

**On-line learning from input versus off-line memory
evolution in adult word learning: Effects of neighborhood
density and phonologically-related practice**

Journal:	<i>Journal of Speech, Language, and Hearing Research</i>
Manuscript ID:	JSLHR-L-13-0150.R2
Manuscript Type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Storkel, Holly; University of Kansas, Speech-Language-Hearing: Sciences and Disorders Bontempo, Daniel; University of Kansas, Bureau of Child Research Pak, Natalie; University of Kansas, Speech-Language-Hearing: Sciences and Disorders
Keywords:	Vocabulary, Word Learning, Adults, Neighborhood Density, Input, Memory Evolution

Running Head: ADULT WORD LEARNING

RESEARCH ARTICLE

On-line learning from input versus off-line memory evolution in adult word learning: Effects of neighborhood density and phonologically-related practice

Holly L. Storkel
Daniel E. Bontempo
Natalie S. Pak

University of Kansas

Submitted (02/25/2014), *Journal of Speech, Language, and Hearing Research*

Send correspondence to:
Holly L. Storkel
Speech-Language-Hearing: Sciences & Disorders
University of Kansas
1000 Sunnyside Ave
Dole Human Development Center, 3001
Lawrence, KS 66045-7555
784-864-0497
hstorkel@ku.edu

Abstract

1
2
3
4
5 *Purpose.* This study investigates adult word learning to determine how neighborhood density and
6
7 practice across phonologically-related training sets influence on-line learning from input during
8
9 training versus off-line memory evolution during no-training gaps.
10

11
12
13
14 *Method.* Sixty-one adults were randomly assigned to learn low or high density nonwords. Within
15
16 each density condition, participants were trained on one set of words and then were trained on a
17
18 second set of words, consisting of phonological neighbors of the first set. Learning was measured
19
20 in a picture-naming test. Data were analyzed using multilevel modeling and spline regression.
21
22
23

24
25
26 *Results.* Steep learning during input was observed, with new words from dense neighborhoods
27
28 and new words that were neighbors of recently learned words (i.e., second set words) being
29
30 learned better than other words. In terms of memory evolution, large and significant forgetting
31
32 was observed during 1-week gaps in training. Effects of density and practice during memory
33
34 evolution were opposite of those during input. Specifically, forgetting was greater for high
35
36 density and second set words than for low density and first set words.
37
38
39
40

41
42
43 *Conclusion.* High phonological similarity, regardless of source (i.e., known words or recent
44
45 training), appears to facilitate on-line learning from input but seems to impede off-line memory
46
47 evolution.
48

49
50
51
52 Key words: vocabulary, word learning, adults, neighborhood density, input, memory evolution
53
54
55
56
57
58
59
60

1
2
3 On-line learning from input versus off-line memory evolution in adult word learning: Effects of
4
5 neighborhood density and phonologically-related practice
6

7 People typically learn new words through exposure. The prevailing assumption is that
8 words are learned from the input that a person hears. That is, learning a word is, at least partly, a
9 function of experiences with the word. However, there is long-standing evidence that learning
10 words can occur in the absence of additional input (Gaskell & Dumay, 2003; Leach & Samuel,
11 2007; Rice, Oetting, Marquis, Bode, & Pae, 1994; Storkel, 2001, 2003; Storkel & Lee, 2011).
12 Specifically, past studies show that word learning performance may continue to improve after
13 training is withdrawn. This pattern of results suggests that there are (at least) two mechanisms
14 that underlie adult word learning: on-line learning from input as well as “off-line” learning in the
15 absence of input (Davis, Betta, Macdonald, & Gaskell, 2008; Davis & Gaskell, 2009; Dumay &
16 Gaskell, 2007; Gaskell & Dumay, 2003; Leach & Samuel, 2007; Tamminen & Gaskell, 2008;
17 Tamminen, Payne, Stickgold, Wamsley, & Gaskell, 2010). This two mechanism framework fits
18 well with current theories of learning and memory. In particular, theories of learning and
19 memory differentiate between change associated with swift episodic learning from input and
20 change associated with slower memory consolidation during sleep, when it is thought that newly
21 learned information is stabilized and integrated with existing information (McClelland,
22 McNaughton, & O'Reilly, 1995; Norman & O'Reilly, 2003; O'Reilly & Rudy, 2000). Moreover,
23 these theories suggest distinct neuroanatomical structures for each mechanism with the
24 hippocampus being crucial for learning from input and the neocortex being critical for memory
25 consolidation (McClelland et al., 1995; Norman & O'Reilly, 2003; O'Reilly & Rudy, 2000).
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51

52 The discussion of memory consolidation has typically emphasized the benefits for word
53 learning performance, creating a rather positive view of retention during gaps without training.
54 Often overlooked in this discussion is a more negative aspect of withdrawing training: forgetting.
55 In fact, in numerous studies, word learning performance declines when training is withdrawn
56
57
58
59
60

1
2
3 (Storkel, Bontempo, Aschenbrenner, Maekawa, & Lee, 2013; Tamminen & Gaskell, 2013; Vlach
4 & Sandhofer, 2012). Recently, Vlach and Sandhofer (2012) documented a curvilinear pattern of
5 forgetting of a newly learned word for both adults and children. Specifically, for both groups,
6 performance dropped rapidly between an immediate test and a 1-week delayed test, and then
7 dropped less rapidly from the 1-week delayed test to a 1-month delayed test, as an asymptote
8 was reached (i.e., performance at floor). While the evidence of memory consolidation as well as
9 forgetting during gaps in word learning training may seem contradictory, emerging theories of
10 memory suggest these counter processes may be critical in creating stable but flexible
11 representations (Hardt, Nader, & Nadel, 2013; Stickgold & Walker, 2013). In particular,
12 forgetting allows non-essential details of a word learning experience to be lost over time so that
13 subsequent word learning experiences will reinforce the essential elements (Vlach & Sandhofer,
14 2012). Recent theories of memory have referred to this array of off-line memory changes as
15 *memory evolution* to better reflect the idea that memories can be forgotten via decay or
16 interference or retained via consolidation, integration, and generalization (Hardt et al., 2013;
17 Stickgold & Walker, 2013). The purpose of the current study is to build on this framework that
18 differentiates on-line learning during input from off-line memory evolution during no-input gaps
19 by examining two variables that are known to influence adult word learning, specifically
20 neighborhood density of the words to be learned and practice across different training sets, to
21 determine how each variable influences each process.

22 *Neighborhood Density*

23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
Neighborhood density is the number of words that are phonologically similar to a given word (Luce & Pisoni, 1998). During training, adults learn words with many neighbors (i.e., high density) more accurately than words with few neighbors (i.e., low density, Storkel, Armbruster, & Hogan, 2006). Thus, high density appears to facilitate learning from input, possibly due to the influence of density on working memory. That is, high density items are held more accurately in

1
2
3 working memory than low density items (Roodenrys & Hinton, 2002; Thomson, Richardson, &
4 Goswami, 2005; Thorn & Frankish, 2005), potentially leading to the creation of a more accurate
5 and/or detailed representation in long-term memory. Unfortunately, Storkel and colleagues
6 (2006) did not examine retention across a no-training gap. Thus, it is unknown how
7 neighborhood density influences memory evolution in word learning by adults (but see Storkel &
8 Lee, 2011 for discussion of child word learning). Thus, one purpose of the current study was to
9 examine the effect of neighborhood density on adult word learning during training and across no-
10 training gaps to provide a clearer picture of how density influences on-line learning from input
11 versus off-line memory evolution in adult word learning.
12
13
14
15
16
17
18
19
20
21
22

23 *Practice across Training Sets*

24
25
26 Gershkoff-Stowe and Han (2007, 2009) showed that learning of one set of words could
27 facilitate learning of a second set of words for both children and adults. In Gershkoff-Stowe and
28 Han (2009), adults received two training sessions for a 1st set of novel words, which were learned
29 to a relatively high level of accuracy. Then, they were trained on a 2nd set of novel words in a
30 third training session. When performance was compared at similar points in training (i.e., first
31 training session for each set), adults were more accurate for the 2nd set of words than the 1st set of
32 words. This improvement for the 2nd set of words is sometimes referred to as a *practice effect*. It
33 was hypothesized that the 1st training set primed learning of the 2nd training set by activating
34 semantic or conceptual networks or perhaps even the entire lexical network. Thus, it appears that
35 learning from input can be enhanced by repeated experience learning novel words. However, the
36 effect of this repeated practice on off-line memory evolution is not known because retention of
37 novel words after a no-input gap was not compared for the 1st versus 2nd set. In addition, the 1st
38 and 2nd sets were related in a general or contextual way only. That is, the sets were “matched
39 approximately for object category” (p. 686, Gershkoff-Stowe & Hahn, 2007, presumably
40 artifacts and natural kinds) and were trained in the same manner in the same environment (i.e.,
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

laboratory environment). It is unclear whether a more specific similarity relationship, namely phonological similarity, would alter the effect of practice. Based on the authors' account, a specific similarity relationship should still result in a positive effect of practice on learning different sets of training words because practice on the 1st set of words would activate the existing network of phonologically similar words, namely the word's neighborhood, potentially facilitating adding new words from the 2nd set to that same neighborhood. The current study tests this hypothesis by examining learning of a 2nd set of nonwords that are phonological neighbors of a 1st set of taught nonwords. In addition, performance is examined during training and following no-training gaps to elucidate the influence of practice on on-line learning from input versus off-line memory evolution in the absence of input.

