False memories for morphologically simple versus complex words in English

Anne Pycha
University of Wisconsin, Milwaukee

In tasks such as lexical decision, people respond differently to morphologically complex words compared to morphologically simple ones (e.g. in English, *lies* vs. *rise*). These divergent responses could conceivably arise from differences in activation levels, or alternatively, from the additional steps required to decompose complex words. To investigate this issue, we used the Deese-Roediger-McDermott (DRM) false memory paradigm, which probes activation of lexical representations by measuring the probability of recalling or recognizing a word (such as *lies*) after listening to a list of its phonological neighbors (such as *wise*, *lose, lime*, etc.). Our results showed a significant false memory effect for complex words, which demonstrates that similar-sounding words can activate representations for stem-plus-affix combinations. Our results also showed no significant difference between false memory rates for complex versus simple words, which suggests that complex stem-plus-affix representations activate at levels equivalent to those of simple stem representations. These findings indicate that differences in activation level probably do not lie at the source of divergent responses to complex and simple words, and that decomposition is the more likely origin.

Keywords: false memory, word recognition, complex words, lexical activation, stem, affix

Background and motivation

People respond differently to morphologically complex words, such as *lies*, compared to morphologically simple ones, such as *rise*. In reading tasks, for example, participants' initial eye fixations last longer on complex words than on comparable simple words (Hyönä, Laine, & Niemi, 1995). And in lexical decision tasks, participants take longer to respond to complex stimuli than to comparable simple stimuli. This finding has been reported for English (orthographic stimuli: Taft &

ISSN 1871-1340 / E-ISSN 1871-1375

Forster, 1975; auditory stimuli: Wurm, 2000), Dutch (Jarvella & Wennstedt, 1993), Italian (Caramazza, Laudanna, & Romani, 1988), Finnish (Leinonen et al., 2009), and Swedish (Jarvella & Wennstedt, 1993), where nonsense items that contain real stems or affixes take longer to reject than those that do not. This finding has also been reported for Dutch (Bergman, Hudson, & Eling, 1988) and Finnish (Laine & Koivisto, 1998; Niemi, Laine, & Tuominen, 1994), where real complex words take longer to accept than real simple ones do.

What is the source of these divergent responses? Different activation levels for complex versus simple words present one possible scenario, while the decomposition process triggered by complex words presents another. We consider activation first. In many models of word recognition, each individual lexical representation acts as a computational unit that monitors speech input (Luce & Pisoni, 1998; Marslen-Wilson, 1987; Marslen-Wilson & Welsh, 1978; Morton, 1969). When no speech input occurs that matches the information stored in the representation, it maintains its *resting activation level*, which is relatively low. When matching speech input does occur, the resulting *heightened activation level* correlates with the closeness of the match. For example, the activation level for the representation of the word *rise* /ɹaʲz/ would be resting (i.e., low) when the speech input does not match at all (e.g., [fʊt] or [pil]), somewhat heightened when the input matches it partially (e.g., [saʲz] or [ɹaʲs]), and very heightened when the input matches it completely ([ɹaʲz]). When the heightened activation level for a representation reaches a certain threshold – that is, when enough evidence for a particular word accumulates – word recognition occurs.

Although speech input plays the principal role in modulating activation levels and therefore in triggering word recognition, most models permit additional factors – notably, frequency – to play a role as well, in a manner that is potentially applicable to complex words. Some of these models propose inherent differences, such that low-frequency words have either lower resting activation levels or higher thresholds than high-frequency words (e.g., Marslen-Wilson, 1987; Morton, 1969), while other models propose a bias, such that speech input produces similar activation levels for all representations, but high-frequency words can differentially amplify these levels at a later stage (Goldinger, Luce, & Pisoni, 1989; Luce & Pisoni, 1998). In this paper, we will refer to inherent differences in resting activation level, but nothing crucial rests on this decision. This is because all of these models make a similar basic prediction, namely that the same process (i.e., activation) occurs for all words, but that the process proceeds more slowly for low-frequency words because they require more speech input in order for their activation level to heighten enough to reach threshold (see e.g. Forster, 1976, p. 264; Goldinger et al., 1989, pp. 505–506; Morton, 1969, p. 167). Thus, all of these models correctly predict that in lexical decision tasks, participants take longer to respond to low-frequency stimuli than to high-frequency stimuli (e.g., Broadbent, 1967; Glanzer & Bowles, 1976; Solomon & Postman, 1952; Whaley, 1978).

If we extend the activation metaphor to the question of morphologically complex versus simple words, it is conceivable that the representation for a complex English word such as *lies*, /la^{j}/ plus /z/, has a lower resting activation level compared to the representation for a simple word such as *rise* /ra^{j}z/. Presumably, the source of this lower resting activation would be the morpheme /z/: whereas /la^{j}/ and /ra^{j}z/ are comparable insofar as they are both stems, /z/ plausibly has a lower resting activation because it is an affix. Under such a model, just as more speech input is required for low-frequency words to reach threshold compared to high-frequency words, more speech input would be required for complex words such as *lies* (stem plus affix) to reach threshold, compared to simple words such as *rise* (stem alone). This would correctly predict that in tasks such as lexical decision, complex words will produce longer RTs compared to simple ones.

To our knowledge, no previous study has explicitly tested the activation scenario for complex words, a gap that the current study seeks to fill. Nevertheless, the existing literature on priming does suggest some tentative support for it, because words with shared affixes (*ailment* → *payment*) do not always prime one another, and/or do not prime one another strongly as words with shared stems (*ailment* → *ail*). For example, in a cross-modal priming study with Polish stimuli, words with a common prefix or suffix primed one another, but the facilitative effect was smaller than that for words with a common stem (Reid & Marslen-Wilson, 2000). In a masked priming study with orthographic English stimuli, words with a common prefix (*unable* → *unfair*) primed each other, but this effect did not appear to be morphological, because unrelated words with common initial segments (*uncle* → *unable*) also primed each other (Chateau, Knudsen, & Jared, 2002). And in a priming study with orthographic French stimuli, although words with a common prefix primed one another, words with a common suffix did not (Giraudo & Grainger, 2003). Furthermore, in an unprimed lexical decision task with orthographic Italian stimuli, participants responded significantly faster to words with high-frequency stems compared to low-frequency stems, regardless of whether the words contained high- or low-frequency suffixes; the authors concluded that affix frequency plays no role in word recognition (Burani & Thornton, 2003).

Taken together, these studies suggests that resting activation levels of affix representations may be lower than those for stem representations, where priming effects (*ailment* → *ail*) have been repeatedly reported (see citations in Amenta & Crepaldi, 2012; Diependaele, Grainger, & Sandra, 2012). If that is the case, the presence of an affix representation such as /z/ could lower the overall resting activation level for a complex word like *lies* in comparison to a simple word like *rise*, which contains no affix. Note, however, that at least one priming study contradicts this line of reasoning. In a cross-modal priming study with English stimuli, Marslen-Wilson, Ford, Older & Zhou (1996), found that words with a common prefix (*rearrange* → *rewrite*) or suffix (*darkness* → *toughness*) primed one another, and the effect was comparable to that found in words with a common stem (*absurdity* → *absurd*).

Decomposition presents a different explanation for why people respond differently to complex versus simple words. Decomposition is a process whereby a complex word is segmented into its component morphemes. Upon hearing the speech input [la^{j}z], for example, the listener must segment the input into [la^{j}] and [z], search the lexicon for the stem entry /la^{j}/, and then check to see if this stem is compatible with an affix entry /z/ (following Taft & Forster, 1975). If each of these steps takes additional time, lexical decision RTs for complex words will be longer than those for simple words, which do not have multiple component morphemes. Unlike the activation scenario, which proposes that the same process occurs for all words (albeit at different speeds), the decomposition scenario proposes that fundamentally different processes occur, with more steps required for complex words compared to simple ones.

