination of two tasks: a Listen-and-Repeat task and a Stem-Completion task. The other 10 nonwords were not presented at this time and served as controls during the testing period. The selection of nonwords as trained or

untrained was counterbalanced among participants (so each nonword appeared in both conditions across participants), and three different random groupings were used across participants for a total of six different lists. After training, participants performed a Visual World Paradigm (VWP) task designed to assess lexical inhibition from the newly learned (trained) nonwords to existing words in the lexicon.

For each of the 20 nonwords, we identified a picturable, real-word counterpart (e.g., the word *job* for the nonword *jod*), which would be the *target word* in the later VWP task. We also identified a word competitor for this target word (e.g., *jog* as competitor for *job*). The nonword together with the two words formed a *triplet*, and there were 20 such triplets (see Appendix A for a full list of the triplets).

An auditory recording of each of these 60 stimuli was created. For each triplet, we then created three spliced auditory recordings of the triplet's target word: a *matching splice*, in which an auditory token of the target word was spliced with a different auditory token of that word, e.g. *jo_bb*; an *other-word splice*, in which an auditory token of the target word was spliced with an auditory token of the competitor word in that triplet, e.g., *jo_gb*; and a *nonword splice*, in which an auditory token of the target word was spliced with an auditory token of the nonword in that triplet, e.g., *jo_db*. Each of the three splices created from a triplet was used as the auditory stimulus in a different trial in the VWP task.

In the VWP task participants were presented with pictures of four objects in a display, and concurrently heard a real word auditory stimulus. Their task was to identify which of the four pictures was the referent of the auditory stimulus. The auditory word was one of the three types of spliced stimuli just described. However, because a nonword splice could come from either a trained nonword or an untrained nonword, there were *four* types of auditory stimuli and hence four splice conditions in the VWP task: the *matching-splice* condition, the *other-word-splice* condition, the *trained-nonword-splice* condition, and the *untrained-nonword-splice* condition.

From the VWP task, we used fixations to the picture of the target word (e.g. *job*) as an estimate of its lexical activation. Crucially, by examining this as a function of the splice condition, we can assess any inhibition from newly learned words (engagement): if a new word is integrated then it can compete with real words during online processing, inhibiting the target word's activation. Therefore, similarly to Dahan, Magnuson, Tanenhaus, et al (2001), we assessed the decrement in looking for the other-word, trained-nonword, and untrainednonword conditions relative to the matching-splice condition.

We expected that any of the three mismatching splices would interfere with activation of the target simply due to subphonemic mismatch alone (e.g., *jo_dg* and *jo_gb* are not good exemplars of *job*). This effect of purely bottom-up mismatch, independent of any lexical competition, can be assessed by comparing of looks to the target in the untrained-nonword-splice and the matching-splice condition provided us with an estimate of this. However, if we find that the difference between the other-wordssplice and matching-splice conditions was greater than the difference between the untrained-nonwordsplice and the matching-splice, this would indicate lexical competition, replicating Dahan, Magnuson, Tanenhaus, et al.,

(2001). Most crucially, this can be extended to the newly learned words by comparing the interference effect of trained and untrained nonwords on the target word. Since the VWP portion of the experiment used all 20 nonwords (assigned to the trained/untrained condition) in a Latin-square design, the effect of training was assessed based on the difference between the trained-nonword-splice and untrained-nonword-splice conditions.

The 20 triplets (word, nonword/newly-learned word, and other-word) were constructed with the above overall design in mind. All nonwords were monosyllables of a CVC, CCVC, or CVCC² structure and ended in a stop consonant (/b/, /d/, /g/, /p/, /t/, /k/). Pilot work (and phonetic sensibility) indicated that subphonemic mismatches in voicing and/or manner cause a much greater reduction in target fixations than those involving place of articulation. As a result all triplets were constrained to have the same voicing and manner and only deviate on place of articulation. However, we were still concerned that even within a voicing class different places of articulation may exert different degrees of coarticulatory strength (e.g., labials could exert a stronger effect than coronals). To verify that this was not the case, we conducted a chi-square analysis to examine whether the type of coarticulatory mismatch differences between the matching-splice and the other-word-splice conditions (the number of labial→velar, coronal→labial splices, etc.) were different from those between matching-splice and the nonword-splice conditions. Differences were not statistically significant ($\chi^2(2) = 2.03$, $p = 0.36$).

From these 20 triplets, three lists of 10 nonwords were randomly selected for training, with complementary lists in which the same nonwords were untrained, creating three pairs of lists. Six participants were run in each of the three list pairs.

Training: Procedure and Stimuli—Participants were seated in a sound-attenuated room and received instructions on the training portion of the experiment. Training consisted of 11 epochs of exposure to the 10 trained nonwords. Within each epoch there was one block of 10 Listen-and-Repeat trials followed by one block of 10 Stem-Completion trials (i.e., one trial of each type for each nonword). Therefore, there were a total of 220 trials (10 words $\times$ 11 epochs $\times$ 2 blocks). During the Listen-and-Repeat task the participants first listened to the nonword and then repeated the nonword into a microphone, cued by a cross in the center of the screen. During the Stem-Completion task, only the first part of the nonword (e.g., *jo-* for the nonword, *jod*) was played and the participant repeated the whole word. Participants had the opportunity to perform five practice trials at the beginning of the training session in order to make sure they had understood the task. The presentation order was randomized within each block and for each participant separately. Training lasted approximately 10-20 minutes.

Stimuli: The Listen-and-Repeat stimuli consisted of the entire nonword and were recorded by a male native speaker of American English in a sound-attenuated room, sampling at 44100Hz. Stemcompletion stimuli came from separate recordings in which the speaker only spoke the first part of the nonword (up through the vowel or the approximant preceding the last consonant). Auditory stimuli were delivered over high quality headphones.

²When coda clusters were used (CVCC words) the first consonant in the coda was always an approximant.