Purpose

The purpose of this study was to determine how neighborhood density and practice across phonologically-related training sets influence on-line learning from input during training versus off-line memory evolution during no-training gaps. These two particular variables are of interest because they index different aspects of phonological similarity. In particular, neighborhood density indexes the phonological similarity between a new word and already known words. Thus, this variable captures similarity between the new input and the existing contents of memory. In contrast, the relationship between the training sets indexes phonological similarity between new words, capturing similarity within the recent and current input. Examining how these two variables influence on-line learning from input during training versus off-line memory evolution during no-training gaps will shed light on whether the effect of phonological similarity is the same on these two processes regardless of the locus of the similarity: long-term memory in the case of density versus recent past input in the case of phonological practice. Based on studies of adults, it is predicted that high density nonwords will be learned better than low density nonwords during training and that 2nd set nonwords will be

1
2 learned better than 1st set nonwords during training. Thus, both types of phonological similarity
3
4 are expected to facilitate on-line learning from input during training. Prior studies of adults have
5
6 not examined retention of nonwords varying by density or practice so it is unclear whether either
7
8 or both types of similarity will enhance off-line memory evolution.
9
10

11 Method

12 *Participants*

13
14
15
16
17 Sixty-one adults between the ages of 18 – 31 years participated. The participants were
18
19 recruited from the university community and received either payment or partial course credit.
20
21 Based on self-report, all participants had normal hearing and were monolingual native English
22
23 speakers. Likewise, none of the participants reported a history of speech, language, or cognitive
24
25 delay. The participants scored within normal limits (i.e., received a standard score of 85 or
26
27 higher) on a receptive vocabulary test (i.e., Peabody Picture Vocabulary Test - 4th edition, Dunn
28
29 & Dunn, 2007). Participants were randomized to one of two stimulus conditions: (1) low density;
30
31 (2) high density. The participant characteristics for the low versus high density groups are shown
32
33 in Table 1. As shown in Table 1, the groups were comparable on all areas measured.
34
35
36
37

38 *Stimuli*

39
40 The stimuli used for this study were 48 nonwords, consisting of consonant-vowel-
41
42 consonant (CVC) sequences that were not found in an online corpus of real English words
43
44 (Storkel, 2013; Storkel & Hoover, 2010). Therefore, these nonwords should not be part of a
45
46 native English speaker's lexicon. The 48 nonwords are shown in Table S1 of the online
47
48 supplemental materials. Nonwords were selected from the corpus to manipulate neighborhood
49
50 density, one of the independent variables. Neighborhood density refers to the total number of
51
52 words in a corpus that are phonologically similar to a given word (Luce & Pisoni, 1998).
53
54 Phonological similarity is defined as all the words that differ from the target word by exactly one
55
56 sound segment (i.e., a substitution, addition, or deletion in any word position). Low density
57
58
59
60

1
2
3 neighborhoods have few phonologically similar neighbors, whereas high density neighborhoods
4
5 have many phonologically similar neighbors. One set of 24 nonwords from low density
6
7 neighborhoods and one set of 24 nonwords from high density neighborhoods were selected. Low
8
9 density was defined as a density $\leq 25^{\text{th}}$ percentile of the CVC pool (cf., Storkel & Lee, 2011).
10
11 High density was defined as a density $\geq 75^{\text{th}}$ percentile of the CVC pool (cf., Storkel & Lee,
12
13 2011). The low density nonwords had 2-7 real-word neighbors ($M = 5.42$, $SD = 1.35$), whereas
14
15 high density nonwords had 15-22 neighbors ($M = 17.04$, $SD = 2.12$).
16
17
18

19
20 Each set of 24 nonwords (i.e., low vs. high density) consisted of 12 pairs of phonological
21
22 neighbors, as shown in Table S1 of the supplement. This pairing was needed to examine the
23
24 effect of phonologically related training sets. That is, one member of the pair was trained first
25
26 (e.g., Group A /ver/) during initial training sessions, followed by training on the second member
27
28 of the pair (e.g., Group B /ner/) during later training sessions. This creates the right conditions
29
30 for testing the effect of practice learning phonologically related nonwords (i.e., does learning
31
32 /ver/ subsequently influence learning of /ner/). The set of items that were trained first are referred
33
34 to as the 1st set. In complement, the set of items that were trained second are referred to as the 2nd
35
36 set. Note that half of the participants in each density condition were trained in the Group A then
37
38 Group B order, whereas the other half of the participants in each condition were trained in the
39
40 reverse order: Group B then Group A. These groupings are shown in Table S1 of the supplement,
41
42 and the neighborhood density associated with each group is shown in Table 2. As shown in
43
44 Table 2, Group A was similar in density to Group B within a given density condition (low or
45
46 high). Thus, density is matched across the 1st and 2nd set. Of the 12 pairs of phonological
47
48 neighbors in each density condition, four pairs were taken from each of three types of neighbors:
49
50 (1) neighbors that shared the initial consonant and vowel (i.e., CV_ neighbors), such as /fub/-
51
52 /fuk/ in the low density condition; (2) neighbors that shared the vowel and final consonant (i.e.,
53
54 _VC neighbors), such as /ver/ - /ner/ in the low density condition; (3) neighbors that shared the
55
56
57
58
59
60

1
2
3 consonants (i.e., C_C neighbors), such as /zis/ - /zos/ in the low density condition. Each neighbor
4
5 pair came from a different phonological neighborhood. That is, aside from these neighbor pairs,
6
7 the nonwords in a condition were not neighbors of one another.
8

9
10 Stimuli also were selected to control phonotactic probability, namely the likelihood of
11
12 occurrence of a sound or sound sequence in words of the language. This is based on two
13
14 measures: positional segment sum and biphone sum (Storkel, 2004). The *positional segment sum*
15
16 is the sum of the positional segment frequency of each sound in the word. *Positional segment*
17
18 *frequency*, in turn, is the sum of the log frequency of every word in the corpus with the specific
19
20 sound in the target word position, divided by the sum of the log frequency of every word in the
21
22 corpus with any sound in the target position. *Biphone sum* is the sum of the biphone frequency of
23
24 each adjacent pair of sounds in the word. Similar to positional segment frequency, *biphone*
25
26 *frequency* is the sum of the log frequency of every word in the corpus with the target pair of
27
28 sounds in the target word position, divided by the sum of the log frequency of every word in the
29
30 corpus containing any sound in the target word position.
31
32
33
34
35

36 The nonwords used as stimuli in this study had “mid” phonotactic probability, defined as
37
38 adult segment sum and biphone sums between the 25th and 75th percentiles. Table 2 provides the
39
40 phonotactic probability values for each nonword group for low and high density conditions. The
41
42 phonotactic probability for low density and high density conditions were compared to ensure
43
44 minimal variability. There was no significant difference for the segment sum, $F(1, 44) = 0.33, p$
45
46 $> 0.50, \eta_p^2 < .01$. Segment sum for low density nonwords ($M = .10, SD = .02, \text{range } .08 - .13$) was
47
48 similar to that of high density nonwords ($M = .11, SD = .02, \text{range } .08 - .13$). In contrast, there
49
50 was a significant effect for biphone sum, $F(1, 44) = 11.57, p = 0.001, \eta_p^2 = .21$. Here, the low
51
52 density nonwords ($M = .0020, SD = .0008, \text{range } .0012 - .0035$) tended to be lower in biphone
53
54 sum than the high density nonwords ($M = .0027, SD = .0007, \text{range } .0013 - .0036$). However, the
55
56 effect size for biphone sum was relatively small, especially when compared to the effect size for
57
58
59
60

1
2
3 density, $F(1, 44) = 503.38, p < 0.001, \eta_p^2 = .92$. Although biphone sum is not perfectly matched
4
5 across the low and high density conditions, the two conditions vary more in density than biphone
6
7 sum and are well matched on segment sum.
8

9
10 Each nonword was paired with a nonobject. The same group of 24 nonobjects was used
11
12 for the low and high density conditions. Thus, there is no difference in nonobject characteristics
13
14 across the two density conditions. The nonobjects were selected from those published by Kroll &
15
16 Potter (1984). The reference number of the selected nonobjects are shown in Table S1 of the
17
18 online supplement. Within each density condition, one group of 12 nonwords (e.g., Group A)
19
20 was paired with a group of 12 nonobjects (e.g., Group 1), and the second group of 12 nonwords
21
22 (e.g., Group B) was paired with the other 12 nonobjects (e.g., Group 2). Recall that these
23
24 groupings correspond to the order of training, namely 1st set versus 2nd set. Consequently, two
25
26 nonobject characteristics were matched across groups/sets. Adult semantic set size for each
27
28 nonobject refers to the number of different semantic neighbors reported by two or more adult
29
30 participants (Storkel & Adlof, 2009a). Group 1 nonobjects ($M = 10.5, SD = 0.5$, range 10-11)
31
32 had similar set sizes to Group 2 nonobjects ($M = 10.5, SD = 0.5$, range 10-11). Objectlikeness is
33
34 the degree to which the nonobject resembles a real object on a 7-point scale, according to ratings
35
36 published by Kroll & Potter (1984). Group 1 nonobjects ($M = 4.4, SD = 0.9$, range 3.2-5.6) were
37
38 similar to Group 2 nonobjects on objectlikeness ratings ($M = 4.5, SD = 0.9$, range 3.2-5.9). Note
39
40 that half of the participants received Group A nonwords paired with Group 1 nonobjects and
41
42 Group B nonwords paired with Group 2 nonobjects. The other half of the participants received
43
44 the reverse pairing (i.e., Group A with Group 2; Group B with Group 1).
45
46
47
48
49
50

51 52 *Procedures*

53
54 Each participant attended four sessions, at approximately seven-day intervals ($M = 7.6$,
55
56 $SD = 2.8$, range 1-26). Recall that participants were randomized to receive either low density or
57
58 high density stimuli. Each session consisted of multiple training-testing cycles administered via a
59
60

1
2 computer with DirectRT software (Jarvis, 2002). During the *training phase*, participants
3 passively listened to recordings of the nonword stimuli and viewed the nonobject on the
4 computer. The nonword stimuli were presented in sentences in the word-final position. For each
5 nonword in the set, the participant heard “This is a _____. Listen closely, it’s called a _____.
6 Remember, it’s a _____. Look at the _____. Don’t forget the _____.” Thus, each participant received
7 five exposures to each nonword per training cycle. The *testing phase* consisted of a production
8 task, in which the participant was presented with a picture of each nonobject and asked to name
9 the corresponding nonword. The accuracy of the participants’ responses on the production task
10 served as the primary outcome measure of the study.
11
12
13
14
15
16
17
18
19
20
21
22
23

24 Session 1 focused exclusively on training and testing of the 1st set of stimuli (i.e., Group
25 A or B nonwords paired with Group 1 or 2 nonobjects) for the participant’s assigned density
26 condition (low vs. high density). Session 1 began with a baseline naming test. In this test, the
27 nonobject pictures were presented and participants were asked to guess the name. As expected,
28 no participant produced the correct nonword for a nonobject at baseline. Instead, participants
29 tended to use real words to describe some aspect of the nonobject (e.g., “wheels” for a nonobject
30 that appeared to have wheels) or to name a real object that they perceived as similar to the
31 nonobject (e.g., “pretzel” for a nonobject with a twisty shape like a pretzel). Baseline testing was
32 followed by three cycles, each including one training phase and one testing phase. Tests during
33 this session measure learning from input. Session 2 began with a test for the 1st set stimuli, which
34 measured the participant’s retention of 1st set stimuli after the one week break. This test taps
35 memory evolution over the gap in training. This post-test was followed by an additional three
36 cycles of training and testing for 1st set stimuli. Again, these tests during training measure
37 learning from input. Session 3 also began with a post-test to measure retention of 1st set stimuli
38 following the one week break. As with the first retention test, this second retention test taps
39 memory evolution in the absence of input. 2nd set stimuli were then trained following the same
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

procedures (i.e., baseline testing followed by three cycles of training and testing). These tests tap learning from input for the 2nd set stimuli. In Session 4, participants performed a post-test for the 2nd set nonwords. This retention test taps memory evolution for the 2nd set stimuli. The receptive vocabulary test (Dunn & Dunn, 2007) also was administered during Session 4.