The literature suggests ample support for the fact that decomposition occurs. Numerous priming studies have demonstrated that exposure to one word (e.g., *pay*) can facilitate subsequent responses to a morphologically related word (e.g., *payment*) (e.g., Crepaldi, Rastle, Coltheart, & Nickels, 2010; Feldman & Larabee, 2001; Fowler, Napps, & Feldman, 1985; Marslen-Wilson, Tyler, Waksler, & Older, 1994; Pastizzo & Feldman, 2002; Rastle, Davis, Marslen-Wilson, & Tyler, 2000; Rastle, Davis, & New, 2004). These morphological priming effects (e.g., *vowed* → *vow*) supersede those observed for words with shared form (*vowel* → *vow*) and shared meaning (*pledge* → *vow*) (Feldman, 2000; Pastizzo & Feldman, 2009), and they remain robust even when the stem undergoes changes due to allophony or allomorphy (Marslen-Wilson, Hare, & Older, 1993; Marslen-Wilson et al., 1994; Marslen-Wilson & Zhou, 1999; Perlak, Feldman, & Jarema, 2008; Scharinger, Reetz, & Lahiri, 2009). Brain imaging studies (for review, see Bozic & Marslen-Wilson, 2010) also support the idea that people break complex words down into morphemes. On the basis of such evidence, almost all models of word recognition incorporate decompositional mechanisms as well as whole-word mechanisms (for reviews, see Amenta & Crepaldi, 2012; Diependaele et al., 2012).

Which scenario, activation or decomposition, is correct? The answer to this question has implications for our understanding of the mental lexicon. The activation scenario would suggest a stem-centric view of the mental lexicon such that, even though all morphemes undergo the same process for recognition, affixes nevertheless play a diminished role compared to stems. Affixes (and by extension, complex words that include affixes) would differ from stems by virtue of their low resting activations, just as low-frequency words differ from high-frequency words in the same manner. In the activation scenario, then, the algorithm for word recognition remains equivalent across situations, and differences between words and morphemes are captured using different representations.

By contrast, the decomposition scenario would suggest a combinatorial view of the lexicon in which stems and affixes play equivalent roles (Marslen-Wilson et al., 1996). Affixes would not differ fundamentally from stems. Instead, complex words would require additional steps compared to simple words. Note that this scenario is roughly analogous to Forster's (1976) serial model of frequency effects, where – in contrast to activation models – the differences between low- and high-frequency words arose from the additional steps required to search an ordered list in which low-frequency words occur at the end. In the decomposition scenario, then, lexical representations remain equivalent across situations, and differences between words and morphemes are captured using different algorithms.

As should be clear, reaction time analyses cannot distinguish between these two alternatives, because either low resting activation or decomposition could conceivably slow participant responses to the same extent. To address this question, then, we need a different experimental paradigm that measures a different outcome variable. In the current study, we used a false memory experiment to probe the activation levels of complex words. In false memory experiments with the Deese-Roediger-McDermott (DRM) paradigm, participants see or hear lists of words that are associates or neighbors of a critical item, and subsequently try to remember those words (seminal papers include Deese, 1959; Roediger & McDermott, 1995; for reviews, see Gallo, 2006, 2010). The key result is that participants often (falsely) remember the critical item, even though it did not occur in the lists. For example, the words *rack, pack, bake, book, bag, bat*, etc., are all phonological neighbors of the critical item *back*, because they differ from it by the substitution of one phoneme. After hearing such a list, participants falsely remembered the unheard word *back* approximately 65 to 70% of the time (Sommers & Lewis, 1999). Several studies have reported similar results, for both serial recall and yes/no recognition tasks, and established the robustness of false phonological memories for morphologically simple words (Amberg, Yamashita, & Wallace, 2004; Ballardini, Yamashita, & Wallace, 2008; Ballou & Sommers, 2008; Garoff-Eaton, Kensinger, & Schacter, 2007; McDermott & Watson, 2001; Schacter, Verfaellie, & Anes, 1997; Wallace, Shaffer, Amberg, & Silvers, 2001; Wallace, Stewart, Shaffer, & Wilson, 1998; Wallace, Stewart, & Malone, 1995; Wallace, Stewart, Sherman, & Mellor, 1995; Watson, Balota, & Roediger, 2003; Westbury, Buchanan, & Brown, 2002). In this study, we extended the DRM paradigm to morphologically complex words. We investigated the probability that people falsely recall and/or recognize a complex word such as *lies*, after listening to lists of neighbors such as *wise* [waʲz], *lose* [luz], *lime* [laʲm], etc. For comparison, we investigated the probability that people falsely recall and/

or recognize a comparable simple word such as *rise*, after listening to *size* [sa^{j}z], *rouse* [ɹa^{w}z], *rice* [ɹa^{j}s], etc.

Following Sommers & Lewis (1999) as well as the work of Wallace and colleagues (e.g., Wallace et al., 1998), we interpret the probability of falsely remembering a critical item as an indication of the activation level of its representation. In most word recognition studies, speech input triggers heightened activation levels in a direct manner (e.g., input [ra^{j}z] for *rise*); in DRM studies, by contrast, speech input does so in an indirect manner. Specifically, the source of heightened activation is the list of heard words: for a critical item such as *rise*, the list will contain spoken phonological neighbors such as *size* [sa^{j}z], *rouse* [ɹa^{w}z], *rice* [ɹa^{j}s], etc. With the speech input from one of these words, such as *size* [sa^{j}z], the fully-matching target representation /sa^{j}z/ heightens its activation level to a large degree, while the partially-matching representations for unheard neighbors, such as those for *rise* /ɹa^{j}z/, *wise* /wa^{j}z/, *seize* /siz/, and so on, also heighten their activation levels, although to a somewhat lesser degree (Luce & Pisoni, 1998; see also Chan & Vitevitch, 2009; Goh, Suárez, Yap, & Tan, 2009; Goldinger et al., 1989; Magnuson, Dixon, Tanenhaus, & Aslin, 2007; Stockall, Stringfellow, & Marantz, 2004; Vitevitch, 2007; Luce, Goldinger, Auer, & Vitevitch, 2000). On a given list, this type of spreading activation occurs repeatedly, and eventually converges on the single word which is a neighbor to all of the words on that list, namely the critical item (*rise* /ɹa^{j}z/). If this converging activation level reaches the threshold for word criterion, it gives the listener the experience of thinking that they heard the critical item, even though they did not (see also Collins & Loftus, 1975; Roediger, Balota, & Watson, 2001; for a different view of false memories, see Kroll, Knight, Metcalfe, Wolf, & Tulving, 1996; Reinitz, 2001).

Thus, a crucial feature of the DRM false memory paradigm is that it probes activation of lexical representations without relying on reaction times. Importantly, inherent differences in resting activation levels, if present, would modulate the likelihood of experiencing a false memory. For example, if a complex word such as *lies* (/la^{j}/ plus /z/) has a low resting activation compared to a simple word such as rise (/ɹa^{j}z/), more speech input should be required to heighten the activation level of *lies* to threshold. In the DRM false memory paradigm, the speech input occurs in the form of heard phonological neighbors, and when activation levels reach threshold, they trigger a false memory for an unheard word. Therefore, if differences in resting activation account for differences between complex and simple words, we would predict significantly different rates of false memories for *lies* compared to *rise*.