Testing: Stimuli and Procedure

Overview and design: After the training phase participants immediately began the VWP Task. A VWP set was constructed from each of 20 the triplets previously described. Each VWP set consisted of the target word from each triplet together with three other real words that were not in any triplet (fillers). On each trial, the participant saw a visual display of the four pictures from one VWP set, then heard the name of one of the pictures and clicked on the named picture. Each auditory stimulus was presented in three different spliced forms. Thus, a complete design entailed 20 VWP sets $\times$ 4 pictures $\times$ 3 splices = 240 trials. Sixty of these trials involved a target word picture, 20 each, with an auditory stimulus that was a matching splice, an other-word splice, or a nonword (either trained or untrained) splice. For any particular participant, half of these 20 nonword splice trials were with a trained nonword splice and half with an untrained nonword splice.

Stimuli: In each VWP set of four word-picture pairs, one of the words was the target word from the triplet (e.g., *job* from the *jod-job-jog* triplet). The other real word in the triplet (*jog*) was never played and no image of it was ever shown. The three additional filler words in the VWP set were semantically unrelated to the target word. One of the fillers had an initial-phoneme overlap with the target word (e.g., *jet* for *jog*), which was done following Dahan, Magnuson, Tanenhaus, et al., (2001) to slow fixations to the target; the other two were phonologically unrelated (e.g. *duck* and *book*). Each of the four pictures in the set (*job*, *jet*, *duck* and *book*) had an equal probability of being the referent of the played word. For each VWP set, both visual and auditory stimuli were constructed. A complete list of the stimuli used in the VWP is shown in Appendix B.

As described, auditory stimuli for the target words were constructed by cross-splicing the final stop consonant of the target word onto the initial portion of each of the three word forms in the corresponding splicing triplet. For each stimulus, the same talker recorded the three complete words in the triplet in a sound attenuated room at 44,100 Hz. For each word we recorded multiple tokens, and the stems were listened to by a small team of 3-4 phonetically trained listeners (including the first and last authors) to identify the single token with the strongest co-articulation. Splicing occurred at the zero-crossing closest to the onset of the release. The average duration was 377 ms for the pre-splice sequences and 88ms for the post-splice sequences. There were no significant differences in duration between splice conditions ($F < 1$). The auditory stimuli for the three filler items in each VWP set were constructed in the same way as the experimental stimuli and the only difference was that each word was spliced either with another token of itself or one of two nonwords (*je_t*, *je_kt*, *je_pt*). None of the nonwords spliced with the filler words occurred in any to-be-trained word sets; the purpose of these splices was simply to control generally for the splicing in the auditory stimuli presented for the VWP targets.

Visual stimuli consisted of pictures of the referent of each word in the VWP set. For each of these 80 words (20 sets of four words: a target and three fillers), a picture was developed using a standard lab procedure (Apfelbaum, Blumstein, & McMurray, 2011; McMurray, Samelson, Lee, & Tomblin, 2010). For each word, a set of 5-10 candidate images were downloaded from a commercial clipart database and viewed by a small focus group of 3-5

undergraduate and graduate students. From this, one image was selected as the most prototypical exemplar of that word. These were subsequently edited to remove extraneous elements, adjust colors and ensure an even more clear depiction of the intended word. The final images were approved by a lab member with extensive experience using the VWP.

Procedure: Testing in the VWP began immediately after the training procedure. First, participants were familiarized with the 80 pictures used in the VWP task, seeing each of the pictures accompanied by its orthographic label³. Next, they were fitted with an SR Research Eyelink II headmounted eye-tracker. After calibration, participants were given instructions for the testing phase, and testing began.

At the beginning of each trial, a set of four pictures was presented in four corners of a 19" monitor operating at 1280 × 1204 resolution. Simultaneously, a small red circle appeared at the center of the screen. After 500 ms, the circle turned blue, cueing the participant to click on it to start the trial. This allowed the participants to briefly look at the pictures before hearing anything, thus minimizing eye-movements due to visual search (rather than lexical processing). As soon as participants clicked on it, the blue circle disappeared and an auditory stimulus corresponding to one of the four words was played (either the target word, or one of the three fillers). Participants then clicked on the picture corresponding to the word, and the trial ended as soon as a click was made. There was no time limit on the trials, and participants were not encouraged to respond quickly. Participants typically responded in less than 2 sec ($M = 1079.76$, $SD = 150.03$ ms), which is typical in these experiments.

Eye-tracking Recording and Analysis: Eye-movements were recorded at 250 Hz using an SR Research Eyelink II head-mounted eye-tracker. Whenever possible both corneal reflection and pupil were used. Participants were calibrated using the standard 9-point display. The Eyelink II compensates for head-movements to yield a real-time record of gaze in screen coordinates. As in prior studies (McMurray, et al, 2010; McMurray, Tanenhaus, & Aslin, 2002), this was automatically parsed into saccades and fixations using the default psychophysical parameters, and adjacent saccades and fixations were combined into a single "look" that started at the onset of the saccade and ended at the offset of the fixation. In converting the coordinates of each look to the object being fixated, the boundaries of the ports containing the objects were extended by 100 pixels in order to account for noise and/or head-drift in the eye-track record. This did not result in any overlap between the objects (the dead space between pictures was 124 pixels vertically and 380 pixels horizontally).

Results

During the training phase, participants' average accuracy was 96.89% ($SD = 3.53\%$) in the listenand- repeat trials and 71.39% ($SD = 20.08\%$) in the stem completion trials. In the VWP testing phase, participants were 99.04% accurate ($SD = 1.74\%$) in clicking on the right picture. Thus, participants learned the words to a fairly high level and had no trouble with

³Familiarization to the pictures could not interfere with the experimental design since participants were only visually exposed to the orthographic labels of the pictures - no auditory stimuli were presented at this phase. In addition, participants were equally pre-exposed to all the visual stimuli that were subsequently used in the VWP task (i.e. the target items corresponding to a trained nonword, [job if jog was trained] as well as the target items for the untrained nonword [net if nep was untrained]).

these tasks. Looking only at accuracy on the 60 experimental trials during the VWP phase, there was a significant effect of splice condition ($F(3,140) = 6.24, p < .01$). Bonferroni adjusted post-hoc comparisons revealed that accuracy in the matching-splice condition ($97.5\% \pm .028$) was significantly lower than both the untrained-nonword-splice ($99.44\% \pm .023, p = .013$) and the trained-nonword-splice ($99.72\% \pm .017\%, p = .003$) conditions. Also, the other-word splice condition accuracy ($97.92\% \pm .028\%$) was significantly lower than that in the trained-nonword-splice condition ($p = .026$). No significant differences in accuracy were found between the trained-nonword- and untrained-nonword-splice conditions ($p = 1$). The explanation for these differences was not immediately obvious, but we note that despite the statistical significance of these differences, they are numerically quite small.