In total, each participant received 30 exposures to each nonword in the 1st set and 15 exposures to each nonword in 2nd set during training. The 1st set received greater training than the 2nd set to ensure that the 1st set would be learned to a high level, establishing the necessary conditions for the practice effect. Excluding baseline testing, each participant performed eight production tests for 1st set stimuli (i.e., 6 tests during input; 2 tests of memory evolution in the absence of input) and four production tests for 2nd set stimuli (i.e., 3 tests during input; 1 test of memory evolution in the absence of input).

Procedural reliability was calculated for 13% of the sample using a checklist of items related to correct examiner behavior (e.g., appropriate instructions provided, appropriate nonword-nonobject set administered) and correct equipment function (e.g., all items presented appropriate number of times). Procedural reliability was computed as the percent of items on the checklist that were scored as appropriately administered divided by the total number of items on the checklist. Procedural reliability was high ($M = 96\%$, $SD = 6\%$, range 82% - 100%). The one instance of low reliability (i.e., 82%) was due to equipment issues (i.e., issues with the quality of the recording for the production test).

Scoring

All sessions were audio- and video-recorded. Participant responses on the production task were phonetically transcribed and scored. Interjudge transcription reliability was computed for 13% of the sample. Mean percent agreement was 96% ($SD = 2\%$, range 93% - 98%). Based on these transcriptions, responses were scored as correct or incorrect, and this served as the dependent variable for the primary analysis. Correct responses were those with all three

1
2
3 phonemes produced in the correct positions. Responses were scored incorrect if they contained
4
5 substitutions or deletions of any of the target phonemes. Interjudge scoring reliability was
6
7 computed for 13% of the sample, with mean agreement of 100% ($SD = 1\%$, range 98% - 100%).
8

9
10 Incorrect responses were further scored for a secondary descriptive analysis. First,
11
12 incorrect responses were coded for the number of phonemes correct (i.e., 0, 1, or 2) to quantify
13
14 phonological similarity between the target and the incorrect response. Second, incorrect
15
16 responses that shared 2 of 3 phonemes with the target (e.g., response /ves/ for target /ver/) were
17
18 compared to the list of real word neighbors of the target. Incorrect responses were scored as a
19
20 real word neighbor if they shared 3 of 3 phonemes with one of the items on the real word
21
22 neighbor list (e.g., response /ves/ for target /ver/ scored as real word neighbor “vase”). Likewise,
23
24 incorrect responses in the 2nd set that shared 2 of 3 phonemes with the target (e.g., response /ner/
25
26 for target /ver/) were compared to the 1st set neighbor and scored as a 1st set neighbor if 3 of 3
27
28 phonemes were shared (e.g., /ner/ is a set neighbor for /ver/). These two additional measures for
29
30 phonologically similar errors quantify the influence of real word neighbors and 1st set trained
31
32 items. Third, all incorrect responses were compared to the list of trained nonwords and scored as
33
34 a trained nonword error if 3 of 3 phonemes were shared. This measure quantifies the extent to
35
36 which participants are confusing nonwords that are being trained together. Lastly, for incorrect
37
38 responses that shared 0 of 3 phonemes, it was noted whether the participant gave no response
39
40 (e.g., said nothing, “I can’t remember”) to quantify the extent to which participants had not
41
42 formed any representation or had experienced a retrieval failure. Error coding was completed by
43
44 a first judge and then by a second judge with discrepancies resolved through discussion.
45
46
47
48
49
50
51

52 Results

53 54 *Primary Analysis Approach*

55
56 The accuracy data were analyzed using multilevel modeling. Multilevel modeling
57
58 (MLM), also called mixed effects modeling, hierarchical linear modeling, or random coefficient
59
60

1
2
3 modeling, is becoming a preferred method for analyzing repeated measures data because it
4
5 allows for a variety of variance/covariance structures, thus being more flexible regarding
6
7 dependencies arising from repeated measures as well as accommodating missing and/or
8
9 unbalanced data (Cnaan, Laird, & Slasor, 1997; Gueorguieva & Krystal, 2004; Hoffman &
10
11 Rovine, 2007; Misangyi, LePine, Algina, & Goeddeke, 2006; Nezlek, Schroder-Abe, & Schutz,
12
13 2006; Quene & van den Bergh, 2004). Moreover, random effects of participants and items can be
14
15 accommodated in the same analysis by incorporating crossed random intercepts, and this is
16
17 becoming the favored analysis approach for psycholinguistic data (cf., Baayen, Davidson, &
18
19 Bates, 2008; Locker, Hoffman, & Bovaird, 2007; Quene & van den Bergh, 2008). Note that the
20
21 dependent variable for this study was accuracy (i.e., correct or incorrect), which is a binary
22
23 variable. Thus, a logistic MLM was used. The analysis proceeded in several model building
24
25 steps; however, only the final model is reported here because certain variables are less
26
27 interpretable in preliminary models that exclude other variables (e.g., main effects are less
28
29 interpretable/meaningful when a significant interaction is present). As shown in Table 3 (i.e.,
30
31 rows 16 and 17), McKelvey and Zavoina's r-squared (Hox, 2010; Windmeijer, 1995) and
32
33 Somers' D (Group, 2014; Newson, 2002) were used as indicators of model fit. The reported final
34
35 model in Table 3 was among the models with the highest r-squared and Somers' D, with very
36
37 minimal change in these values when other non-significant effects were added to the model.
38
39
40
41
42
43
44

45 *Participant and Item Effects*

46
47 Table 3 shows the final model. First, the crossed random effects of participants and items
48
49 (i.e., nonwords) were examined to better understand variability. To facilitate insight into the
50
51 magnitude of individual differences, participant and item level variance was expressed as a
52
53 median odds ratio (MOR, Merlo et al., 2006). Conceptually, the MOR conveys the median
54
55 increase in the odds of a correct response between a pair of participants or items that are alike on
56
57 all other covariates. The MOR has the further advantage of being on the same scale as the odds
58
59
60

1
2 ratio (OR), which was used as the effect size for the fixed effects (e.g., neighborhood density).

3
4
5 In terms of interpreting the magnitude of an effect stated in an odds ratio metric, there are not
6
7 straightforward guidelines because the interpretation is somewhat dependent on the rate of the
8
9 event of interest (Chen, Cohen, & Chen, 2010). However, generally values near 1.5 are
10
11 interpreted as small effects, those near 2-3 are interpreted as moderate effects, and those near 5-6
12
13 and higher are considered large effects (Chen et al., 2010). As shown in Table 3 (i.e., rows 14
14
15 and 15), the MOR for participants was 1.92 (95% CI = 1.70 – 2.22) and the MOR for items was
16
17 3.00 (95% CI = 2.44 – 3.88). Thus, the variability between participants is associated with a
18
19 median difference of 1.92 in the odds of a correct response when comparing two randomly
20
21 drawn participants who are alike on other covariates. This is a moderate effect size. Likewise, the
22
23 variability between nonwords is associated with a median difference of 3.00 in the odds of a
24
25 correct response between two randomly drawn nonwords that are alike on other covariates.
26
27 Again, this is a moderate effect size. Taken together, random effects of participants and items
28
29 were moderate in size with the variance of items being slightly larger than that of participants.
30
31
32
33
34

35
36 Significant predictors of participant and item variability also were included in the model,
37
38 as shown in Table 3 (rows 1-3). In terms of participant characteristics, the odds of a correct
39
40 response were 1.60 (95% CI = 1.00 – 2.56) times higher for males relative to females. In terms of
41
42 mean accuracy, males had a mean accuracy of 66% (95% CI = 60-73%) compared to 58% (95%
43
44 CI = 54-62%) for females. In the final model reported in Table 3, the effect of gender was no
45
46 longer significant. However, it was significant in earlier models and just missed traditional
47
48 criteria for significance in the final model. For this reason, gender was retained in the model as a
49
50 reference point for future research. Turning to vocabulary scores, the odds of a correct response
51
52 were 1.02 (95% CI = 1.01 – 1.04) times higher for each one point increase in standard score on
53
54 the *Peabody Picture Vocabulary Test*. Splitting the sample at the median standard score of 113
55
56 showed that mean accuracy of participants with scores at or above the median was 64% (95% CI
57
58
59
60

1
2 = 59-69%), whereas mean accuracy of participants with scores below the median was 55% (95%
3
4 CI = 51-59%). In terms of item characteristics, the odds of a correct response were 1.22 (95% CI
5
6 = 1.06 – 1.39) times higher for a one unit decrease in semantic set size, replicating Storkel and
7
8 Adlof (2009b). However, mean accuracy values were quite similar. Specifically, mean accuracy
9
10 was 59% (95% CI = 53-65%) for higher set size items compared to 60% (95% CI = 54-67%) for
11
12 lower set size items. This small mean difference suggests that this term might be tapping other
13
14 characteristics of the nonobjects that were not entered in the model (e.g., strength of the first
15
16 semantic associate, Storkel & Adlof, 2009a). This effect should be interpreted with caution. Note
17
18 that the magnitude of these participant and item level predictors was relatively small.
19
20
21
22

23 *Input versus Memory Evolution Tests*

24
25
26 Turning to variables that address the research questions, Figure 1 shows the data for test
27
28 (each individual panel), neighborhood density as a continuous variable (x-axis), and set (top
29
30 versus bottom row). Starting with the effect of test, recall that the absolute timing of different
31
32 tests varied. That is, the first three tests (excluding baseline) occurred during input (i.e., first
33
34 training phase). Thus, these tests are separated by minutes. In contrast, the fourth test occurred
35
36 after a gap without any training. Thus, test 3 and test 4 are separated by approximately 1 week.
37
38 The next set of tests (i.e., tests 5, 6, and 7) again occurred during input (i.e., second training
39
40 phase), returning to a separation in tests on the order of minutes. Finally, the last test (i.e., test 8)
41
42 again occurred after a no training gap, creating a separation of approximately 1 week from the
43
44 previous test. To address this variability in absolute timing of tests and in the mental activities
45
46 that occurred between tests (i.e., on-line learning from input during training vs. off-line memory
47
48 evolution during no-training gaps), spline regression and multiple intercepts were used to model
49
50 the effect of test. With linear splines, the effect of an explanatory variable (i.e. test) is assumed to
51
52 be piecewise linear on a specified number of segments separated by knots where the slope
53
54 changes (Gould, 1993; Panis, 1994). Coding can be for the slope or the change in slope (cf.,
55
56
57
58
59
60

1
2
3 Storkel et al., 2013). To capture abrupt changes in level (i.e., after one week of memory
4
5 evolution), an additional intercept parameter can be paired with each knot. Coding for this model
6
7 is shown in Table S2 of the online supplemental materials.
8