Thus, the key questions of our study were (a) whether complex words can undergo heightened activation levels in the DRM paradigm – that is, can phonological neighbors trigger sufficiently high activation levels for stem-plus-affix representations to produce false memories for words like *lies*, and (b) if so, do the representations for complex words exhibit activation levels comparable to those of simple words like *rise*? To our knowledge, this is the first DRM false memory experiment conducted with morphologically complex words. We focused on words like *lies* because they are regularly inflected (in this example, for either plural or present tense), and therefore highly likely to consist of distinct stem and affix representations, rather than whole-word representations (Allen & Badecker, 2002; Fowler et al., 1985; Kempley & Morton, 1982; Stanners, Neiser, Hernon, & Hall, 1979; although see also Baayen, Dijkstra, & Schreuder, 1997). To preview the results, participants produced a robust false memory effect for complex words. And, they did so at rates comparable to those for simple words. This finding suggests that inherent differences in resting activation level probably do not lie at the source of divergent responses to complex and simple words, and that decomposition is the more likely origin.

Method

Stimuli

Target stimuli

As critical items, we selected forty monosyllabic English words. Twenty of the words were morphologically simple, and twenty were morphologically complex. As shown in Table 1, the complex words contained one of three suffix realizations: present tense or plural [z], past tense [d], or past tense [t]. The simple words contained phonologically identical pseudo-suffixes. We selected the simple words such that they did not contain embedded pseudo-stems; for example, removing [z] from *cheese* produces a non-stem [t͡ʃi].

Table 1. Critical items used in the design of the experiment

Condition	Suffix type	Critical items
Simple	Present tense or plural [z]	cheese, ease, phase/faze, rise
	Past tense [d]	aid, bride, fade, glade, trade
	Past tense [t]	fact, fast, haste, jest, list, lust, pest, raft, rest, rust, wrist
Complex	Present tense or plural [z]	blows, grows, lies, ties
	Past tense [d]	cried, died/dyed, laid, lied, paid, played, tried
	Past tense [t]	hacked, laced, lost, pieced, poked, tipped, ticked, tucked, wrecked

We were obligated to exclude critical items with the voiceless [s] realization of present tense/plural (e.g., *caps*) because the English lexicon does not include enough simple words with equivalent characteristics (e.g., *lapse*) to match across conditions. We purposely excluded complex words that exhibited stem changes upon suffixation, but ultimately included one such word (*lost*) in order to balance lexical statistics across conditions. A couple of the critical items have homophones; importantly, the homophones were always of the same morphological type (i.e., *died* and *dyed* are both complex, while *phase* and *faze* are both simple).

Across the two conditions, we balanced the critical items for frequency, familiarity, and phonological neighborhood density, to the extent possible, as shown in Table 2. Two-tailed t-tests showed no significant differences between simple versus complex critical items for any of these characteristics.

Table 2. Lexical statistics (means and standard deviations) for the forty words used as critical items in the DRM false memory paradigm, across complex versus simple conditions

Condition	Frequency	Familiarity	Number of neighbors
Simple	1.19 (0.62)	6.85 (0.29)	19.75 (6.74)
Complex	1.19 (0.68)	6.92 (0.14)	18.85 (9.29)

Note. Log frequency is the base-10 log of the overall corpus frequency (Marian, Bartolotti, Chabal, & Shook, 2012). Familiarity ratings represent judgments on a scale from 1 to 7, from a large sample of American English speakers (Nusbaum, Pisoni, & Davis, 1984). Density refers to the total number of words that differ from the item by the substitution of one phoneme (Marian et al., 2012). Note familiarity ratings are not generally available for complex words, so we report the familiarity of the relevant stems instead.

Note that the syllable structure of the critical items varied. The maximal structure was $C_1C_2VC_3C_4$, although most words had either C_1VC_4 (i.e., simple onset and simple coda) or $C_1VC_3C_4$ (i.e., simple onset and complex coda). Coda structure varied systematically with the voicing of the suffix, such that words with voiced suffixes or pseudo-suffixes all contained a single coda consonant (for [z], *lies* [laʲz] and *phase* [feʲz]; for [d], *played* [pleʲd] and *glade* [gleʲd]), while those with voiceless suffixes or pseudo-suffixes all contained two coda consonants (for [t]: *tipped* [tɪpt], *list* [lɪst]). Onset structure varied on an item-by-item basis, with some words containing no onset (*ease* [iz]), some containing a single onset consonant (*lies* [laʲz], *cheese* [t͡ʃiz]), and others containing two onset consonants (*cried* [kɹaʲd]), *bride* [bɹaʲd]). Ideally, we would have imposed a constant syllable structure across all critical items, but it was not possible to do this while also balancing them for lexical statistics. In keeping with the goals of the current study, however, both the complex and simple conditions contained critical items with one and two onset consonants, and critical items with one and two coda consonants. Thus, syllable structures varied in ways that were comparable across complex versus simple conditions.

As discussed in the introduction, people respond differently to complex versus simple words in traditional recognition tasks, such lexical decision (Bergman et al., 1988; Laine & Koivisto, 1998; Niemi et al., 1994). To verify that this was also the case for our experimental materials, we compiled previously-reported reaction times on lexical decision and naming for our forty critical items (Balota et al., 2007). As Table 3 shows, mean reaction times were greater for complex critical items, compared to simple ones. Two-tailed t-tests indicated that the difference in reaction times was significant for lexical decision ($p < 0.05$) although not for naming ($p = 0.25$).

Table 3. Mean reaction times for the forty words used as critical items, in lexical decision and naming tasks. Data are from the English lexicon project (Balota et al., 2007)

Condition	Lexical decision	Naming
Simple	600.86 (42.94)	595.65 (41.14)
Complex	653.79 (59.54)	612.30 (45.23)

Previous work has shown that concrete critical items (such as *bread*) exhibit lower rates of false recall than abstract ones (such as *sleep*) (Pérez-Mata, Read, & Diges, 2002). To examine whether this lexical characteristic could potentially be relevant for our experimental materials, we compiled previously-reported concreteness ratings for our forty critical items (Brysbaert, Warriner, & Kuperman, 2014), shown in Table 4. A two-tailed t-test indicated that the difference between ratings for simple versus complex critical items was not significant. Thus, we do not expect concreteness to impact our results, although if it were to do so, the effect should be to lower false recall rates in the complex condition, where concreteness ratings are (descriptively) higher.

Table 4. Mean concreteness for the forty words used as critical items, on a scale from 1 (most abstract) to 5 (most concrete) (Brysbaert et al., 2014)

Condition	Concreteness rating
Simple	3.46 (1.01)
Complex	3.66 (0.76)

For each critical item, we constructed a list of nine phonological neighbors that differed by the substitution of a single phoneme. Sample lists are displayed in Table 5, and the complete list of stimuli is in the Appendix.

Table 5. Sample lists of phonological neighbors of the indicated critical item (CI). Each neighbor differed from the indicated critical item by the substitution of a single phoneme, which was located in one of five possible positions

Condition	CI	Edge onset	Medial onset	Vowel	Medial coda	Edge coda
Simple	*phase/ faze*	haze, maze, pays		fees, fizz, foes		faith, fake, fame
	glade	blade, flayed, played, slayed	grade/ grayed	glad, glide, glued		glaze
	list	fist, hissed		laced, last, lost, lust	lift, lint	lisp
Complex	*lies*	light, lime, live		laws, lose		dies, guys/ guise, vies/ vise, wise
	played	blade, flayed	prayed	plead, plowed/ ploughed		place, plague, plane/plain, plate
	tipped	ripped, shipped, whipped, zipped		taped, topped, typed	tint, tilt	

We excluded any phonological neighbors that were also morphologically related to the critical item; for example, we excluded *lied* from the *lies* list, and we excluded *tips* from the *tipped* list. Note that when calculating phonological neighborhoods, the Clearpond website counts r-colored vowels as a single segment; we followed the same convention here, such that e.g. *ford* is a neighbor of *fade*.

Across the two conditions, we balanced the neighbors for frequency, familiarity, and number of phonological neighbors, to the extent possible, as shown in Table 6. Two-tailed t-tests showed no significant differences between simple versus complex critical conditions for any of these characteristics.