Evaluating competition from newly learned words—To test our primary hypotheses, we examined the fixation data from the experimental trials of the testing phase (i.e. trials in which the auditory stimulus was one of the 20 experimental targets). We restricted analysis to only those trials in which participants selected the correct picture, excluding on average 1.0 trial (1.67%) per participant. We started by computing the proportion of trials on which participants were fixating the target picture at each 4 ms time-slice for each of the four splicing conditions (Figure 2). This shows a clear effect of splicing condition, with matching-splice targets showing the fastest responding (the quickest increase in looks to the targets), followed by the untrained-nonword-splice, trained-nonword-splice, and other-word-splice conditions. Given that the untrained-nonword- and trained-nonword-splice conditions were comprised of the same nonwords across subjects, the difference between them appears to support our hypothesis that newly learned word forms can inhibit known word recognition with minimal training.

To evaluate this statistically, we computed the average proportion of fixations to the target picture in a time window between 600 ms and 1600 ms post stimulus onset as our dependent variable⁴. This was analyzed with linear mixed effects models using the lme4 (version 0.999999-2) (Bates, Maechler, & Dai, 2009) and languageR (Baayen, 2009) packages in R (R Development Core Team, 2009). Since our dependent variable was a proportion, we transformed these proportions using the empirical logit transformation. Before examining fixed effects, we constructed several models to determine the appropriate random effects structure. In these models, subjects and items were entered both as random intercepts and with splice condition as a random slope on either or both. From this, the model with random intercepts for both subjects and items was the most complex model supported by data—adding random slopes did not yield a significantly better model fit. Lastly, because there is uncertainty about the number of d.f. in linear mixed effects models, statistical significance was assessed using the Monte-Carlo Markov Chain estimation procedure with 20,000 iterations.

⁴The 600ms were chosen based on the fact that the stem duration (i.e. pre-splice sequence) had an average duration of ~400ms plus the 200ms needed to plan an eye-movement. The average subject RT across experiments was ~1350ms, however among our subjects some had average RT ~1000ms and some as high as ~1900ms. Therefore, since participants' speed varied, we decided to use a broad time window in order make sure we capture the effect in both fast and slow participants.

Two models were run using this structure. First, to determine whether our paradigm was sensitive to competition between the two known words (replicating the key finding of Dahan, Magnuson, Tanenhaus, et al., (2001)), we used a subset of our data, excluding the trained-nonword-splice condition trials. Splice condition was the only fixed effect and was coded as two variables; 1) matching- versus untrained-nonword-splice (+.5 / -.5) and 2) untrained-nonword- versus other-word-splice (+.5 / -.5). The proportions of looks to the target were higher in the matching-splice condition than the untrained-nonword-splice condition ($b = -0.687$, $SE = .228$, $p_{MCMC} < 0.001$), suggesting that participants were sensitive to the subphonemic mismatch even when it did not predict another real word. In addition, the untrained-nonword-splice was significantly different from the other-word-splice condition ($b = -0.788$, $SE = .228$, $p_{MCMC} < 0.001$). These results together indicate a replication of Dahan, Magnuson, Tanenhaus, et al., (2001) critical finding of greater looks to the target in the matching-splice than in the other-word-splice condition, and are thus overall in agreement with Dahan, Magnuson, Tanenhaus, et al., (2001).

The second model focused on our key comparison. Here, we limited our analysis to the trained-nonword- and untrained-nonword-splice condition trials. Splice condition was again the only fixed effect and was coded as one factor: untrained-nonword-splice versus trained-nonword-splice. The difference between the two conditions was significant ($b = -0.542$, $SE = .235$, $p_{MCMC} = 0.021$), indicating a larger interference effect for the trained-nonwords than the untrained.

Time course of competition from newly learned words: One question that arises from these findings is whether the immediate competition from newly learned, unconsolidated words demonstrated in the prior analysis may be due to a different mechanism than previously reported competition effects that emerge only following a consolidation period or interleaved training. Davis and Gaskell (2009) raise the possibility that two learning/memory mechanisms could be at work in novel word learning and engagement. First, episodic hippocampal representations of newly learned words could be available even without a consolidation period and might create competition during known-word recognition with a slightly delayed time course. Second, truly lexical competition from new competitors may also arise but only after a consolidation period (or with interleaved exposure) and this latter form of competition would have a faster time course. According to this view, previously reported competition effects would be of the latter kind, whereas the results we have reported would be of the former kind, and thus based on a different mechanism.

On such a view, we would predict *later* effects of inhibition from newly learned words than from known (by definition, consolidated) words. This provides a way of testing the idea of different underlying mechanisms. To address this possibility we therefore examined the timing of the competition effect for both newly learned and previously known words. We first quantified the competition effect induced by newly learned words by subtracting the proportion of looks to the target (over time) in the trained-nonword- splice condition from that in the untrained-nonword-splice condition. Similarly, to quantify competition from previously known words we subtracted the proportion of looks to the target in the other-word- splice condition from that in the untrained-nonword-splice condition. These are

shown in Figure 3, and indicate how strong the interference effect was at each moment in time.

To assess statistically assess the onsets of these effects, we adopted the approach of McMurray, Clayards, Tanenhaus and Aslin (2008; see also Toscano & McMurray, 2012, 2014), by computing the time at which each competition effect crossed 20% of its maximum. Because individual subject's time course effects were quite variable, we jackknifed the data prior to computing the effect onsets. Here, we averaged the time courses of each participant less one, and computed the onset. This was then repeated excluding each participant in turn, attaining dataset the same size as the original. The results can be compared using a modified T-test. Using the jackknifed data, we computed the point in time at which the competition effect for each subject crossed 20% of that subject's maximum value, with the requirement that it stayed over 20% for at least 40ms. Paired samples t-test revealed that the competition from newly learned words was slightly faster (509ms) than that from previously learned words (640ms), but this difference was not significant, $t_{\text{jackknifed}}(35) = 1.03$, $p = .312$. This is not consistent with a slower episodic representations interpretation.