9
10 The fixed effects of test are shown in rows 4-7 in Table 3. The first coefficient (row 4 of
11
12 Table 3) captures the slope across all tests that were administered, excluding baseline. This term
13
14 generally captures learning from input because additional coefficients capture memory evolution,
15
16 as described later. This all-test slope was significant, indicating that the odds of a correct
17
18 response were 3.75 (95% CI = 3.36 – 4.18) times higher for each subsequent test relative to the
19
20 immediately preceding test. Note that this odds ratio indicates a moderate effect of training. This
21
22 also can be seen in Figure 1, where accuracy generally increases across each panel. The second
23
24 coefficient (row 5 of Table 3) captures the abrupt change across the first 1-week gap in training.
25
26 This is the first retention test, capturing off-line memory evolution. As shown in Table 3, this
27
28 effect was significant with a large effect size. Specifically, the odds of a correct response were
29
30 25.00 (95% CI = 18.18 – 34.48) times lower for the test point following the no training gap
31
32 relative to the test point preceding the no training gap. Thus, participants forgot the newly
33
34 learned words when training was withdrawn for 1-week. This is apparent in Figure 1 by
35
36 comparing the Test 3 and Test 4 panels, which clearly show a precipitous drop in performance
37
38 from Test 3 to Test 4. Note that performance at Test 4 looks quite similar to performance at Test
39
40 1, indicating that the forgetting that occurred during the 1-week gap negated much of the
41
42 learning that occurred during input. The third coefficient (row 6 of Table 3) was intended to
43
44 capture any change in slope that might occur in during later training after this first gap. As shown
45
46 in Table 3, there was a non-significant change in slope in these later tests, with the trend being
47
48 for a slightly shallower slope across this training-testing period. As shown in the Test 4-8 panels
49
50 of Figure 1, accuracy began to approach ceiling in later tests, accounting for this non-
51
52 significantly shallower slope during later training and testing. Although this effect was non-
53
54
55
56
57
58
59
60

1
2
3 significant, it was retained in the model because the full set of predictors is needed to describe
4
5 the complexity of time in this design. Note that models dropping these non-significant effects did
6
7 not lead to appreciable changes in odds ratios or p-values for the other effects in the model. The
8
9 fourth coefficient (row 7 of Table 3) captured abrupt change across the second 1-week gap in
10
11 training. This is the second retention test, providing an additional window into off-line memory
12
13 evolution. As shown in Table 3, this effect was significant with a large effect size and again
14
15 indicated forgetting when training was withdrawn. Specifically, the odds of a correct response
16
17 were 29.41 (95% CI = 18.52 – 45.45) times lower for the test point following the no training gap
18
19 relative to the test point preceding the no training gap. Comparison of the Test 7 and Test 8
20
21 panels in Figure 1 also shows a clear drop in accuracy consistent with forgetting. Once again the
22
23 forgetting that occurred during the 1-week gap in training negated much of the learning that
24
25 occurred during training, as can be seen by comparing Test 8 to Tests 4 and 5. Note that the odds
26
27 ratio for the two retention coefficients (i.e., 25.00 versus 29.41) is quite similar, indicating
28
29 similarly large effects of forgetting across each 1-week gap.
30
31
32
33
34

35 *Neighborhood Density*

36
37
38 Turning to neighborhood density, the effect of density was modeled using both main
39
40 effects and interactions (see rows 8-11 of Table 3). For all of these terms, density was modeled
41
42 as a continuous predictor (i.e., raw density), rather than as dichotomously coded variable (i.e.,
43
44 low vs. high). As shown in row 8 of Table 3, the main effect of density was not significant. The
45
46 direction of the main effect is for high density nonwords to be responded to more accurately than
47
48 low density nonwords. In particular, the odds of a correct response were 1.03 (95% CI = 0.97 –
49
50 1.10) times higher for each one neighbor increase in density. This non-significant main effect of
51
52 density was retained in the model because interaction terms involving density were significant.
53
54 First, the interaction between density and tests (row 9 of Table 3) was significant with a small
55
56 effect size, indicating that the difference between low and high density words widens as training
57
58
59
60

1
2 progresses. The fit line in each panel of Figure 1 depicts the effect of density. Across panels,
3 which correspond to test, the steepness of this line generally increases, indicating a greater
4 advantage for high over low density words as training progresses. Second, the interaction
5 between density and the first retention test (row 10 of Table 3) is significant with a small effect
6 size. Here, there is a greater drop in word learning performance across the no-training gap for
7 higher density nonwords than for lower density nonwords. As shown in Figure 1, the fit line for
8 density has a shallower slope at Test 4 than at Test 3, indicating a more precipitous drop in
9 accuracy for high density nonwords across the no-training gap, yielding accuracy that is more
10 similar across low and high density nonwords after memory evolution has taken place. Third, the
11 interaction between density and the second retention test (row 11 of Table 3) is significant with a
12 small effect size. The interpretation is similar to the first retention test. As shown in Figure 1, the
13 fit line for density has a shallower slope at Test 8 than at Test 7, indicating greater forgetting for
14 high density nonwords across the second no-training gap. Lastly, an interaction between density
15 and later tests (i.e., Tests 4-8 that occurred after the first gap in training) was tested but found to
16 be non-significant ($OR = 0.99$, $95\% CI = 0.97 - 1.01$). Thus, this interaction was deleted from
17 the final model and is not shown in Table 3. This non-significant interaction suggests that the
18 relationship between density and test is consistent across all tests, without any significant change
19 occurring in the later tests as ceiling is approached. To summarize, high density nonwords were
20 learned better than low density nonwords during training, but high density nonwords experienced
21 greater forgetting during the no-training gaps than low density nonwords.

22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50 Turning to the error data, Table 4 shows the error pattern for input versus retention tests
51 for low and high density items. As shown in Table 4, errors during input testing (i.e., Tests 1, 2,
52 3, 5, 6, and 7) were primarily close phonological approximations of the target, sharing 2 of 3
53 phonemes with the target. Notably, the percentage of close phonological approximations was
54 relatively similar across low and high density items. However, these phonological
55
56
57
58
59
60

1
2 approximations were frequently real word neighbors for high density items but rarely real word
3
4 neighbors for low density items. The percentages of other error types during input testing were
5
6 similar across low and high density items. Turning to retention testing (i.e., Tests 4, 8), errors
7
8 tended to be dissimilar to the target with the predominant error pattern being 0 of 3 phonemes
9
10 shared. This fits well with the accuracy analysis by further showing the impact of forgetting.
11
12 Specifically, participants have now either completely lost even partial representations for an item
13
14 or have lost access to partial (or full) representations of an item. Like the accuracy analysis, there
15
16 is evidence in the error analysis that the impact of forgetting is greater for high density than for
17
18 low density. During training, close phonological approximations (i.e., 2 of 3 phonemes shared)
19
20 were relatively similar across low and high density items. However, at retention, fewer close
21
22 phonological approximations are observed for high than low density. Likewise, the no response
23
24 rate during retention testing is higher for high density items than for low density items, indicating
25
26 that forgetting has a larger impact on high than low density items. Overall, phonological
27
28 approximation errors were observed for both low and high density items during training, and
29
30 these were most often real words for high density items. In contrast, after memory evolution,
31
32 errors were dissimilar from the targets for both low and high density items, with fewer
33
34 phonological approximations and more no response errors for high than low density items.
35
36
37
38
39
40
41

42 *Practice across Phonologically-Related Training Sets*

43
44
45 Turning to the influence of practice on learning phonologically similar nonwords, the
46
47 effect of Set was modeled using both main effects and interactions (see rows 12-13 of Table 3).
48
49 The main effect of Set was significant with a small effect size (see row 12 of Table 3).
50
51 Specifically, the odds of a correct response were 1.23 (95% CI = 1.06 – 1.43) times higher for
52
53 the 2nd set relative to the 1st set. Thus, learning the 1st set appeared to facilitate learning of the 2nd
54
55 set. Although this can be seen by comparing the upper (Set 1) and lower (Set 2) rows of plots in
56
57 Figure 1, it is a bit difficult. For clearer illustration with mean values, at Test 1, mean accuracy is
58
59
60

1
2
3 36% (95% CI = 30-42%) for 2nd set nonwords compared to 31% (95% CI = 25-36%) for 1st set
4
5 nonwords. Likewise, at Test 2, mean accuracy is 63% (95% CI = 55-70%) for 2nd set nonwords
6
7 compared to 60% (95% CI = 53-67%) for 1st set nonwords. At Test 3, mean accuracy is 73%
8
9 (95% CI = 65-81%) for 2nd set nonwords compared to 71% (95% CI = 64-79%) for 1st set
10
11 nonwords. Although it appears that the effect of set may be diminishing across tests, this
12
13 interaction was not significant (OR = 0.93, 95% CI = 0.77 –1.12) and is not shown in the Table 3
14
15 final model. However, there was a significant interaction of set with the first retention test (see
16
17 row 13 of Table 3), and this effect had a moderate effect size. Here, the effect of set reverses,
18
19 with the 1st set showing a smaller drop in accuracy compared to the 2nd set across the no-training
20
21 gap. This again may be difficult to ascertain from Figure 1. In terms of mean values for clearer
22
23 comparison, mean accuracy for 1st set at Test 3 is 71% (95% CI = 64-79%) versus 39% (95%CI
24
25 = 33-44%) at Test 4. In comparison, mean accuracy for the 2nd set at Test 3 is 73% (95% CI =
26
27 = 65-81%), but mean accuracy drops more dramatically to 29% (95%CI = 24-34%) at Test 4.
28
29
30
31
32
33 Overall, the 2nd set appears to be learned more accurately than the 1st set during training but the
34
35 1st set appears to be retained more accurately than the 2nd set across the no training gap.
36
37

38
39 Turning to the error data, Table 5 shows the error pattern for input versus retention tests
40
41 for 1st and 2nd set items. As shown in Table 5, errors during input testing (i.e., Tests 1, 2, 3, 5, 6,
42
43 and 7) were primarily close phonological approximations of the target, sharing 2 of 3 phonemes.
44
45 In general, the error types were similar across 1st and 2nd set items during training. Turning to
46
47 retention testing (i.e., Tests 4, 8), the predominant error is 0 of 3 phonemes shared. As previously
48
49 discussed for density, this reflects forgetting across the gap in training. More important for
50
51 understanding the effect of set is that differences between 1st set and 2nd set errors begin to
52
53 emerge at retention. Specifically, 0 of 3 phoneme errors are more prevalent for the 2nd set than
54
55 the 1st set. In complement, 2 of 3 phoneme errors are more prevalent for the 1st set than the 2nd
56
57 set. These data fit well with the hypothesis that forgetting had a greater impact on the 2nd set than
58
59
60

1
2
3 the 1st set. Here, partial representations (or access to partial representation) appears to be retained
4
5 better for the 1st set than the 2nd set. In addition, the number of 1st set neighbors that are produced
6
7 as errors to 2nd set targets doubled from the input tests to the retention tests. This suggests that 1st
8
9 set neighbors may be interfering with retention of 2nd set targets. A related observation is that the
10
11 2nd set shows more other trained nonword errors than the 1st set at retention. This could indicate
12
13 confusion between 1st set and 2nd set items that are not neighbors of one another and/or confusion
14
15 among 2nd set items. Either scenario fits the hypothesis that forgetting has a greater impact on 2nd
16
17 set than 1st set items, potentially due to confusion among trained items during the gap.
18
19
20