Table 6. Lexical statistics (means and standard deviations) for the 360 words used as neighbors, across complex versus simple conditions

Condition	Frequency	Familiarity	Number of neighbors
Simple	1.02 (0.81)	6.71 (0.67)	16.01 (8.14)
Complex	1.02 (0.78)	6.73 (0.63)	16.75 (9.53)

Note. See Table 2 for sources of this information. For neighbors with homophones, the frequency statistic reflects the summed frequency of all homophones (e.g., the frequency of *size* = frequency of *size* + frequency of *sighs*), while the familiarity statistic reflects the familiarity of simple homophones only. The number of neighbors was calculated on a strictly phonological basis (e.g., what are the neighbors of [sa^{j}z]), so the calculation of this statistic was not dependent on the presence or absence of homophones.

Our goal was that each list of nine words should include a roughly equal number of phoneme substitution types. That is, for critical items with C_1VC_4 structure such as *lies* [laʲz], there are three phoneme substitution types, and we aimed to include three neighbors that substituted onset C_1, three that substituted the vowel, and three that substituted the coda C_4. For critical items with $C_1VC_3C_4$ structure such as *tipped* [tɪpt], there are four phoneme substitution types and thus we obviously could not divide the list of nine neighbors so neatly, but we nevertheless aimed to include approximately equal numbers of neighbors that substituted onset C_1, V, coda C_3, and coda C_4. Similarly for critical items with $C_1C_2VC_4$ structure such as *cried* [kɹaʲd], we aimed to include approximately equal numbers of phoneme substitutions that substituted onset C_1, onset C_2, vowel, and coda C_4. Table 7 displays the mean proportion of substitution types included on the lists. A chi-square test (conducted on raw counts of neighbor types, rather than proportions) showed no significant dependence between critical item complexity and phoneme substitution types ($\chi^2 = 2.25$, $p = 0.81$).

Table 7. Mean proportion of phoneme substitution types included on lists of nine words, across conditions

Condition	Onset			Coda	
	C_1	C_2	V	C_3	C_4
Simple	0.35	0.02	0.34	0.18	0.11
Complex	0.40	0.01	0.32	0.17	0.09

Gaps in the English lexicon limited the extent to which we were able to achieve our goal. The proportion of C_2 neighbors was extremely low in both conditions, partly because this neighbor type was not applicable to the many critical items that did not contain complex onsets, and partly because of English syllable phonotactics: complex onsets must generally rise in sonority, so C_2 neighbors could typically substitute only [l, ɹ] in this position (e.g., *prayed* is a neighbor of the critical item *played*). The proportion of C_4 neighbors was also low in both conditions, again primarily because of phonotactics: complex codas must generally fall in sonority, but additionally, complex obstruent codas must agree in voicing. For $C_1VC_3C_4$ critical items such as *tipped* [tɪpt], then, there are no possible neighbors with rising coda sonority (*[tɪpɹ]), and also none with conflicting voice specifications in the coda (*[tɪpd], *[tɪbt]). Of the possibilities that remained, we excluded words like *tips* [tɪps] because they were morphologically related to the critical item, further lowering the number of neighbors with C_4 substitution.

Given that our stimulus limitations were not haphazard but resulted primarily from systematic gaps in the lexicon, we felt reasonably confident that our lists of neighbors could produce the converging association that is typically required for phonological false memories to occur. The fact that our list words were equivalently distributed among the broader neighbor types of onset substitution (collapsing C_1 and C_2), vowel substitution, and coda substitution (collapsing C_3 and C_4) supports this idea. Furthermore, and most importantly, the distribution of neighbor types, while unequal within lists, was nevertheless nearly identical across the complex and simple conditions, so it is unlikely that any difference between these conditions would be due to differences in list composition.

The neighbor lists in both conditions included both complex and simple words. For example, the neighbor list for the critical item *lies* includes *laws* as well as *lose*; similarly, the neighbor list for the critical item *phase* includes *fees* as well as *fizz*. Our goal was to include roughly equivalent numbers of complex and simple neighbors on each list. As shown in Table 8, this effort was only partially successful. A chi-square test (conducted on raw counts of neighbor types, rather than proportions) showed no significant dependence between critical item complexity and neighbor types ($\chi^2 = 5.07$, $p = 0.08$).

Table 8. Mean number of neighbor types (standard deviations) included on each list of nine words, across conditions

CI type	Complex neighbor *laws, fees*	Simple neighbor *lose, fizz*	Homophonous neighbor *guise/guys, grade/grayed*
Complex CI *lies*	3.80 (2.26)	3.95 (1.93)	1.25 (1.07)
Simple CI *phase*	2.90 (1.21)	5.00 (1.41)	1.10 (0.79)

In the complex CI condition, the relatively higher number of complex neighbors was ultimately driven by gaps in the English lexicon – in particular, the generally sparse neighborhoods of many words with past-tense [t]. For the critical item *tipped*, for example, the Clearpond website returns fourteen neighbors: thirteen of these are complex (*ripped*, *shipped*, etc.), and only one (*tilt*) is simple.

Filler stimuli

We selected forty monosyllabic English words as filler critical items, with two goals in mind. On the one hand, we wanted to distract participants from focusing strategically on complex critical items, which obviously stand out by virtue of their contrast with simple critical items. On the other hand, we also wanted to distract participants from focusing strategically on any target critical item, which stand out by virtue of their distinct phonotactic structure. That is, as Table 1 shows, all of our target critical items, whether simple or complex, ended in a sequence of vowel-[z], vowel-[d], or consonant-[t], and this is not a representative sample of the English monosyllabic lexicon. With these factors in mind, we selected fifteen fillers that were complex, irregular past-tense words with C(C)VC shape (e.g., *came*, *stole*). And we selected twenty-five fillers that were simple words with different phonotactic structures, including CVC words with a voiced fricative coda besides [z] (e.g., *dive*), CVC words ending with a voiced stop coda besides [d] (e.g., *rib*), and C(C)VCC words with consonant-voiceless stop coda sequences besides consonant-[t] (e.g., *stark*). Thus, although the fillers were not split evenly between complex and simple conditions in the way that targets were, their selection addressed the two major characteristics (morphological structure and phonotactic structure) that might have otherwise made the targets appear distinct. For each critical item, we selected nine neighbors, including roughly equivalent numbers of neighbors that differed by an onset phoneme, a vowel, or a coda phoneme.

Recording

A speaker recorded each word in a sound-proof booth with a head-mounted microphone. The speaker was a male native speaker of the Midwestern variety of American English, unaware of the purpose of the experiment. He recorded the words in a random order. The audio recording was digitized at a sampling rate of 44.1 kHz, and segmented into individual files using the Praat program (Paul Boersma & David Weenink, 2014).

Procedure

The eighty lists (forty target lists plus forty filler lists) were divided into four sets of twenty (A, B, C, D), each containing five lists from the complex condition, five lists from the simple condition, and ten lists from the control condition. Note that in the master set of eighty lists, certain critical items also occurred as phonological neighbors on other lists. For example, the word *phase* served as a critical item, but it also occurred as a phonological neighbor on the list of words for the critical item *fade*. Importantly, we divided the lists so that on any given set, no critical items also occurred as a neighbor. For example, the *phase* list was assigned to set B, while the *fade* list was assigned to set A. Thus, any participant who was assigned to set A would actually hear the word *phase* during the experiment and his or her responses to this particular word would count as "veridical". By contrast, any participant who was assigned to set B would not hear *phase* during the experiment, but would hear its list of phonological neighbors, so his or her responses to this word would count as "critical item". Each participant was randomly assigned to one of the four sets.