Discussion

The results of Experiment 1 suggest that previous exposure to a word form enables it to compete with other words in real time, after only minimal training which included no opportunity for consolidation. This in turn suggests that novel words can start inhibiting other known words immediately after they are learned – without sleep, meaning, or spaced or interleaved training. Moreover, this inhibition can be observed quite early in the time course of processing, as early as the competition derived from known words.

In Experiment 1, training on the novel word forms included their overt production as well as effortful recollection. Experiment 2 was designed to evaluate whether these processes are necessary during training for competitive engagement to develop. Specifically, in Experiment 2 the Listen-and-Repeat and Stem-Completion tasks were replaced by a phoneme monitoring task as the training procedure, and instead of single recording tokens we used multiple tokens. The testing was identical to that of Experiment 1.

Experiment 2

Participants

Seventy-two native speakers of English participated in this experiment. All were students at the University of Iowa and received either a gift card or course credit as compensation. Twelve of them were excluded from the analyses due to problems with the eye-tracking.

Training: Procedure and Stimuli

Training consisted of 22 blocks of phoneme monitoring. Within each block participants were asked to monitor for a specific phoneme (/k/, /p/, /t/, /b/, /d/, or /g/) while listening to each of the ten nonwords (i.e. $22 \times 10 = 220$ trials in total). Participants had the chance to perform five practice trials at the beginning of the training session in order to make sure they

had understood the task. The presentation order was randomized within each block and for each participant separately. Training lasted approximately 10-15 minutes.

Stimuli—The same 20 nonwords of Experiment 1 were used. However instead of using the same recordings, we used 5 different tokens for each nonword during training. All recording and stimulus preparation procedures were identical to Experiment 1.

VWP Task: Stimuli and Procedure

The VWP testing immediately followed the phoneme monitoring training, and was identical (in both stimuli and procedure) to Experiment 1.

Results

In the training phase the participants' average phoneme monitoring accuracy was 94.54% ($SD = 11.96$). This excludes training data for three subjects who misplaced their fingers on the buttons during the training (but were kept in the analyses of testing). In the VWP testing phase they selected the correct picture at 99.5% ($SD = 0.88\%$) accuracy. Looking only at the 60 experimental trials, across all conditions accuracy was over 99% and no significant effect of splice condition was found on accuracy ($F(3,284) = 1.02, p = .39$).

Evaluating competition from newly learned words—Figure 4 shows the proportion of fixations to the target as a function of time for each of the four splicing conditions. As in Experiment 1, analysis was restricted to only those trials in which participants selected the correct picture ($M = 0.3$ [.5%] trials / participant). Again we see a clear effect of splicing condition, with matching-splice targets showing the fastest responding, followed by the untrained-nonword- splice condition and then the trained-nonword- and other-word-splice conditions. Given that (across subjects) the trained and untrained splice conditions were the same nonwords, this appears to support our hypothesis that novel words can inhibit known word recognition with minimal training.

To evaluate this statistically we conducted the same linear mixed model as in Experiment 1. The same (600-1600 ms) time window was used and again proportions were transformed with the empirical logit function. Again, we first compared models on their random effects structure and found that the model with random intercepts for subjects and items was the best fitting model. Statistical significance was assessed using the Monte-Carlo Markov Chain estimation procedure with 20,000 iterations.

As in Experiment 1 we started by documenting that known words interfered with each other in our VWP task (as in Dahan, Magnuson, Tanenhaus, et al., (2001)). To do this, we excluded the trained-nonword-splice condition trials and used Splice condition as the only fixed effect, coded as two factors; 1) matching- versus untrained-nonword-splice ($+/- .5$) and 2) untrained-nonword- versus other-word-splice ($+/- .5$). Proportions of looks to the target were higher in the matching-splice condition compared to the untrained-nonword-splice condition ($b = -0.838, SE = .168, p_{MCMC} < 0.001$) and the untrained-nonword-splice was significantly different from the other-word-splice condition ($b = -0.738, SE = .168, p_{MCMC} < 0.001$). In the second analysis, examining the effect of learning, we again included only

the trained-nonword- and untrained-nonword-splice conditions and splice-condition was the only fixed effect (untrained-nonword-splice versus trained-nonword-splice: $\pm .5$). As in Experiment 1, the difference between the two conditions was significant ($b = -0.415$, $SE = .190$, $p_{MCMC} = 0.029$) with more interference for trained nonwords.

Time course of competition from newly learned words: Similarly to Experiment 1, we compared the time course of the two competition effects (competition from newly learned and previously known words) in order to address the question of whether the novel words compete with other words via a slowed mechanism (Figure 5). The procedure was identical to Experiment 1. Paired samples t-test again showed that the competition from newly learned words was slightly faster (500ms) than that from previously learned words (640ms), but not significantly, $t_{\text{jackknifed}}(59) < 1$.

Discussion

The results of Experiment 2 confirm the pattern of results from Experiment 1; novel words engage in competition with other words immediately after initial exposure and this competition is as fast as in the case of previously known words. Moreover, we found this effect even after removing the demands of speech production from the training and substituting the original training with a less effortful and more passive exposure (phoneme monitoring) to multiple recordings of the nonwords. These findings suggest that mere exposure to the phonological word form is sufficient for immediate lexical engagement to occur.

General discussion

This study asked whether the emergence of lexical configuration and engagement call for different learning conditions. Previous research suggests that lexical engagement via inhibition develops more slowly than lexical configuration, and also requires additional criteria to be met, such as sleep-based consolidation (Gaskell & Dumay, 2003; Dumay & Gaskell, 2007; Tamminen et al, 2010), or training that explicitly interleaves exposure to the novel words and their familiar competitors along with non-sleep-based consolidation (Lindsay & Gaskell, 2013). As a result, these prior findings appear to suggest that different learning mechanisms support the acquisition of lexical configuration and engagement. However, in contrast to prior methods of assessing inhibitory engagement, we used a method that 1) experimentally manipulates activation of the inhibiting word and 2) measures specific activation of the inhibited word (rather than general activation across words). Using this more sensitive method, we observed the effects of inter-lexical inhibition after a very brief training that incorporated none of the above factors. This showed evidence for immediate lexical engagement of novel word forms after only a very brief (10-20 minute) training, without the need for special learning procedures other than, or additional to, those required for lexical configuration. This was replicated in two different tasks (and also in a third, unreported experiment based on Experiment 2). In addition, our time course analyses suggest that this competition effect is not slower than the competition from previously known words. This is inconsistent with the idea that unconsolidated lexical representations compete with other words via an entirely different, slower mechanism (Davis & Gaskell,

2009). These results provide a clear answer to the central empirical question that motivated this study.