21 *Additional Models*

22
23
24 An interaction between density and set was explored but this term was not significant
25
26 (OR = 1.01, 95% CI = 0.99 – 1.03). Thus, this interaction was not retained in the final model.
27
28 Likewise, an interaction between density and vocabulary scores was explored to address the
29
30 possibility of individual differences in the effect of density. However, this interaction term was
31
32 not significant (OR = 0.999, 95% CI = 0.997 – 1.000) and, therefore, was not retained. In
33
34 addition, the influence of the type of neighbor on the effect of set was explored. These data are
35
36 shown in Table S3 of the on-line supplement. Recall that there were three ways that 1st set and
37
38 2nd set nonwords could be related. Specifically, they could share the initial consonant and vowel
39
40 (i.e., CV_ neighbors), the vowel and final consonant (i.e., _VC neighbors), or the consonants
41
42 (i.e., C_C neighbors). It is possible that the effect of set could be stronger or weaker across these
43
44 different types of relationships so models incorporating interactions between set and type of
45
46 neighbor were explored (see Table S3 of the supplement). These models showed that the effect
47
48 of set was consistent across these three types of neighbors.
49
50
51
52
53
54

55 Discussion

56
57 The purpose of this study was to determine how neighborhood density and practice
58
59 across phonologically-related training sets influence on-line learning from input during training
60

1
2
3 versus off-line memory evolution during no-training gaps. In terms of learning from input, steep
4
5 learning during training was observed with improvements diminishing slightly (albeit non-
6
7 significantly) at later test points as ceiling performance was approached. During learning from
8
9 input, new words from dense neighborhoods were responded to more accurately than new words
10
11 from sparse neighborhoods, replicating prior work (Storkel et al., 2006). In addition, words that
12
13 were neighbors of previously learned words (i.e., 2nd set words) were learned better than words
14
15 that were not neighbors of previously learned words (i.e., 1st set words), replicating prior work
16
17 (Gershkoff-Stowe & Hahn, 2007, 2009). In terms of the overall pattern of memory evolution,
18
19 large and significant forgetting was observed across both 1-week gaps in training. Effects of
20
21 density and practice during memory evolution were opposite of those observed during input.
22
23 Specifically, high density words experienced greater forgetting than low density words, and
24
25 words that were neighbors of previously learned words (i.e., 2nd set words) experienced greater
26
27 forgetting than words that were not neighbors of previously learned words (i.e., 1st set words).
28
29 This asymmetry in the effect of the variables across on-line learning and off-line memory
30
31 evolution supports the need to differentiate these two mechanisms to understand word learning.
32
33

34 *On-line learning during training versus off-line memory evolution*

35
36
37
38 As in most prior studies, adult participants demonstrated strong and robust on-line
39
40 learning during training. In contrast, off-line memory evolution appears to vary across studies
41
42 with some studies showing improvements in performance, namely memory consolidation, across
43
44 gaps in training (Gaskell & Dumay, 2003; Leach & Samuel, 2007; Rice et al., 1994; Storkel,
45
46 2001, 2003; Storkel & Lee, 2011), and others showing declines in performance, namely
47
48 forgetting, across gaps in training (Storkel et al., 2013; Tamminen & Gaskell, 2013; Vlach &
49
50 Sandhofer, 2012). There are numerous differences in methodology that could account for why
51
52 some studies show evidence of memory consolidation while others show evidence of forgetting.
53
54
55
56
57
58
59 For the current study, which showed clear evidence of forgetting during gaps in training, one
60

1
2
3 contributing factor likely was the size of the gap in training, which was 1-week. As shown by
4
5 Vlach and Sandhofer (2012) for word learning, forgetting is a curvilinear function such that the
6
7 rate of forgetting is initially rapid but then slows over time. Vlach and Sandhofer tested word
8
9 learning immediately following training, 1-week after training, and 1-month after training.
10
11 Visual inspection of their figure suggests that 1-week after training approximates the turn in the
12
13 curvilinear function that marks the end of the rapid phase of forgetting. Thus, forgetting may
14
15 have been minimized if the gap in training had been smaller. An additional factor that may have
16
17 contributed to forgetting in the current study is that the training exposure was not particularly
18
19 rich. That is, the training solely relied on passive auditory exposure, rather than more active
20
21 participation (e.g., repeating the word) or more detailed description of the nonobjects (e.g.,
22
23 highlighting perceptual or semantic characteristics) or nonwords (e.g., highlighting phonological
24
25 characteristics). In addition, although participants in this study had an opportunity to generate the
26
27 words during repeated naming testing, they did not receive any feedback on the accuracy of their
28
29 responses. Thus, a richer and more interactive training procedure may have minimized forgetting
30
31 or may have even resulted in memory consolidation (Vlach & Sandhofer, 2012).
32
33
34
35
36
37

38
39 Regardless of the reason that forgetting was observed in this study rather than memory
40
41 consolidation, the impact of forgetting on performance was striking. That is, visual inspection of
42
43 Figure 1 showed that forgetting after a gap in training lead to performance that was similar to
44
45 performance observed at the start of training. Moreover, comparison of odds ratios further shows
46
47 that the impact of forgetting was 6 to 7 times larger than the impact of one additional training
48
49 (i.e., the inverse-retention odds ratios of 25-29 are approximately 6-7 times larger than the
50
51 testing odds ratio of 3.75). In essence, forgetting negated the effects of several training cycles.
52
53
54 This underscores the importance of understanding what can be done during training to either
55
56 minimize forgetting or enhance consolidation during gaps in training, especially in educational
57
58
59
60

1
2
3 and clinical contexts. In these types of naturalistic contexts, the magnitude of forgetting
4
5 observed in the current study would severely limit overall gains.
6

7 *Neighborhood density*

8
9
10 As in prior studies, words with many neighbors (i.e., high density) were learned more
11
12 accurately than words with few neighbors (i.e., low density, Storkel et al., 2006). The prior
13
14 account of this effect is that high density sound sequences are held more accurately in working
15
16 memory than low density sound sequences (Roodenrys & Hinton, 2002; Thomson et al., 2005;
17
18 Thorn & Frankish, 2005). The more accurate representation in working memory may then
19
20 support the creation of a more accurate and/or detailed representation in long-term memory,
21
22 leading to more accurate responding for high density words during training than for low density
23
24 words. One might be inclined to find support for this hypothesis in the error analysis, where
25
26 more real words were reported for high density than for low density words, which could suggest
27
28 that the real words are activated and providing support for learning. However, it is unclear how
29
30 to interpret these real word errors. By definition, high density words have more real word
31
32 neighbors than low density words. Thus, a one phoneme change to a high density word has a
33
34 greater probability of yielding a real word neighbor than that same change to a low density word.
35
36 For example, the substitution of /b/ for /z/ results in a real word for the high density nonword
37
38 /zin/ (i.e., “bean”) but not for the low density nonword /zis/ (i.e., “beas” is not a real word). For
39
40 this reason, it is unclear whether real word errors reflect an interaction between the target and the
41
42 neighbors or are actually phonological substitutions that result in real words by chance.
43
44
45
46
47
48

49
50 Turning to memory evolution, this facilitation of learning high density words during
51
52 training is then countered by greater forgetting of high density words during the 1-week gap in
53
54 training. This greater forgetting of high density words than low density words suggests that
55
56 interference may occur between newly learned words and existing known words in long-term
57
58 memory (see Hardt et al., 2013 for a discussion of interference in neocortical memory systems).
59
60

1
2
3 Because a high density word is similar to many more existing known words than a low density
4
5 word is, interference would be greater for high density than low density words, accounting for
6
7 the greater forgetting of high than low density words. This hypothesis is supported by the error
8
9 analysis which showed fewer phonological approximations and more no responses for high
10
11 density than for low density items, indicating that partial representations or access to partial
12
13 representations had been lost for high density items after memory evolution. This pattern
14
15 indicates greater interference for high density than for low density items, although it fails to
16
17 reveal the source of that interference. Thus, whether real word neighbors are actually the cause
18
19 of greater interference for high density words requires additional verification.
20
21
22

23 24 *Practice across Training Sets*

25
26 This study extends prior work on learning similar sets of words (Gershkoff-Stowe &
27
28 Hahn, 2007, 2009) to the learning of phonologically similar sets of words. As with conceptually
29
30 similar set of words, learning a second set of phonologically similar words facilitated word
31
32 learning relative to learning a first set of words. In keeping with the hypothesized mechanism for
33
34 conceptually similar word sets, learning the first set of words presumably primes the existing
35
36 phonological neighborhoods of these words. Thus, when a second word from the same
37
38 phonological neighborhood is taught (i.e., 2nd set), it is easier to learn the second word than when
39
40 the neighborhood has not been previously primed (i.e., 1st set). A more novel observation from
41
42 the current study is that this facilitative effect of practice is reversed when retention is examined.
43
44 Here, 2nd set words experienced greater forgetting than 1st set words across a 1-week gap in
45
46 training. This finding suggests that retention of the words learned in the 1st set interfered with
47
48 retention of the words learned in the 2nd set. That is, the phonological overlap between two
49
50 recently learned words leads to confusion between the two words, negatively impacting the later
51
52 learned (i.e., 2nd set) word. This hypothesis is supported by the finding that 1st set nonwords were
53
54 reported in place of 2nd set nonwords more frequently after the gap.
55
56
57
58
59
60

1
2
3 It is important to note that it cannot be unequivocally determined whether the
4
5 phonological overlap between the sets was crucial in obtaining the effects observed. A control
6
7 condition where phonologically unrelated words were taught in each set was not included. Thus,
8
9 it is possible that the effect might be attributable to a broader mechanism. That is, activating the
10
11 lexicon to learn any kind of word may facilitate learning any type of subsequent word, regardless
12
13 of whether these words are related in some way or not. Likewise, the 1st set of words might
14
15 induce greater forgetting in the 2nd set of words, even if the words are unrelated. There is some
16
17 suggestion of this in the error analysis where other trained nonwords were reported in place of
18
19 2nd set nonwords more frequently after the gap. The other trained nonwords are unrelated to the
20
21 target (due to the way the stimuli were selected). Taken together, there is evidence of
22
23 interference from 1st set related nonwords as well as unrelated nonwords on the 2nd set nonwords
24
25 after memory evolution. This suggests the need to further tease apart the role of similarity among
26
27 trained items on learning from input and memory evolution.
28
29
30
31
32