During the experiment, participants were seated in an individual carrel within a quiet laboratory setting, in front of a computer equipped with a mouse, keyboard, and high-quality headphones. Printed instructions on the computer screen guided them through each step. In the study phase, participants listened to twenty lists of nine spoken words. Each word on a list was played individually, followed by 1 second of silence before the onset of the next word. After each list, participants did a free recall task, in which they were given 45 seconds to type as many words as they could remember from the list, in any order. After 45 seconds, they proceeded to the next list. The overall order of the twenty lists, as well as the order of the nine words within each list, was randomized for each participant.

In the test phase, after listening to all twenty lists, participants did a recognition task in which they listened to an individual spoken word, and made a yes/no judgment as to whether they had heard the word previously in the experiment.

There were 96 items in the recognition task, which included forty words that the participant actually heard (two from each of the twenty heard lists), plus fifty-six that the participant had not heard. The unheard words included the twenty critical items from the participant's own set. In addition, the unheard words included thirty-six foils, consisting of twelve critical items from other experimental sets (one from each of twelve unheard lists, which included three complex, three simple, and six filler lists), and twenty-four neighbor words from other sets (two from each of twelve unheard lists, which again included three complex, three simple, and six filler lists). The order of items in the recognition task was randomized for each participant.

Participants

Participants were native speakers of the Midwest variety of American English (n = 86), between the ages of 18 and 30, approximately half female and half male, with no history of problems in speech, language, or hearing. The experiment took approximately forty-five minutes of their time, and in exchange for participating, they received either cash compensation or extra credit points in a linguistics course.

Results

Recall task

The recall task yielded a total of 10,369 responses. Thus, on average, listeners responded with 6.03 items per list (= 10,369 / (20 lists per participant * 86 participants)). Once the fillers were removed, there were 5,037 target responses. Participants sometimes typed the same word twice for one list, resulting in 239 duplicates (129 in the complex condition, and 110 in the simple condition), which were removed. Many of the list items had legitimate alternative spellings, such as *airs* instead of *heirs* or *waste* instead of *waist*, which were counted as accurate responses (112 responses in the complex condition, and 110 in the simple condition). Participants also sometimes typed true mis-spellings, which were included in the data set as long as their orthographic-to-phonetic conversion produced an actual word of English. Thus, for example, *reaked* was accepted for *reeked, tride* was accepted for *tried, jist* was accepted for *gist*, and *kart* was accepted for *cart* (33 such responses in the complex condition, and 56 in the simple condition). There were 20 typos which did not produce a real English word, such as *shve* and *ylkd* (8 in the complex condition, 12 in the simple condition), and these were removed from the data set. After removal of fillers, duplicates, and typos, 4,778 responses were included in the final analysis.

Following the methodology used in previous studies on false recall (Roediger & McDermott, 1995; Sommers & Lewis, 1999) we classified a response as "veridical" if it corresponded to a word that actually occurred on the list (e.g., *light, lime, live, laws, lose, dies, guys, vies, wise*, etc.), "intrusion" if it did not occur on the list (e.g., random intrusions such as *child, coffee, table*, etc.), and "critical item" if it corresponded to the critical item (e.g., *lies*).

For descriptive statistics, we calculated proportions. For veridical proportions, again following previously established methodology, we divided the number of veridical responses per list by nine, which was the number of words that actually occurred on each list. For example, if a participant provided four veridical responses for a list, the proportion of veridical responses would be 0.44 = 4/9. For intrusion proportions, we divided the number of intrusion responses by the total number of responses per list. For example, if a participant provided one intrusion response for a list, plus four veridical responses and one critical item, the proportion of intrusion responses would be 0.17 = 1/(1 intrusion + 4 veridical + 1 critical item). Finally, the critical item proportion was calculated as 0 if the participant did not respond with the critical item, and 1 if they did. Table 9 displays the mean proportion of veridical, intrusion, and critical item responses given across the three conditions.

Table 9. Proportions of response types provided by participants (means, standard deviations) in recall task, across conditions

Condition	Veridical	Intrusion	Critical item
Simple	0.47 (0.19)	0.19 (0.20)	0.24 (0.43)
Complex	0.45 (0.20)	0.21 (0.21)	0.28 (0.45)

For inferential statistics, we conducted two different types of analyses, a repeated-measures ANOVA and a mixed-effects logit model. The advantage of ANOVA is that we can readily compare our results with those from previous DRM false memory studies, which typically employ either ANOVA or t-tests. One disadvantage of ANOVA, however, is that it may not be generally appropriate for the analysis of categorical outcome variables (Jaeger, 2008). Another disadvantage of ANOVA, specific to the DRM recall task, is that it may not be appropriate for situations in which the denominator changes from one trial to the next (as in the calculation of intrusion proportions, where the denominator was the total number of words the participant wrote down during a given recall trial). Logit models can address these problems, although they make it more difficult to compare our results with those from previous studies. As one way to adjudicate between the different strengths and weaknesses of these analyses, we ran them both. Importantly, they yielded the same conclusions.

Repeated-measures ANOVA analysis was conducted on the proportion of responses as the outcome variable, using the predictors response type (Veridical vs. Intrusion vs. Critical Item) and critical item complexity (Simple vs. Complex). Results revealed a main effect of response type ($F(2, 170) = 77.11, p < 0.05$). To understand the nature of this main effect, we performed further analyses on response type. In DRM false memory tasks, veridical responses are typically significantly higher than other response types, because these responses correspond to words that participants actually heard on the lists. However, the crucial comparison occurs between intrusion and critical item responses, because the false memory effect occurs whenever participants recall the critical item significantly more often than (random) intrusions. Consistent with this, our planned comparisons revealed a significant difference between responses to veridical words compared to both other word types ($t = -14.56, p < 0.05$), and between responses to intrusions compared to critical items ($t = 2.35, p < 0.05$). No other results were significant.

A mixed logit model was conducted using the glmer function from the R package lme4. Following Jaeger (2008), we counted each response as a "success", and we counted each possible lack of response as a "failure". For example, if a participant provided four veridical responses for a list, we counted four successes, plus five failures, corresponding to the five words on the list of nine that he or she did not recall. If a participant provided one intrusion response, we counted one success, and the remaining responses (veridical plus critical item) as failures. If a participant provided the critical item as a response, we counted one success, and zero failures. We used predictor variables of response type (Veridical vs. Intrusion vs. Critical item) and condition (Simple vs. Complex), with random intercepts for participants. We used treatment coding. "Simple" served as the baseline for condition, because previous studies on phonological false memories have typically used critical items of this type (e.g., morphologically simple words like *rise*), and we were interested in how "Complex" would deviate from this baseline. "Critical item" served as the baseline for response type, which allowed us to make comparisons between veridical versus critical item responses and, crucially, between intrusion versus critical item responses.

The model revealed two simple effects. Response type exerted a simple effect for both levels of the predictor. The odds of a response increased in the Veridical condition compared to the Critical Item baseline, by a factor of approximately 2.89

($\beta = 1.06$, std. error = 0.12, $z = 8.99$, $p < 0.05$). The odds of a response decreased in the Intrusion condition compared to the Critical Item baseline, by a factor of approximately 0.73 ($\beta = -0.31$, std. error 0.12, $z = -2.47$, $p < 0.05$). These effects replicate previous findings in the literature and indicate that the false memory paradigm produced the intended result: that is, participants were crucially less likely to respond with a random intrusion compared to a critical item. No other results were significant.

To summarize the recall results, both repeated-measures ANOVA and mixed logit models revealed a significant difference between critical item responses compared to random intrusions, yet found no significant difference between the simple versus complex conditions. That is, people were more likely to falsely recall a complex critical item (such as *lies*) than a random intrusion (such as *child, coffee, table*, etc.), and this false memory effect was equivalent to that found for simple critical items (such as *rise*).