Moreover, considering just real-time word recognition, at face value inter-lexical inhibition does not appear necessary for word recognition, and indeed several models do not include it (Gaskell & Marslen-Wilson, 1997; Marslen-Wilson, 1987). Perhaps this is the case if we consider only a small set of distinct words. However, in interactive activation models such as TRACE (McClelland & Elman, 1986) and Shortlist (1994), inhibition is essential for dealing with embedded words (e.g., the *ham* in *hamster*) and segmentation ambiguities (*car go* / *cargo*) as in these circumstances. In such cases, since there is no bottom-up phonetic information to disqualify *ham*, only inhibition from the longer word can do it. So inhibition may in fact be necessary for these sorts of problems (see Gow & Gordon, 1995), and the availability of inter-lexical inhibition from the earliest moments of learning may help listeners recognize newly learned words in more complex, realistic situations.

What causes the discrepancy between the findings reported here and previous work (such as Davis, Di Betta, Macdonald, & Gaskell, 2009; Dumay & Gaskell, 2007; Gaskell & Dumay, 2003)? Why did we find such robust evidence for inter-lexical inhibition under minimal exposure conditions? We believe that there are two crucial differences between our paradigm and those used in past studies. Both of these derive from the fundamental notion that to establish lexical inhibition we must experimentally enhance activation of the inhibiting (newly learned) word, and observe its specific effects on the inhibited (familiar) word. First our use of the VWP allows a more specific measure of the activation of the inhibited word. In contrast, lexical decision, pause detection, or word-spotting may be sensitive to the activation of the target word, but they are also sensitive to other words (e.g., the competitors) and to other factors (e.g. task demands); therefore it is difficult to clearly interpret what they measure. Second, we used Marslen-Wilson's and Warren's (1994) spliced stimulus manipulation to specifically enhance activation of the inhibiting word (the newly learned nonword). This may be essential to observing our effect, because if newly learned words are only weakly active (e.g., configuration has not been completely formed), they may not exert much of an inhibitory effect on familiar words – even if such connections have been formed. It is crucial to point out that if no inhibitory links between the newly learned and familiar words were formed, then boosting the activation of the newly learned word should not have any effect on the activation of the familiar words.

A number of caveats to our interpretation are worth mentioning. First, one might argue that our manipulated stimuli are less representative of natural language processing, and therefore our results may not generalize. We cannot rule this out, but as we've argued, this manipulation was crucial for the logic of our measure, and there is a long history of using unnatural or manipulated stimuli in psycholinguistics for the purpose of disentangling theories. Indeed, psycholinguistics often finds itself in tension between naturalistic observation, and experimental manipulation (c.f., Brown-Schmidt, Gunlogson, & Tanenhaus, 2008), and it is an important goal of future work to determine how and what we believe are the key aspects of this approach to measuring inhibition.

Second, as our task analysis of pause detection, word spotting and lexical decision suggests, precisely what connections are being formed with and without sleep depends largely on our understanding of the testing task. In this regard, a richer understanding of the VWP, particularly with respect to newly learned words, is warranted (c.f., Magnuson, Tanenhaus, Aslin, & Dahan, 2003; Salverda, Brown, & Tanenhaus, 2011). It is not clear if standard critiques of the VWP (e.g., pre-naming of the objects) apply in this case, as the competing words were not displayed. But nonetheless, there is much we still do not understand about this task, even less with newly learned words.

One important concern with our paradigm may come from Dumay and Gaskell (2012) who posit that a form of episodic familiarity with newly learned words may dominate behavior in the earliest periods after learning; given the repetition of items in the VWP task, such concerns may appear to apply – perhaps our results can be attributed to inhibition between episodic memory traces of phonological sequences, and not between consolidated, abstract lexical representations. This seems unlikely to us for several reasons. First and foremost, the time course of the newly learned inhibition effect is quite early (relative to that of known words). Second, such an account predicts *facilitation* (instead of the reported inhibitory effect) during this stage (based on Dumay & Gaskell, 2012), and, if this account is true, it is not clear why such effects have not been observed in other tasks. Third, some of these concerns are motivated by the fact that during the VWP participants were exposed multiple (a total of 12) times to each of the 20 picture quartets, therefore possibly forming audio-visual episodic memories of the relevant stimuli. Our experimental design does not, however, allow for this, as participants were exposed equally to all four splice conditions in each set, so this alone could not yield differences between the experimental conditions. If anything it should have expanded listeners' tolerance for phonetic mismatch, reducing differences between conditions. In addition, participants were exposed to each splicing condition only once, therefore there was no chance for a previously presented stimulus to be repeated. Fourth, it may be suggested that episodic traces formed during training were actually responsible for the observed inhibitory effects. However, given the vastly different contexts between learning and test (the presence of the pictures, the absence of the newly learned words) it seems even less likely that such episodic representations would be engaged during testing, than during more decontextualized tasks like lexical decision. Fourth, Experiment 2 somewhat mitigates this issue by exposing participants to multiple tokens of the new words, promoting more abstraction. Finally, a number of popular accounts of lexical organization suggest that such episodic traces *constitute* the lexicon (Goldinger, 1998), and no current theories of learning or lexical organization posit a clear relationship between episodic memory and lexical organization, consequently we must leave this discussion as speculative at best. However, the broader, and more important point, is that we need a clearer understanding of what factors contribute to performance in the VWP and how lexical organization is related to other cognitive and behavioral processes.

These caveats notwithstanding, our results have significantly enriched our understanding of word learning. First, they have implications for the mechanism(s) underlying word learning: if the two lexical properties (configuration and engagement) emerge concurrently, they can clearly be acquired via the same types of experiences and may also share a common learning

mechanism. The framework we offered in the Introduction provides a useful way to conceptualize this learning process as one in which weights change with experience in roughly the same manner across various sets of connections, even as the specific sets of weights may serve different functions (bottom-up access, inhibition, feedback, etc.).