33 *Conclusion*

34
35
36 Two types of phonological similarity were examined: (1) similarity between a new word
37
38 and existing words in long-term memory (i.e., neighborhood density); (2) similarity between
39
40 recent and current input (i.e., 1st set vs. 2nd set neighbors). The results across these two types of
41
42 phonological similarity were the same. Specifically, phonological similarity, regardless of
43
44 source, appears to facilitate on-line learning from input, replicating prior studies. In contrast,
45
46 phonological similarity, regardless of source, appears to impede off-line memory evolution. That
47
48 is, greater forgetting across a gap in training was observed when similarity was high than when it
49
50 was low. This suggests that high similarity may increase interference compared to low similarity.
51
52 Overall, the degree of forgetting was striking. When combined with the other findings, there is a
53
54 clear need to explore memory evolution in word learning to determine how to minimize
55
56 forgetting and enhance memory consolidation.
57
58
59
60

Author Note

The project described was supported by Grants DC 08095, DC 05803, and HD02528 from NIH. The contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. The second author was supported by the Analytic Techniques and Technology (ATT) Core of the Center for Biobehavioral Neurosciences of Communication Disorders (BNCD, DC05803). We would like to thank staff of the Participant Recruitment and Management Core (PARC) of the Center for Biobehavioral Neurosciences of Communication Disorders (BNCD, DC05803) for assistance with recruitment of preschools and children; staff of the Word and Sound Learning Lab (supported by DC 08095) for their contributions to stimulus creation, data collection, data processing, and reliability calculations.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

References

- 1
2
3
4
5 Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed
6
7 random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390-412.
8
9
10 Chen, H., Cohen, P., & Chen, S. (2010). How big is a big odds ratio? Interpreting the magnitudes
11
12 of odds ratios in epidemiological studies. *Communication in Statistics - Simulation and*
13
14 *Computation*, *39*, 860-864.
15
16
17 Cnaan, A., Laird, N. M., & Slasor, P. (1997). Using the general linear mixed model to analyze
18
19 unbalanced repeated measures and longitudinal data. *Statistics in Medicine*, *16*, 2349-
20
21 2380.
22
23
24 Davis, M. H., Betta, A. M. D., Macdonald, M. J. E., & Gaskell, M. G. (2008). Learning and
25
26 consolidation of novel spoken words. *Journal of Cognitive Neuroscience*, *21*, 803-820.
27
28
29 Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning:
30
31 neural and behavioural evidence. *Philosophical Transactions of the Royal Society of*
32
33 *London, Series B, Biological Science*, *364*(1536), 3773-3800. doi: 364/1536/3773 [pii]
34
35 10.1098/rstb.2009.0111 [doi]
36
37
38 Dumay, N., & Gaskell, M. G. (2007). Sleep-Associated Changes in the Mental Representation of
39
40 Spoken Words. *Psychological Science*, *18*(1), 35-39.
41
42
43 Dunn, L. M., & Dunn, L. M. (2007). *Peabody picture vocabulary test—4th edition*. San Antonio,
44
45 TX: Pearson.
46
47
48 Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words.
49
50 *Cognition*, *89*(2), 105-132.
51
52
53 Gershkoff-Stowe, L., & Hahn, E. R. (2007). Fast mapping skills in the developing lexicon.
54
55 *Journal of Speech, Language, and Hearing Research*, *50*, 682-697.
56
57
58
59
60

- 1
2
3 Gershkoff-Stowe, L., & Hahn, E. R. (2009). *Differential effects of lexical neighborhoods on*
4
5 *receptive and productive word learning*. Paper presented at the Symposium for Research
6
7 on Child Development, Denver, CO.
8
9
10 Gould, W. (1993). Linear splines and piecewise linear functions. *Stata Technical Bulletin*, 15,
11
12 13-17.
13
14 Group, U. S. C. (2014). How can I get a Somers' D after logistic regression in Stata? Retrieved
15
16 January 8, 2014, from <http://www.ats.ucla.edu/stat/stata/faq/somersdlogit.htm>
17
18
19 Gueorguieva, R., & Krystal, J. H. (2004). Move over ANOVA: Progress in analyzing repeated
20
21 measures data and its reflection in papers published in the Archives of General
22
23 Psychiatry. *Archives of General Psychiatry*, 61, 310-317.
24
25
26 Hardt, O., Nader, K., & Nadel, L. (2013). Decay happens: The role of active forgetting in
27
28 memory. *Trends in Cognitive Sciences*, 23(3), 111-120.
29
30
31 Hoffman, L., & Rovine, M. (2007). Multilevel models for the experimental psychologist:
32
33 Foundations and illustrative examples. *Behavior Research Methods*, 39, 101-117.
34
35
36 Hox, J. (2010). *Multilevel analysis: Techniques and applications* (2nd edition ed.). New York:
37
38 Routledge.
39
40 Jarvis, B. G. (2002). DirectRT research software (Version 2002). New York, NY: Empirisoft.
41
42
43 Kroll, J. F., & Potter, M. C. (1984). Recognizing words, pictures, and concepts: A comparison of
44
45 lexical, object, and reality decisions. *Journal of Verbal Learning & Verbal Behavior*,
46
47 23(1), 39-66.
48
49
50 Leach, L., & Samuel, A. G. (2007). Lexical configuration and lexical engagement: When adults
51
52 learn new words. *Cognitive Psychology*, 55, 306-353.
53
54
55 Locker, L., Jr., Hoffman, L., & Bovaird, J. A. (2007). On the use of multilevel modeling as an
56
57 alternative to items analysis in psycholinguistic research. *Behavior Research Methods*,
58
59 39(4), 723-730.
60

- 1
2
3 Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation
4
5 model. *Ear & Hearing, 19*, 1-36.
6
- 7 McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary
8
9 learning systems in the hippocampus and neocortex: Insights from the successes and
10
11 failures of connectionist models of learning and memory. *Psychological Review, 102*,
12
13 419-457.
14
15
- 16 Merlo, J., Chaix, B., Ohlsson, H., Beckman, A., Johnell, K., Hjerpe, P., . . . Larsen, K. (2006). A
17
18 brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of
19
20 clustering in multilevel logistic regression to investigate contextual phenomena. *Journal*
21
22 *of epidemiology and community health, 60*, 290-297.
23
24
25
- 26 Misangyi, V. F., LePine, J. A., Algina, J., & Goeddeke, F. (2006). The adequacy of repeated
27
28 measures regression for multilevel research: Comparisons with repeated measures
29
30 ANOVA, multivariate repeated measures ANOVA, and multilevel modeling across
31
32 various multilevel research designs. *Organizational Research Methods, 9*, 5-28.
33
34
35
- 36 Newson, R. (2002). Parameters behind "nonparametric" statistics: Kendall's tau, Somers' D and
37
38 median differences. *Stata Journal, 2*, 45-64.
39
- 40 Nezlek, J. B., Schroder-Abe, M., & Schutz, A. (2006). Multilevel analyses in psychological
41
42 research. Advantages and potential of multi-level random coefficient modeling.
43
44 *Psychologische Rundschau, 57*, 213-223.
45
46
- 47 Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions
48
49 to recognition memory: A complementary-learning-systems approach. *Psychological*
50
51 *Review, 110*, 611-646.
52
53
- 54 O'Reilly, R. C., & Rudy, J. W. (2000). Computational principles of learning in the neocortex and
55
56 hippocampus. *Hippocampus, 10*, 389-397.
57
58
59
60

- 1
2
3 Panis, C. (1994). Linear splines and piecewise linear functions. *Stata Technical Bulletin*, 18, 27-
4
5 29.
6
7
8 Quene, H., & van den Bergh, H. (2004). On multi-level modeling of data from repeated
9
10 measures designs: A tutorial. *Speech Communication*, 43, 103-121.
11
12 Quene, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed
13
14 random effects and with binomial data. *Journal of memory & language*, 59(4), 413-425.
15
16
17 Rice, M. L., Oetting, J. B., Marquis, J., Bode, J., & Pae, S. (1994). Frequency of input effects on
18
19 word comprehension of children with specific language impairment. *Journal of Speech &*
20
21 *Hearing Research*, 37(1), 106-121.
22
23
24 Roodenrys, S., & Hinton, M. (2002). Sublexical or lexical effects on serial recall of nonwords?
25
26 *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(1), 29-33.
27
28 doi: 10.1037/0278-7393.28.1.29
29
30
31 Stickgold, R., & Walker, M. P. (2013). Sleep-dependent memory triage: Evolving generalization
32
33 through selective processing. *Nature Neuroscience*, 16, 139-145.
34
35
36 Storkel, H. L. (2001). Learning new words: Phonotactic probability in language development.
37
38 *Journal of Speech, Language, and Hearing Research*, 44(6), 1321-1337.
39
40
41 Storkel, H. L. (2003). Learning new words II: Phonotactic probability in verb learning. *Journal*
42
43 *of Speech, Language, and Hearing Research*, 46(6), 1312-1323.
44
45
46 Storkel, H. L. (2004). Methods for minimizing the confounding effects of word length in the
47
48 analysis of phonotactic probability and neighborhood density. *Journal of Speech,*
49
50 *Language, and Hearing Research*, 47(6), 1454-1468.
51
52
53 Storkel, H. L. (2013). A corpus of consonant-vowel-consonant (CVC) real words and nonwords:
54
55 Comparison of phonotactic probability, neighborhood density, and consonant age-of-
56
57 acquisition. *Behavior Research Methods*, 45, 1159-1167.
58
59
60

1
2
3 Storkel, H. L., & Adlof, S. M. (2009a). Adult and child semantic neighbors of the Kroll and
4
5 Potter (1984) nonobjects. *Journal of Speech, Language and Hearing Research, 52*(2),
6
7 289-305. doi: 1092-4388_2009_07-0174 [pii]

8
9
10 10.1044/1092-4388(2009/07-0174) [doi]

11
12 Storkel, H. L., & Adlof, S. M. (2009b). The effect of semantic set size on word learning by
13
14 preschool children. *Journal of Speech, Language and Hearing Research, 52*(2), 306-320.
15
16 doi: 1092-4388_2009_07-0175 [pii]

17
18
19 10.1044/1092-4388(2009/07-0175) [doi]

20
21 Storkel, H. L., Armbruster, J., & Hogan, T. P. (2006). Differentiating phonotactic probability and
22
23 neighborhood density in adult word learning. *Journal of Speech, Language, and Hearing*
24
25 *Research, 49*(6), 1175-1192.
26
27

28
29 Storkel, H. L., Bontempo, D. E., Aschenbrenner, A. J., Maekawa, J., & Lee, S. Y. (2013). The
30
31 effect of incremental changes in phonotactic probability and neighborhood density on
32
33 word learning by preschool children. *Journal of Speech, Language, and Hearing*
34
35 *Research, 56*, 1689-1700.
36
37

38
39 Storkel, H. L., & Hoover, J. R. (2010). An on-line calculator to compute phonotactic probability
40
41 and neighborhood density based on child corpora of spoken American English. *Behavior*
42
43 *Research Methods, 42*, 497-506.
44