Additional analyses for recall

In free recall tasks, participants typically exhibit better performance for words that occur at the beginning of the study list compared to those that occur in the middle ("primacy effect"), and they also exhibit superior performance for words that occur at the end ("recency effect") (Gallo, 2006, pp. 26–28). To investigate whether simple versus complex critical items modulated these effects, we tabulated the total number of veridical responses provided by participants according to the serial position in which they originally occurred at study (each list contained nine items, so the positions are 1 through 9), as seen in Table 10. For both rows, a primacy effect is apparent, because positions 1 and 2 have more veridical responses than later positions. A recency effect is also apparent, because positions 8 and 9 have more veridical responses than earlier positions. A chi-square test showed no significant dependence between critical item complexity and study list position ($\chi^2 = 1.10$, $p = 0.99$); thus, critical item complexity did not significantly modulate these effects.

Table 10. Number of veridical responses provided during free recall test, according to the word's original serial position, 1 through 9, in the study list

Condition	1	2	3	4	5	6	7	8	9
Simple	243	223	184	153	152	164	175	217	310
Complex	233	210	177	147	141	145	170	203	313

Intrusions are important in DRM false memory experiments, primarily because they indicate the baseline probability that participants will (falsely) recall any item that did not occur on the study list; if critical item responses exceed this baseline, we can claim a false memory effect. Yet intrusions also reflect a participant's active efforts to reconstruct the list of words he or she heard at study, and as such, they merit some attention in their own right. In our results, participants provided both morphologically simple intrusion responses (such as *old* for the list associated with critical item *aid*, and *lag* for the list associated with critical item *laid*) and morphologically complex ones (such as *aged* for the *aid* list, and *played* for the *laid* list). We were interested to know whether the simplicity versus complexity of the critical item modulated the simplicity versus complexity of the intrusions, so we tabulated the types of intrusion responses provided in each condition, as shown in Table 11. In both conditions, participants provided more simple responses than complex ones. A chi-square test showed no significant dependence between critical item complexity and complexity of intrusions ($\chi^2 = 3.66$, $p = 0.89$); thus, critical item complexity did not significantly modulate intrusion type.

Table 11. Number of intrusion responses, according to whether they were simple or complex

Condition	Simple intrusion	Complex intrusion
Simple CI list	282	123
Complex CI list	339	124

In their study of false phonological memories, Sommers & Lewis (1999, p. 88) found that when participants falsely recalled a critical item at test, they were approximately equally likely to do so at the beginning, middle, and end of their responses. We were interested to know whether the same effect occurred in our own results, and whether this effect was modulated by critical item complexity, so we tabulated the critical item responses according to the order in which participants provided them at test, as shown in Table 12. Unlike those of Sommers & Lewis (1999), our critical item responses are not evenly distributed, but rather show a tendency to occur early. This difference is probably due to task instructions: Sommers & Lewis instructed their participants to recall the last few items on the study list first, whereas we did not. A chi-square test showed no significant dependence between critical item complexity and order of recall ($\chi^2 = 3.66$, $p = 0.89$); thus, critical item complexity did not significantly modulate response order.

Table 12. Number of critical item responses, according to order in which they were recalled by subjects

Condition	1	2	3	4	5	6	7	8	9
Simple	23	14	27	12	16	4	5	2	1
Complex	24	23	29	16	13	8	6	1	1

Recognition task

In the recognition task, participants responded to three different types of items. "Veridical" items actually occurred on a list that the participant heard (e.g., *light*, *lime, live, laws, lose, dies, guys, vies, wise*, etc.). "Intrusion" items did not occur on any list that the participant heard (foils drawn from one of the three experimental sets that the participant was not assigned to, e.g., *fizz, post, shaft* etc.). "Critical items" were critical items for which the participant heard lists of neighbors (e.g., *lies*).

Following previous work, we calculated proportions of "yes" responses to each type of item. For veridical proportions, we divided the number of "yes" responses by the total number of items of this type that were presented. For example, participants responded to two veridical items from each of the twenty lists they heard; if they responded "yes" to one of these, their veridical proportion for this list was 0.50 (= 1/2). Participants responded to three intrusions from each of twelve lists that they did not hear; if they responded "yes" to one of these, their intrusion proportion was 0.33 (= 1/3). Finally, participants responded to one critical item from each of twenty lists that they heard; if they responded "yes" to it, their critical item proportion was 1 (= 1/1). Table 13 displays the mean proportions of "yes" responses for the three item types across conditions.

Table 13. Proportions of "yes" responses (means, standard deviations) given in recognition task, across conditions

Condition	Veridical	Intrusion	Critical item
Simple	0.69 (0.46)	0.30 (0.46)	0.49 (0.50)
Complex	0.75 (0.43)	0.34 (0.47)	0.53 (0.50)

Although the order of items in the recognition test was randomized for each participant, we nevertheless wanted to ensure that item sequence did not interact with any of our other predictors. Therefore, we coded each item according to whether it occurred in the first, second, or third sequence of the recognition list (items 1 through 32, 33 through 64, and 65 through 96, respectively). A linear mixed-effects model with participant and item as random variables showed no main effect of sequence, and no interaction with other predictors. Subsequent analyses therefore excluded sequence as a predictor.

Repeated-measures ANOVA analysis was conducted on the proportion of 'yes' responses as the outcome variable, using the predictors response type (Veridical vs. Intrusion vs. Critical Item) and critical item complexity (Simple vs. Complex). Results revealed main effects of response type ($F(2, 170) = 162.90$, $p < 0.05$) and of complexity ($F(1,85) = 7.29$, $p < 0.05$), but no interaction between them. Planned comparisons revealed a significant difference between responses to veridical words compared to both other word types ($t = -17.33$, $p < 0.05$), and between responses to intrusions compared to critical items ($t = 9.36$, $p < 0.05$).

As with recall, we also analyzed the recognition results with a mixed logit model. We counted each "yes" response as a success and each "no" response as a failure. We used predictor variables of response type (Veridical vs. Intrusion vs. Critical item) and condition (Simple vs. Complex), with treatment coding using "Critical item" and "Simple" as baselines. We included random intercepts for participants. The model revealed two simple effects. Response type exerted a simple effect for both levels of the predictor. The odds of a response increased in the Veridical condition compared to the Critical Item baseline, by a factor of approximately 2.36 ($\beta = 0.86$, std. error = 0.12, $z = 6.90$, $p < 0.05$). The odds of a response decreased in the Intrusion condition compared to the Critical Item baseline, by a factor of approximately 0.42 ($\beta = -0.86$, std. error = 0.13, $z = -6.74$, $p < 0.05$). These effects replicate previous findings in the literature and indicate that the false memory paradigm produced the intended result: that is, participants were crucially less likely to respond with a random intrusion compared to a critical item. No other results were significant.

To summarize the recognition results, both repeated-measures ANOVA and mixed logit models revealed a significant difference between critical item responses compared to intrusions, yet found no significant difference between the simple versus complex conditions. That is, people were more likely to falsely recognize a complex critical item (such as *lies*) than a random intrusion, and this false memory effect was similar to that found for simple critical items (such as *rise*).

Discussion

Our results demonstrate a robust false memory effect for morphologically complex words such as *lies*, as indicated by the fact that, in both recall and recognition tasks, participants were significantly more likely to provide critical item responses than intrusion responses. Such an effect crucially depends upon sufficiently heightened activation levels for affix representations like /z/, in addition to stem representations like /la^j/. Thus, one new finding of the current study is that, after hearing a list of phonological neighbors, people can activate more than one lexical representation (i.e., stem-plus-affix) in order to give rise to a false memory for a complex word. Furthermore, false memory rates for complex critical items such as *lies* did not differ significantly from those for simple critical items such as *rise*. With the caveat that we are interpreting a null result, this finding suggests that, given equivalent levels of speech input, the representations for complex and simple words produce equivalently heightened levels of activation, and therefore people's divergent responses to complex versus simple words probably originate elsewhere, presumably in the extra steps required for decomposition. In order to reach such a conclusion, however, we would need to consider several alternative interpretations of our findings, as well as several avenues for further research.