Second, more broadly, these findings bear on what a lexical representation consists of; if words start to engage with other words immediately as they are learned, this may indicate that inter-lexical inhibition is a core rather than a subsequent, secondary characteristic of words. Inter-lexical connections, thus, may not just indicate the strengthening and/or deepening of word learning that follows initial lexical configuration; rather the ability to interact with other words may be fundamental to the instantiation of lexical representations (Luce & Pisoni, 1998; Vitevitch & Luce, 1998; 1999).

Third, the present results clarify and extend the work of Gaskell, Dumay and colleagues in indicating that lexical competition can occur at the phonological level without requiring any meaning. Lexical inhibition has been documented in several studies (Dahan, Magnuson, Tanenhaus, et al., (2001); Luce & Pisoni, 1998), and predicted by several models (Chen & Mirman, 2012; McClelland & Elman, 1986). However, in earlier studies the locus of the observed effect was left unclear, because the stimuli were typically known words, incorporating both phonology and semantics (e.g. Dahan, Magnuson, Tanenhaus, et al., (2001)), and the locus of inhibition could therefore have been at either phonology or semantics, or at some representational level such as the lemma that is linked to both. The work of Gaskell, Dumay, and colleagues began to indicate that lexical inhibition can arise even without meanings. However, the stimuli used by Gaskell, Dumay and colleagues, even though without explicit meaning, all bore a strong resemblance to one particular familiar word (e.g. *cathedrue* and *cathedral*), so that conceivably, the novel word could have been linked to that known word's meaning, and the inhibition, thus, mediated by semantics after all. The present findings reduce the possibility of such mediation, because the nonword stimuli did not bear a strong resemblance to any *one particular* familiar word (*jod* is similar to many words and hence unlikely to inherit the semantics of any specific one of them). Therefore, the present study offers a clearer answer to what has previously been indicated: newly learned words with no associated meaning can inhibit other known words at the lexical-phonological or lexeme level.

While our results show clear evidence that inter-lexical inhibition can be established without consolidation, they appear to fly in the face of the many studies showing that sleep may be required for this. What are we to make of these studies? First, as we described, there are methodological concerns with the tasks used to measure inhibition in prior studies implicating sleep. Consequently, it is not clear that we can attribute the changes that come with sleep to the emergence of lexical inhibition. However, something clearly is changing. We have shown that lexical engagement can be acquired from the earliest stages of exposure to a word, but sleep may still be necessary for this engagement to become observable in coarser measures of lexical activation, such as lexical decision latencies. Sleep may, for example, be necessary for lexical knowledge to become embedded in the complex system that enables listeners to make lexical decision judgments. In addition, sleep could also improve the strength of configuration, and this may have complex consequences for other

words when filtered through the dynamics of decay and feedback. What is needed is a clearer, perhaps computational understanding of these tasks that can explain how simultaneous activation for multiple items, as well as processes like decay and inhibition, can lead to changes in word-spotting, lexical decision or pause-detection times. Lacking such a description, it is clear from this prior work that some change to the word recognition system occurs with sleep; however, given our results, it is not specifically the emergence of inhibition, and it is not entirely clear how to characterize it.

More generally, by bringing to bear real-time measures like eye-tracking into word learning, this study brings together two approaches addressing issues in different time-scales: on-line lexical dynamics, and the slower time-scale trajectory of word-learning (c.f., Gupta & Tisdale, 2009; McMurray, Horst, & Samuelson, 2012). The combination of these two approaches may reveal interesting aspects of both, as one can inform and enrich the other. For example, while we observed an inhibitory effect of newly learned words in both experiments, there may be differences in the emergence of this effect as a function of learning: when the training is weakened (Experiment 2) the inhibitory effect of newly learned words seems to be delayed (compare Figure 3 to Figure 2). In addition, fine-grained observations during online processing may enable us to test computationally derived predictions about the online dynamics of novel lexical representations acquired under different training conditions. Recent computational models have posited explicit interactions between the online dynamics of lexical processing and longer-term learning (McMurray et al., 2012). These models predict that inhibition between items is helpful, or even necessary, for unsupervised learning both in the context of words (McMurray, Zhao, Kucker, & Samuelson, 2013) and other aspects of language (McMurray, Aslin, & Toscano, 2009). If inhibition developed only with large amounts of exposure and consolidation, this would potentially be problematic for those approaches. Our finding, that inhibition can in fact appear with only minimal training, is however quite consistent with these ideas, and indeed lends support to them.

More broadly, the results of our study are clear. First, as previously suggested, lexical engagement does not require semantics, indicating that inhibition can arise at the lexeme level. Second, and more importantly, lexical engagement, as measured via inter-lexical inhibition, emerges immediately after the first exposures to a novel word form and does not require specialized experiences or consolidation. This indicates that during word learning the lateral connections, which are responsible for a word's engagement with the rest of the lexical system, are built simultaneously with the bottom-up pathways that allow the word to be recognized. Finally, it is important to recall our critical idea that configuration and engagement are both instantiated in basically the same way (connections), even though they are different properties of a word. This makes the implications of simultaneity clear: simultaneous emergence of these properties corresponds to simultaneous weight change. It is therefore parsimonious to suppose that the same learning mechanisms are involved in emergence of both the properties of configuration and engagement.

Acknowledgments

The authors would like to thank Keith Apfelbaum for his valuable input and assistance with various aspects of this project, Dan McEchron for technical assistance and for recruiting participants, and Marcus Galle for assistance with auditory stimulus development. This research was supported by NIH Grant NIDCD R01 DC006499 awarded to PG and NIH Grant DC008089 awarded to BM.

Appendix A

Triplets (in IPA) used in Experiments 1 and 2 to construct the spliced stimuli.