45
46 Storkel, H. L., & Lee, S. Y. (2011). The independent effects of phonotactic probability and
47
48 neighborhood density on lexical acquisition by preschool children. *Language and*
49
50 *Cognitive Processes, 26*, 191-211. doi: 10.1080/01690961003787609
51

52
53 Tamminen, J., & Gaskell, M. G. (2008). Newly learned spoken words show long-term lexical
54
55 competition effects. *Q J Exp Psychol (Colchester), 61*(3), 361-371. doi: 783032454 [pii]

56
57 10.1080/17470210701634545 [doi]
58
59
60

- 1
2
3 Tamminen, J., & Gaskell, M. G. (2013). Novel word integration in the mental lexicon: Evidence
4
5 from unmasked and masked semantic priming. *Quarterly Journal of Experimental*
6
7 *Psychology, 66*, 1001-1025.
8
- 9
10 Tamminen, J., Payne, J. D., Stickgold, R., Wamsley, E. J., & Gaskell, M. G. (2010). Sleep
11
12 spindle activity is associated with the integration of new memories and existing
13
14 knowledge. *The Journal of Neuroscience, 30*(43), 14356-14360.
15
16
- 17 Thomson, J. M., Richardson, U., & Goswami, U. (2005). Phonological similarity neighborhoods
18
19 and children's short-term memory: typical development and dyslexia. *Memory &*
20
21 *Cognition, 33*(7), 1210-1219.
22
23
- 24 Thorn, A. S., & Frankish, C. R. (2005). Long-term knowledge effects on serial recall of
25
26 nonwords are not exclusively lexical. *Journal of Experimental Psychology: Learning,*
27
28 *Memory, and Cognition, 31*(4), 729-735. doi: 2005-08130-010 [pii]
29
30
31 10.1037/0278-7393.31.4.729 [doi]
32
- 33 Vlach, H. A., & Sandhofer, C. M. (2012). Fast mapping across time: Memory processes support
34
35 children's retention of learned words. *Frontiers in Psychology, 3*, 1-8. doi:
36
37 10.3389/fpsyg.2012.00046
38
39
- 40 Windmeijer, F. A. G. (1995). Goodness-of-fit measures in binary choice models. *Econometric*
41
42 *Reviews, 14*, 101-116.
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table 1. *Participant characteristics for each stimulus condition.*

	Low Density Stimulus	High Density Stimulus
	Condition	Condition
Number	29	32
Gender	Females = 24	Females = 26
	Males = 5	Males = 6
Age ²	<i>M</i> = 22 years	<i>M</i> = 22 years
	<i>SD</i> = 3 years	<i>SD</i> = 3 years
	Range = 18-29 years	Range = 18-31 years
PPVT-4 ^{1, 2}	<i>M</i> = 109	<i>M</i> = 115
	<i>SD</i> = 13	<i>SD</i> = 12
	Range = 89-136	Range = 88-140
Race	White = 28	White = 28
	Asian = 1	Asian = 2
	American Indian = 0	American Indian = 1
	Unknown = 0	Unknown = 1
Ethnicity	Non-Hispanic = 29	Non-Hispanic = 30
	Hispanic = 0	Hispanic = 1
	Unknown = 0	Unknown = 1

¹Standard score on the *Peabody Picture Vocabulary Test – 4th edition* (Dunn & Dunn, 2007).

²Differences between the groups were not statistically significant, all $t(59) < 1.7$, all $p > 0.10$, $\eta_p^2 < 0.05$.

Table 2. *Stimuli characteristics.*

	Low Density	High Density
	<i>M (SD)</i>	<i>M (SD)</i>
	Group A	
Phonotactic Probability: Segment Sum	0.10 (0.02)	0.11 (0.02)
Phonotactic Probability: Biphone Sum	0.002 (0.001)	0.003 (0.001)
Neighborhood Density	6 (1)	17 (2)
	Group B	
Phonotactic Probability: Segment Sum	0.11 (0.02)	0.11 (0.01)
Phonotactic Probability: Biphone Sum	0.002 (0.001)	0.003 (0.001)
Neighborhood Density	5 (2)	17 (2)

Note: Low Density and High Density do not differ significantly for segment sum, $F(1, 44) = 0.33, p > 0.50, \eta_p^2 < .01$; but do differ significantly for biphone sum, $F(1, 44) = 11.57, p = 0.001, \eta_p^2 = .21$, and neighborhood density (as intended), $F(1, 44) = 503.38, p < 0.001, \eta_p^2 = .92$.

Table 3. *Final Multi-level Model.*

Row #	Variable	Odds Ratio (95% confidence interval)	Reciprocal for OR < 1 ¹ (95% confidence interval)
1	Gender	1.60 (1.00 – 2.56)	
2	PPVT-4 Standard Score	1.02 (1.01 – 1.04)**	
3	Semantic Set Size	0.82 (0.72 – 0.94)** ¹	1.22 (1.06 – 1.39)
4	All Tests (Test 1-8) Slope	3.75 (3.36 - 4.18)***	
5	Retention 1 Intercept (1-week gap 1)	0.04 (0.03 – 0.06)*** ¹	25.00 (18.18 – 34.48)
6	Later Tests (Tests 4-8) Slope	0.89 (0.77 – 1.01) ¹	1.13 (0.99 – 1.30)
7	Retention 2 Intercept (1-week gap 2)	0.03 (0.02 – 0.05)*** ¹	29.41 (18.52 – 45.45)
8	Neighborhood Density	1.03 (0.97 – 1.10)	
9	Density x All Tests Slope	1.04 (1.03 – 1.05)***	
10	Density x Retention 1 Intercept	0.89 (0.85 – 0.92)*** ¹	1.13 (1.09 – 1.17)
11	Density x Retention 2 Intercept	0.89 (0.85 – 0.94)*** ¹	1.12 (1.07 – 1.18)
12	Set	1.23 (1.06 – 1.43)**	
13	Set x Retention 1 Intercept	0.29 (0.22 – 0.38)*** ¹	3.45 (2.61 – 4.55)
14	MOR for Participants	1.92 (1.70 – 2.22)	
15	MOR for Items	3.00 (2.44 – 3.88)	
16	McKelvey & Zavoina's r^2	0.27	
17	Somers' D	0.50 (0.46 - 0.54)	

* $p < .05$, ** $p < .01$, *** $p < .001$; All models included an overall intercept term that serves as the traditional constant. ¹For OR < 1, a correct response is less likely for a higher than a lower value of the variable. The third column shows the reciprocal of the OR for ease of interpretation.

Table 4. *Descriptive error analysis for input versus retention tests by density.*

Phonemes Correct	Tests During Input (i.e., Tests 1, 2, 3, 5, 6, 7)		Retention Tests (i.e., Tests 4, 8)	
	Low Density	High Density	Low Density	High Density
	Stimulus	Stimulus	Stimulus	Stimulus
	Condition	Condition	Condition	Condition
0 of 3	$M = 26\%$ $SD = 18\%$	$M = 34\%$ $SD = 16\%$	$M = 46\%$ $SD = 28\%$	$M = 65\%$ $SD = 22\%$
1 of 3	$M = 18\%$ $SD = 8\%$	$M = 15\%$ $SD = 9\%$	$M = 14\%$ $SD = 8\%$	$M = 11\%$ $SD = 11\%$
2 of 3	$M = 56\%$ $SD = 18\%$	$M = 51\%$ $SD = 17\%$	$M = 40\%$ $SD = 24\%$	$M = 25\%$ $SD = 20\%$
2 of 3: Real Word Neighbors	$M = 6\%$ $SD = 6\%$	$M = 53\%$ $SD = 19\%$	$M = 6\%$ $SD = 9\%$	$M = 49\%$ $SD = 33\%$
0-2 of 3: Other Trained Nonword	$M = 6\%$ $SD = 8\%$	$M = 12\%$ $SD = 11\%$	$M = 9\%$ $SD = 14\%$	$M = 20\%$ $SD = 26\%$
0 of 3: No Response	$M = 45\%$ $SD = 33\%$	$M = 42\%$ $SD = 30\%$	$M = 48\%$ $SD = 38\%$	$M = 61\%$ $SD = 35\%$