Similarities in false remembering of simple versus complex words

Based on previous findings, we have assumed that complex words such as *lies* generally undergo decomposition. It remains possible, however, that participants in the false memory task actually activated whole-word representations for these words. If so, there would be no reason to posit separate, lower resting activations for affix representations compared to stem ones, and therefore no reason to suppose lower overall activation for representations like /la^{j}/ plus /z/ compared to /ra^{j}z/. This could account for the similarity in our results for complex versus simple words.

As noted in the introduction, we purposely selected regularly-inflected words as critical items because evidence shows that listeners are highly likely to decompose such words during listening and reading (Allen & Badecker, 2002; Fowler et al., 1985; Kempley & Morton, 1982; Stanners et al., 1979). Yet the nature of the false memory task may nevertheless foster whole-word activation, even if more traditional word recognition tasks show evidence for decomposition. This is because participants heard lists of words that were, by definition, phonologically similar to one another, which could encourage them to attend to the sounds of words, rather than to their meanings or constituent morphemes. To determine the plausibility of this scenario, future research could use the DRM false memory paradigm with semantic associates. These experiments are similar to the current study, except that participants see or hear a list of words (such as *thread, pin, eye, sewing, sharp, point*, etc.) whose meaning is associated with the critical item (such as *needle*) (Deese, 1959; Roediger & McDermott, 1995; and citations in Gallo, 2006, 2010). Previous work strongly suggests that semantic false memories operate via a mechanism that is distinct from that of false phonological memories (Ballardini et al., 2008; Ballou & Sommers, 2008; Garoff-Eaton et al., 2007; McDermott & Watson, 2001; Watson et al., 2003). For our purposes, potential differences between simple and complex words might become more apparent in a paradigm that encourages listeners to attend to meanings (rather than sounds) and, by extension, to morphemes.

The nature of the list words employed in the current study may have also fostered a whole-word approach. As Table 8 showed, while complex words were included on every list, the majority of list words were either simple (*lose, fizz*) or homophonous between simple and complex interpretations (*guise/guys*). The disproportionate number of simple words, as well as the exclusion of morphological relatives (*lied* did not occur on the *lies* list), may have also discouraged activation of decomposed representations. To help determine if this was the case, future research could manipulate the proportion of simple versus complex words on lists.

The role of monitoring

The similarity in false memory rates for complex versus simple words could also be due to a confound between activation levels on the one hand, and monitoring processes on the other. The DRM false memory paradigm raises the activation level of representations indirectly via similar-sounding words. During a process called monitoring, participants reflect upon that activation in order to decide whether it corresponds to a real event or not, effectively asking themselves: Did I really hear *lies*? Did I really hear *rise?* (Roediger & McDermott, 1995; and citations in Gallo, 2006, pp. 97–108). Strictly speaking, of course, the correct answer to this question is always no, but when the monitoring process fails, participants believe the answer to be yes. Thus, any increase in monitoring failures will give rise to an increase in false memory responses. Several previous studies have demonstrated that activation and monitoring can operate independently of one another (e.g., Cortese, Khanna, White, Veljkovic, & Drumm, 2008; Dodson & Schacter, 2001; Hicks & Marsh, 1999), so it is logically possible that our results could reflect (a) lower resting activation for complex words than for simple ones, combined with (b) more failures in the monitoring of complex words compared to simple ones, which conspired to produce (c) roughly equivalent rates of false memories for complex and simple words. As we discuss below, however, this possibility seems unlikely.

In the framework developed by Marcia Johnson and colleagues (Johnson, 2006; Johnson, Hashtroudi, & Lindsay, 1993; Johnson & Raye, 1981), people use several heuristics to distinguish between real versus false memories. Real memories typically contain more surface detail, and more information about surrounding context, than false memories do (Norman & Schacter, 1997). In addition, real memories typically lack a cognitive record of how the event arose, while false memories possess such a record. Johnson proposes that monitoring fails precisely whenever a false memory contains increased amounts of surface detail, increased contextual information, or decreased cognitive records – in other words, whenever false memories most closely resemble real ones.

Following Johnson's proposals, the different representational possibilities for words like *rise* versus *lies* suggest that, if anything, we should expect more failures in monitoring for simple words than for complex ones. For example, the representation for the word-final consonant in *rise* is uncontroversially /z/, which contains only somewhat less surface detail than a spoken token of [z]. But the representation for the word-final consonant in *lies* alternates according to its phonological environment (e.g., *lies* [laʲz] vs. *lights* [laʲts]) and is therefore plausibly /S/, namely an alveolar fricative underspecified for voicing. (A similar logic applies to word-final consonants for critical items such as *cried* [kɹaʲd] vs. *laced* [leʲst] where the word-final consonant is plausibly /T/, namely an alveolar stop underspecified for voicing). By definition, /S/ contains significantly less surface detail than a spoken token of [z]. The upshot is that false memories for *rise* closely resemble real memories for *rise* (that is, /z/ resembles [z]), while false memories for *lies* do not (/S/ does not closely resemble [z]). On this basis, we might expect monitoring to fail more often for simple words compared to complex words. As another example, because inflectional morphemes must fit grammatically into their host sentences, complex words arguably require people to think more about surrounding context than simple words do. Our DRM false memory experiment presented words in isolation, devoid of sentence contexts. Again, the upshot is that false memories for *rise* resemble real memories for *rise*, because the lack of context presents no particular impediment to believing that the word is a real memory, whereas false memories for *lies* do not necessarily resemble real memories for lies, because real instances of *lies* would occur in a sentence context. Therefore, we might expect monitoring to fail more often for simple words like *rise*, compared to complex words like *lies*. These statements are speculative and would require further research to substantiate. Nevertheless, they both suggest that, if there was any difference in monitoring failures across conditions, such failures would have been more likely for simple words compared to complex ones – ruling out a scenario in which our results are due to a confound between activation and monitoring, which would require a difference in the opposite direction.

Alternatives to converging activation

The similarity in results for complex versus simple words could also arise if memory conjunction errors, rather than converging neighborhood activation, are responsible for the phenomenon of false memories. Memory conjunction errors occur when people claim to remember a stimulus that they did not perceive, but that is constructed entirely from parts of other stimuli that they did perceive (for a review, see Reinitz, 2001). For example, participants in several studies have claimed to remember words like *toothache*, after having been exposed to *toothpaste* and *heartache* (Reinitz & Demb, 1994; Underwood, Kapelak, & Malmi, 1976). This finding suggests that, rather than (or in addition to) storing words like *toothache* as a monolithic whole-word representation, people encoded each compound element separately. Similarly, participants in another study have claimed to remember words like *barley*, after having been exposed to *barter* and *valley* (Kroll et al., 1996). This finding suggests that, rather than (or in addition to) storing *barter* and *valley* as whole-word representations, people encoded each component syllable separately.

Results from at least two DRM false memory studies suggest that participants might also encode each phoneme component separately, at least under certain experimental conditions (Ballardini et al., 2008; McDermott & Watson, 2001). Both studies used printed stimuli, and exposed participants to phonological neighbors for durations as brief as 20 milliseconds. Participants did not have a conscious recollection of seeing the words, and veridical recall for list words such as *deep,* *weep*, etc. was near zero. Despite this, false recall for critical items such as *sleep* was robust, with an average probability of about 50%. Converging neighborhood activation does not offer a convincing explanation for these findings: given the low rates of veridical recall, it seems unlikely that words such as *deep* were truly activated, and hence unlikely that exposure to *deep* activated the critical item *sleep* as a word. Memory conjunction, on the other hand, offers a plausible account: during the very brief exposures, participants may have encoded individual phonemes such as /s/, /l/, /i/, and /p/, and conjoined them.