Matching-splice (real word spelling)	Word-splice (real word spelling)	Nonword-splice (either trained or untrained)
bɛɪt (bait)	bɛɪk (bake)	bɛɪp
bɹaɪd (bride)	bɹaɪb (bribe)	bɹaɪg
kæt (cat)	kæp (cap)	kæk
tʃɪk (chick)	tʃɪp (chip)	tʃɪt
dɑrk (dark)	dɑrt (dart)	dɑrp
fɔrk (fork)	fɔrt (fort)	fɔrp
græd (grad)	græb (grab)	græg
hɪp (heap)	hɪt (heat)	hɪk
dʒɑb (job)	dʒɑg (jog)	dʒɑd
lɪp (leap)	lɪk (leak)	lɪt
mʌg (mug)	mʌd (mud)	mʌb
nɛk (neck)	nɛt (net)	nɛp
pɑrt (part)	pɑrk (park)	pɑrp
rɑd (rod)	rɑb (rob)	rɑg
ʃeɪk (shake)	ʃeɪp (shape)	ʃeɪt
steɪt (state)	steɪk (steak)	steɪp
sut (suit)	sup (soup)	suk
tɑrp (tarp)	tɑrt (tart)	tɑrk
wɛb (web)	wɛd (wed)	wɛg
zɪt (zit)	zɪp (zip)	zɪk

Appendix B

Stimuli used in the VWP tasks in Experiments 1 and 2

Target word	Cohort	Unrelated 1	Unrelated 2
bait	boot	head	jug
bride	bread	yacht	vote
cat	cord	beard	blood
chick	check	pig	hook
dark	dog	cloud	ride
fork	fog	god	side
grad	gripe	boat	stork
heap	hood	yard	maid

Target word	Cohort	Unrelated 1	Unrelated 2
job	jet	duck	book
leap	lark	wig	peg
mug	milk	truck	spark
neck	nut	goat	wet
part	pad	trout	black
rod	root	feet	bet
shake	shed	keg	dead
state	stick	red	chart
suit	sword	kid	reed
tarp	toad	vet	jeep
web	wood	drop	cook
zit	zap	raid	feed

References

- Allopenna PD, Magnuson JS, Tanenhaus MK. Tracking the Time Course of Spoken Word Recognition Using Eye Movements: Evidence for Continuous Mapping Models. *Journal of Memory and Language*. 1998; 38(4):419–439.10.1006/jmla.1997.2558
- Apfelbaum K, Blumstein SE, McMurray B. Semantic priming is affected by real-time phonological competition: evidence for continuous cascading systems. *Psychonomic Bulletin & Review*. 2011; 18(1):141–9.10.3758/s13423-010-0039-8 [PubMed: 21327343]
- Baayen, RH. languageR: Data sets and functions with “Analyzing Linguistic Data: A practical introduction to statistics” R package version 0.955. 2009. Retrieved February 04, 2014 from <http://cran.rproject.org/web/packages/languageR/index.html>
- Brown-Schmidt S, Gunlogson C, Tanenhaus MK. Addressees distinguish shared from private information when interpreting questions during interactive conversation. *Cognition*. 2008; 107(3): 1122–34.10.1016/j.cognition.2007.11.005 [PubMed: 18164700]
- Chen Q, Mirman D. Competition and cooperation among similar representations: toward a unified account of facilitative and inhibitory effects of lexical neighbors. *Psychological Review*. 2012; 119(2):417–30.10.1037/a0027175 [PubMed: 22352357]
- Dahan D, Magnuson JS, Tanenhaus MK. Time course of frequency effects in spokenword recognition: evidence from eye movements. *Cognitive Psychology*. 2001; 42(4):317–67.10.1006/cogp.2001.0750 [PubMed: 11368527]
- Dahan D, Magnuson JS, Tanenhaus MK, Hogan EM. Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*. 2001; 16(5-6):507–534.10.1080/01690960143000074
- Davis MH, Di Betta AM, Macdonald MJE, Gaskell MG. Learning and consolidation of novel spoken words. *Journal of Cognitive Neuroscience*. 2009; 21(4):803–20.10.1162/jocn.2009.21059 [PubMed: 18578598]
- Davis MH, Gaskell MG. A complementary systems account of word learning: neural and behavioural evidence. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences*. 2009; 364(1536):3773–800.10.1098/rstb.2009.0111
- Dumay N, Gaskell MG. Sleep-associated changes in the mental representation of spoken words. *Psychological Science*. 2007; 18(1):35–9.10.1111/j.1467-9280.2007.01845.x [PubMed: 17362375]
- Dumay N, Gaskell MG. Overnight lexical consolidation revealed by speech segmentation. *Cognition*. 2012; 123(1):119–32.10.1016/j.cognition.2011.12.009 [PubMed: 22261419]