Table 5. *Descriptive error analysis for input versus retention tests by set.*

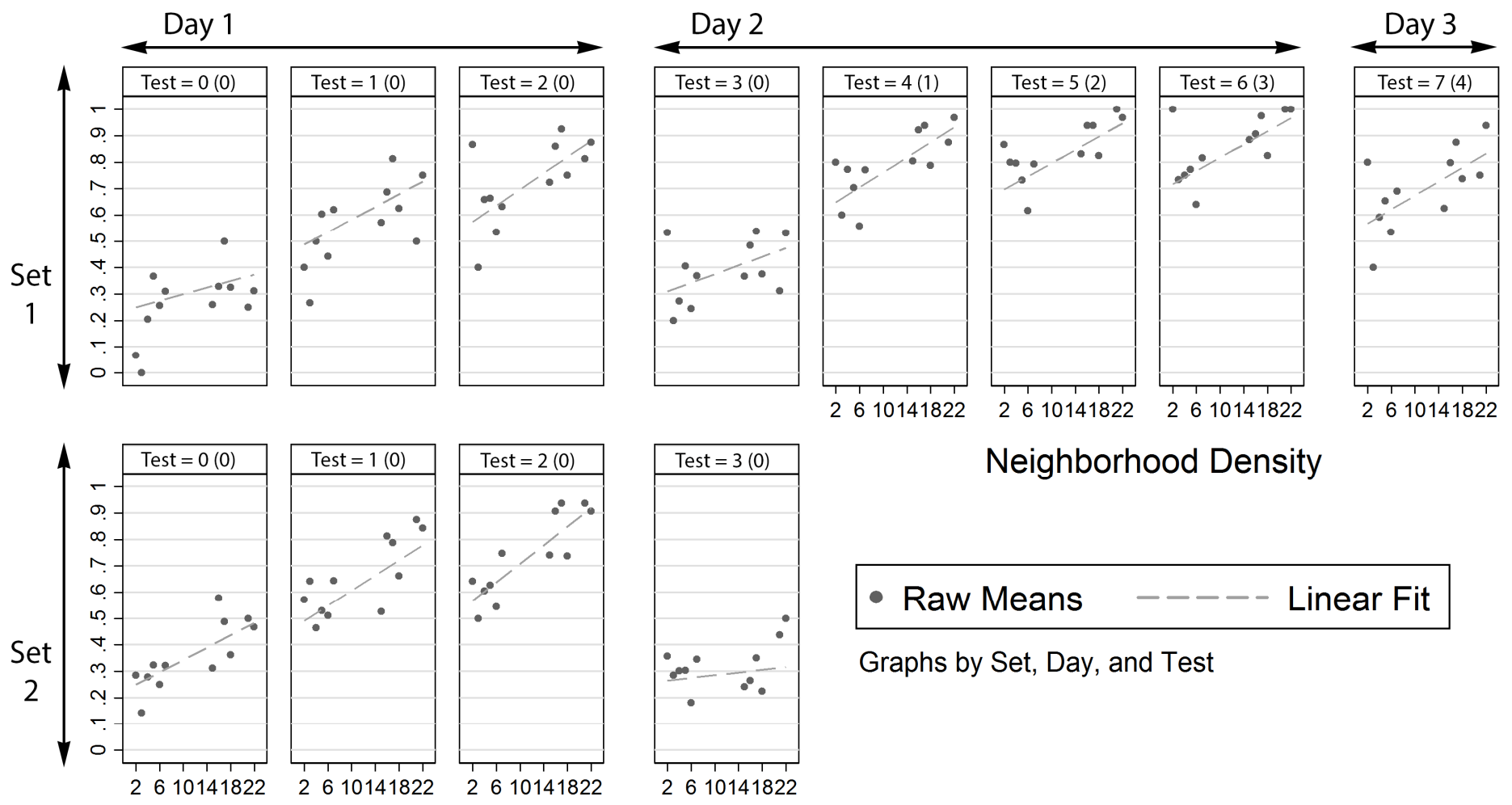
Phonemes Correct	Tests During Input (i.e., Tests 1, 2, 3, 5, 6, 7)		Retention Tests (i.e., Tests 4, 8)	
	1 st Set	2 nd Set	1 st Set	2 nd Set
0 of 3	<i>M</i> = 28%	<i>M</i> = 35%	<i>M</i> = 51%	<i>M</i> = 62%
	<i>SD</i> = 20%	<i>SD</i> = 22%	<i>SD</i> = 30%	<i>SD</i> = 30%
1 of 3	<i>M</i> = 16%	<i>M</i> = 17%	<i>M</i> = 11%	<i>M</i> = 13%
	<i>SD</i> = 11%	<i>SD</i> = 14%	<i>SD</i> = 12%	<i>SD</i> = 13%
2 of 3	<i>M</i> = 56%	<i>M</i> = 48%	<i>M</i> = 37%	<i>M</i> = 25%
	<i>SD</i> = 21%	<i>SD</i> = 25%	<i>SD</i> = 29%	<i>SD</i> = 27%
2 of 3: 1 st Set Neighbor	N/A ¹	<i>M</i> = 7%	N/A ¹	<i>M</i> = 18%
		<i>SD</i> = 13%		<i>SD</i> = 32%
0-2 of 3: Other Trained Nonword	<i>M</i> = 11%	<i>M</i> = 9%	<i>M</i> = 11%	<i>M</i> = 23%
	<i>SD</i> = 13%	<i>SD</i> = 12%	<i>SD</i> = 18%	<i>SD</i> = 33%
0 of 3: No Response	<i>M</i> = 40%	<i>M</i> = 45%	<i>M</i> = 57%	<i>M</i> = 55%
	<i>SD</i> = 32%	<i>SD</i> = 39%	<i>SD</i> = 41%	<i>SD</i> = 40%

¹ 1st set Neighbor was only coded for the 2nd set items because the purpose was to quantify the extent to which the first trained items influenced responding during training of second set items. Note that items within a set were not neighbors of one another. Any confusion within a set is captured by the Other Trained Nonword category.

Figure Caption

1
2
3
4
5 *Figure 1.* Probability of a correct response by the number of neighbors for the 1st set nonwords
6
7 (top) and 2nd set nonwords (bottom). Each panel from left to right corresponds to a given test
8
9 point: test 1, 2, 3, 4 (1-week gap), 5, 6, 7, 8 (1-week gap). The dashed linear fit line in each panel
10
11 illustrates the effect of neighborhood density.
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 1



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Supplemental Materials

Summary Description for Web Portal

This pdf document contains:

- One table providing the nonobject and nonword pairings for each density condition (Table S1).
- One table providing the coding scheme used for test variables in the multi-level model (Table S2).
- One table (Table S3) showing the results of two multi-level models. One model decomposes the effect of Set by neighbor type (see column 2 of Table S2). The other model decomposes the Set x Retention interaction by neighbor type (see column 3 of Table S2).

Table S1: *Nonobject and nonword pairings for low density and high density conditions*

	Low Density	High Density		Low Density	High Density
Nonobjects:	Nonwords:	Nonwords:	Nonobjects:	Nonwords:	Nonwords:
Group 1	Group A	Group A	Group 2	Group B	Group B
46	/ver/	/zin/	88	/ner/	/zen/
37	/zis/	/dʒæd/	53	/zos/	/dʒæf/
26	/fub/	/ðel/	23	/fuk/	/ʃel/
22	/woŋ/	/fuz/	86	/waŋ/	/faiz/
38	/jʌd/	/nʌf/	59	/jʌp/	/nʌd/
67	/ʃædʒ/	/θik/	33	/vædʒ/	/jik/
72	/θeɪ/	/boʃ/	82	/θæf/	/beɪf/
81	/bʊp/	/ʃov/	63	/bʊv/	/tov/
31	/neɪ/	/ʃab/	11	/ʃeɪ/	/fab/
27	/ðɪv/	/wap/	56	/ðɪb/	/wop/
06	/dʒæθ/	/raɪθ/	20	/dʒeθ/	/raɪg/
54	/taz/	/maʊt/	09	/faz/	/haʊt/

Note: All nonobjects come from Kroll & Potter (1984). Nonword to nonobject pairings were counterbalanced across participants. This appendix illustrates one possible pairing. Specifically, Group A nonwords were paired with Group 1 nonobjects. Group B nonwords were paired with Group 2 nonobjects. Group A nonwords were phonological neighbors of Group B nonwords of the same Density condition (Low or High). An individual participant would learn either the low or the high density stimuli based on their randomization to stimulus condition.

Table S2. *Coding of tests for each Set in the multi-level model.*

Set	Session	Test	All Tests (Test 1-8) Slope Code	Retention 1 Intercept Code	Later Tests (Tests 4-8) Slope Code	Retention 2 Intercept Code
1 st Set	Session 1	Test 0	Baseline test that was not included in any models because scores were at floor with 0 variance			
		Test 1	0	0	0	0
		Test 2	1	0	0	0
		Test 3	2	0	0	0
	Session 2	Test 4	3	1	0	0
		Test 5	4	1	1	0
		Test 6	5	1	2	0
		Test 7	6	1	3	0
	Session 3	Test 8	7	1	4	1
	2 nd Set	Session 3	Test 0	Baseline test that was not included in any models because scores were at floor with 0 variance		
Test 1			0	0	0	0
Test 2			1	0	0	0
Test 3			2	0	0	0
Session 4		Test 4	3	1	0	0

Table S2 Note 1: As detailed in the main article, participants were randomly assigned to low or high density stimuli. In the low density condition, both the 1st Set and the 2nd Set consisted of only low density nonwords. For the model, the individual density of each nonword (i.e., a value from 2-7 neighbors) was entered. Likewise, in the high density condition, both the 1st Set and the 2nd Set consisted of only high density nonwords. For the model the individual density of each nonword (i.e., a value from 15-22 neighbors) was entered. In this way, density was modeled as a continuous variable rather than dichotomously coded as low (i.e., 0) and high (i.e., 1).

Two models examined whether the effect of Set (1st vs. 2nd) was influenced by the type of neighbor: CV_neighbor versus VC_neighbor versus C_C neighbor. In the first model, the main effect of set was decomposed into an interaction between set and neighbor type. As can be seen in the second column of Table S2, the coefficients for each of the interactions was similar. Follow-up comparisons showed that none of the types differed significantly from each other, $X^2(2) = 0.80, p = .67$. Thus, the effect of set appeared to be similar for each type of neighbor. In a similar manner, the interaction of set with retention was decomposed into an interaction between set, retention, and neighbor type. As shown in the last column of Table S2, the coefficients for these interactions were similar. Again, follow-up comparisons showed that none of the interactions with type differed significantly from each other, $X^2(2) = 2.55, p = .28$. Once again, the effects observed in the prior analysis of set do not appear to be influenced by the type of neighbor. Overall, the effect of set appeared to be robust across the three different types of neighbors.

Table S3. Models exploring the effect of neighbor type. Estimates are expressed as odds ratios with 95% confidence interval in parentheses.

Variable	Model for Set x Neighbor Types	Model for Set x Retention x Neighbor Types
Gender	1.60 (1.00 – 2.56)	1.60 (1.00 – 2.56)
PPVT-4 Standard Score	1.02 (1.01 – 1.04)**	1.02 (1.01 – 1.04)**
Semantic Set Size	0.82 (0.72 – 0.94)** ¹	0.82 (0.72 – 0.94)** ¹
All Tests (Test 1-8) Slope	3.75 (3.36 - 4.18)***	3.75 (3.36 - 4.18)***
Retention 1 Intercept (1-week gap 1)	0.04 (0.03 – 0.06)*** ¹	0.04 (0.03 – 0.06)*** ¹
Later Tests (Tests 4-8) Slope	0.89 (0.77 – 1.01) ¹	0.89 (0.77 – 1.01) ¹
Retention 2 Intercept (1-week gap 2)	0.03 (0.02 – 0.05)*** ¹	0.03 (0.02 – 0.05)*** ¹
Neighborhood Density	1.03 (0.97 – 1.10)	1.03 (0.97 – 1.10)
Density x All Tests Slope	1.04 (1.03 – 1.05)***	1.04 (1.03 – 1.05)***
Density x Retention 1 Intercept	0.89 (0.85 – 0.92)*** ¹	0.89 (0.85 – 0.92)*** ¹
Density x Retention 2 Intercept	0.89 (0.85 – 0.94)*** ¹	0.89 (0.85 – 0.94)*** ¹
Set	Decomposed by Type	1.23 (1.06 – 1.43)**
Set x Retention 1 Intercept	0.29 (0.22 – 0.38)*** ¹	Decomposed by Type
Set x CV_neighbors	1.33 (1.07 - 1.65)*	
Set x VC_neighbors	1.19 (0.96 – 1.48)	
Set x C_C neighbors	1.19 (0.96 – 1.47)	
Set x CV x Retention 1		0.33 (0.22 – 0.48)*** ¹
Set x VC x Retention 1		0.23 (0.16 – 0.34)*** ¹
Set x C_C x Retention 1		0.32 (0.22 – 0.47)*** ¹
MOR for Participants	1.92 (1.70 – 2.22)	1.92 (1.70 – 2.22)
MOR for Items	3.01 (2.45 – 3.89)	3.01 (2.44 – 3.88)
McKelvey & Zavoina's r^2	0.27	0.27
Somers' D	0.50 (0.46-0.54)	0.50 (0.46-0.54)

* $p < .05$, ** $p < .01$, *** $p < .001$; All models included an overall intercept term that serves as the traditional constant. ¹For OR < 1, a correct response is less likely for a higher than a lower value of the variable.