- Dumay, N.; Gaskell, MG.; Feng, X. A Day in the Life of a Spoken Word; Proceedings of the 26th Annual Meeting of the Cognitive Science Society; Chicago, IL: Mahwah, NJ. Lawrence Erlbaum Associates; 2004. p. 339-344.
- Gaskell MG, Dumay N. Lexical competition and the acquisition of novel words. *Cognition*. 2003; 89(2):105–132.10.1016/S0010-02770300070-2 [PubMed: 12915296]
- Gaskell MG, Marslen-Wilson W. Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive*. 1997
- Goldinger SD. Echoes of echoes? An episodic theory of lexical access. *Psychological Review*. 1998; 105(2):251–279. [PubMed: 9577239]
- Gow DW, Gordon PC. Lexical and prelexical influences on word segmentation: evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance*. 1995; 21(2): 344–59. [PubMed: 7714476]
- Grainger J, Jacobs AM. Orthographic Processing in Visual Word Recognition: A Multiple Read-Out Model. *Psychological Review*. 1996 Apr 29.
- Gupta P, Tisdale J. Word learning, phonological short-term memory, phonotactic probability and long-term memory: towards an integrated framework. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences*. 2009; 364(1536):3755–71.10.1098/rstb.2009.0132
- Karpicke JD, Roediger HL. The critical importance of retrieval for learning. *Science (New York, N Y)*. 2008; 319(5865):966–8.10.1126/science.1152408
- Leach L, Samuel AG. Lexical configuration and lexical engagement: when adults learn new words. *Cognitive Psychology*. 2007; 55(4):306–53.10.1016/j.cogpsych.2007.01.001 [PubMed: 17367775]
- Lindsay S, Gaskell MG. Spaced Learning and the Lexical Integration of Novel Words. 2009:2517–2522.
- Lindsay S, Gaskell MG. Lexical Integration of Novel Words Without Sleep. *Journal of Experimental Psychology Learning, Memory, and Cognition*. 2013; 39(2):608–622.
- Luce PA, Pisoni DB. Recognizing spoken words: the neighborhood activation model. *Ear and Hearing*. 1998; 19(1):1–36. [PubMed: 9504270]
- Magnuson JS, Dixon J, Tanenhaus MK, Aslin RN. The dynamics of lexical competition during spoken word recognition. *Cognitive Science*. 2007; 31(1):133–56.10.1080/03640210709336987 [PubMed: 21635290]
- Magnuson JS, McMurray B, Tanenhaus MK, Aslin RN. Lexical effects on compensation for coarticulation: a tale of two systems? *Cognitive Science*. 2003; 27(5):801–805.10.1016/S0364-0213(03)00067-3
- Magnuson JS, Tanenhaus MK, Aslin RN, Dahan D. The Time Course of Spoken Word Learning and Recognition: Studies With Artificial Lexicons. *Journal of Experimental Psychology: General*. 2003; 132(2):202–227.10.1037/0096-3445.132.2.202 [PubMed: 12825637]
- Marslen-Wilson W. Functional parallelism in spoken word-recognition. *Cognition*. 1987; 25(1-2):71–102. [PubMed: 3581730]
- Marslen-Wilson W, Warren P. Levels of Perceptual Representation and Process in Lexical Access: Words, Phonemes, and Features. *Psychological Review*. 1994; 101(4):653–675. [PubMed: 7984710]
- Mattys SL, Clark JH. Lexical activity in speech processing: evidence from pause detection. *Journal of Memory and Language*. 2002; 47(3):343–359.10.1016/S0749-596X0200037-2
- McClelland JL, Elman JL. The TRACE model of speech perception. *Cognitive Psychology*. 1986; 18(1):1–86.10.1016/0010-02858690015-0 [PubMed: 3753912]
- McClelland JL, Mirman D, Holt LL. Are there interactive processes in speech perception? *Trends in Cognitive Sciences*. 2006; 10(8):363–9.10.1016/j.tics.2006.06.007 [PubMed: 16843037]
- McMurray B, Aslin RN, Toscano JC. Statistical learning of phonetic categories: insights from a computational approach. *Developmental Science*. 2009; 12(3):369–78.10.1111/j.1467-7687.2009.00822.x [PubMed: 19371359]
- McMurray B, Horst JS, Samuelson LK. Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*. 2012; 119(4):831–77.10.1037/a0029872 [PubMed: 23088341]

- McMurray B, Samelson VM, Lee S, Tomblin JB. Individual differences in online spoken word recognition: Implications for SLI. *Cognitive Psychology*. 2010; 60(1):1–39.10.1016/j.cogpsych.2009.06.003 [PubMed: 19836014]
- McMurray B, Tanenhaus MK, Aslin RN. Gradient effects of within-category phonetic variation on lexical access. *Cognition*. 2002; 86(2):B33–42. [PubMed: 12435537]
- Norris D. Shortlist: a connectionist model of continuous speech recognition. *Cognition*. 1994; 52(3): 189–234.10.1016/0010-02779490043-4
- Norris D, McQueen JM, Cutler A. Competition and segmentation in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1995; 21(5):1209–28.
- Norris D, McQueen JM, Cutler A. Merging information in speech recognition: feedback is never necessary. *The Behavioral and Brain Sciences*. 2000; 23(3):299–325. discussion 325–70. [PubMed: 11301575]
- Norris D, McQueen JM, Cutler A. Perceptual learning in speech. *Cognitive Psychology*. 2003; 0285:00006–9.10.1016/S0010-02850300006-9
- Salverda AP, Brown M, Tanenhaus MK. A goal-based perspective on eye movements in visual world studies. *Acta Psychologica*. 2011; 137(2):172–80.10.1016/j.actpsy.2010.09.010 [PubMed: 21067708]
- Schwartz BL, Hashtroudi S. Priming is independent of skill learning. *Journal of Experimental Psychology: Learning Memory and Cognition*. 1991; 17(6):1177–1187.
- Tamminen J, Payne JD, Stickgold R, Wamsley EJ, Gaskell MG. Sleep spindle activity is associated with the integration of new memories and existing knowledge. *The Journal of Neuroscience*. 2010; 30(43):14356–60.10.1523/JNEUROSCI.3028-10.2010 [PubMed: 20980591]
- Tanenhaus MK, Spivey-Knowlton MJ, Eberhard KM, Sedivy JC. Integration of visual and linguistic information in spoken language comprehension. *Science*. 1995; 268(5217):1632–4. [PubMed: 7777863]
- Vitevitch MS, Luce PA. When Words Compete: Levels of Processing in Perception of Spoken Words. *Psychological Science*. 1998; 9(4):325–329.10.1111/1467-9280.00064
- Vitevitch MS, Luce PA. Probabilistic Phonotactics and Neighborhood Activation in Spoken Word Recognition. *Journal of Memory and Language*. 1999; 40(3):374–408.10.1006/jmla.1998.2618
- Vrooomen J, van Zon M, de Gelder B. Cues to speech segmentation: evidence from juncture misperceptions and word spotting. *Memory & Cognition*. 1996; 24(6):744–55. [PubMed: 8961819]

Highlights

- When listeners recognize spoken words, similar sounding words inhibit each other.
- Sleep was thought to be essential for the acquisition of these inhibitory links.
- We report inter-lexical inhibition after 20 mins of training and without sleep.
- Knowledge of a word sound and its links to other words might emerge simultaneously.

Figure 1.

Visualization of a multi-layer lexical network. Word form level representations are shown as localist for ease of depiction rather than a theoretical commitment to such representations.

Figure 2.
Proportion of trials on which participant was fixating the target at each 4 ms time slice as a function of splice condition in Experiment 1.

Figure 3.

Competition effect from newly learned words (untrained-nonword-splice minus trained-nonword-splice; grey line) and previously known words (untrained-nonword-splice minus other-word-splice; black line) in Experiment 1.

Figure 4.
Proportion of trials on which participant was fixating the target at each 4 ms time slice as a function of splice condition in Experiment 2.

Figure 5.

Competition effect from newly learned words (untrained-nonword-splice minus trained-nonword-splice; grey line) and previously known words (untrained-nonword-splice minus other-word-splice; black line) in Experiment 2.