In the current study, our results showed that people claimed to remember the complex critical item *lies*, after having been exposed to *wise* [wa^{j}z], *lose* [luz], *lime* [la^{j}m], etc. Under a memory conjunction account, this occurred because people separately encoded each component phoneme on the word list (e.g., /w/, /a^j/, /z/, /l/, /u/, /z/, /l/, /a^j/, /m/, etc.), and the conjunction of these phonemes is /l/, /a^j/, /z/. False memories for simple critical items like *rise* would arise from the same mechanism, that is, a conjunction of the phonemes /r/, /a^j/, /z/. Importantly, conjunction takes place at a level – namely, the phonemic level – where the different morphological structure of *lies* versus *rise* plays no role, so this account predicts no difference in false memory rates for complex versus simple words, consistent with our results.

Additional research would obviously be required to evaluate the plausibility of memory conjunction as an explanation for our results, but in the meanwhile, we note that the phenomenon of phoneme conjunction, if it exists, seems to be a product of task-specific conditions. The studies of Ballardini et al. (2008) and McDermott & Watson (McDermott & Watson, 2001) presented visual words for durations so brief that participants had no conscious recollection of them, and rates of veridical recall were accordingly near zero. Under these conditions, the presence of false memories cannot be accounted for by spreading activation from neighboring words, and necessitates an appeal to individual phonemes. By contrast, the current study presented auditory words at normal speaking rates, and rates of veridical recall averaged 0.45 to 0.47 (Table 9). Under these conditions, the presence of false memories very plausibly arises from spreading activation from neighboring words.

Conclusion: A combinatorial lexicon

We began this study with the observation that people respond differently to morphologically complex versus simple words, typically taking longer to respond in lexical decision tasks when the stimulus contains morphological constituents compared to when it does not. We outlined two scenarios that could plausibly account for these findings, an activation scenario in which complex words have lower resting activation levels than simple ones, and a decomposition scenario in which complex words require extra processing steps compared to simple ones. We tested the activation scenario in a DRM false memory task, which probes input-triggered activation levels of lexical representations by measuring the probability of recalling or recognizing a word after listening to a list of its phonological neighbors. Our results showed a robust false memory effect for complex words, which demonstrates that similar-sounding words can activate representations for stem and affix morphemes at the same time. Our results also showed no significant difference between false memory rates for complex and simple words, which suggests that complex stem-plus-affix representations heighten their activation levels in a manner equivalent to those of simple stem representations.

How can we reconcile our findings with those of the priming literature? Recall that words with shared affixes (*ail-ment* → *pay-ment*) do not always prime one another in lexical decision tasks, and/or do not prime one another as strongly as words with shared stems (Chateau et al., 2002; Giraudo & Grainger, 2003; Reid & Marslen-Wilson, 2000). It was on this basis that we had speculated that affix representations might have lower resting activation levels than stem representations. Note, however, that priming results reflect the combined effects of the prime word plus the target word. For example, if *ail-ment* is the prime and *pay-ment* is the target, the reaction time to *pay-ment* reflects the interaction between these two words, and is not a simple measure of activation of the suffix *-ment*. This point is made clear by results from cross-modal priming with English stimuli, which have shown that suffixed words prime their stems (*punishment* primes *punish*), but do not prime related suffixed forms (*confession* does not prime *confessor*) (Feldman & Larabee, 2001; Marslen-Wilson et al., 1994; Marslen-Wilson & Zhou, 1999; but see Pastizzo & Feldman, 2002). In pairs such as *confession* → *confessor*, the slow reaction time probably reflects an interaction whereby suffixed words inhibit lexical activation of other suffixed words with the same stem (Marslen-Wilson et al., 1994). Thus, we need not necessarily interpret a slow RT to *confessor* to mean that suffixed stems in general exhibit low activation. By this same logic, in pairs such as *ail-ment* → *pay-ment*, we need not necessarily interpret a slow RT to *pay-ment* to mean that affixes in general exhibit low activation. In fact, different interactions between different types of prime-target pairs may be responsible for the somewhat idiosyncratic results reported across these studies.

When we consider the literature on unprimed lexical decision, which factors out potential interactions between prime and target, a different picture emerges. In one study, for example, participants made judgments for four types of orthographic Italian nonsense words: fully morphological words that contained a real stem plus a real affix (such as *cantevi*, which contains *cant-* 'sing' plus *-evi* '2sg.Past'), partially morphological words that contained either a real stem plus a nonsense affix (*cant-* 'sing' plus *-ovi*) or a nonsense stem plus real affix (*canz-* plus *-evi* '2sg.past'), and fully nonsensical words that contained no morphemes (*canzoni*) (Caramazza et al., 1988). Results showed that participants took significantly longer to reject completely morphological words, compared to completely nonsensical words. Partially morphological words produced reaction times that were between these two extremes, i.e., significantly faster than *cant-evi* ('sing'-'2sg.Past') but significantly slower than *canzoni*. Notably, however, there were no differences between the results for *cant-ovi* ('sing'-NONSENSE) and *canz-evi* (NONSENSE-'2sg.Past'). That is, any morpheme embedded within a nonsense word, whether it was a stem or an affix, exhibited an equivalent effect. Subsequent work has shown that English exhibits a similar pattern of results (Wurm, 2000), as do Swedish and (in a somewhat modified paradigm) Dutch (Jarvella & Wennstedt, 1993). Under these conditions, then, results have shown that input-triggered activation levels of affix representations appear to be equivalent to those for stem representations.

Thus, the findings from unprimed lexical decision studies, which represent a relatively direct measure of activation levels for representations, are consistent with our own findings. Just as the reaction time results for nonce stimuli depend crucially upon sufficiently heightened activation for affix representations like *-evi* ('2sg.Past') as well as for stem representations like *cant-* ('sing'), the false memory effect that we found for complex words like *lies* depends crucially upon sufficiently heightened activation for affix representations like /z/ as well as stem representations like /la^j/. In other words, these results suggest that stems and affixes play equivalent roles in the lexicon; as Marslen-Wilson & colleagues expressed it: "[t]his is a strongly combinatorial view of lexical representation and processing, and it assigns a crucial role not just to stem morphemes but also to the *affixes*…" (1996, p. 224).

Under this view, people's divergent responses to complex versus simple words do not originate in any fundamental difference between their representational characteristics, but presumably arise instead from differences in the word recognition algorithm – that is, from the extra steps required to decompose a complex word into its constituent morphemes, as originally outlined by Taft & Forster (1975). In this sense, complex words have a completely different status in the lexicon than other types of words with slow reaction times, such as low frequency words, which do indeed seem to exhibit lower resting activation levels (or an equivalent bias) compared to their more frequent counterparts (Goldinger et al., 1989; Luce & Pisoni, 1998). One benefit of the current study is that it demonstrates how, paradoxically enough, we can shed light on reaction time results by using an experimental paradigm with non-timed responses.

Acknowledgements

I thank Dylan Pearson, Ellen Abolt, and Amara Sankhagowit for their research assistance, and two anonymous reviewers for useful comments that helped me to improve the manuscript. This research was partially supported by the UWM Research Growth Initiative at the University of Wisconsin, Milwaukee.

References

Allen, M., & Badecker, W. (2002). Inflectional regularity: Probing the nature of lexical representation in a cross-modal priming task. *Journal of Memory and Language*, 46(4), 705–722. doi:10.1006/jmla.2001.2831

Amberg, M., Yamashita, J., & Wallace, W. (2004). Implicit word activation during pre-recognition processing influences correct recognition and estimates of presentation frequency. *Memory*, 12(2), 129–139. doi:10.1080/09658210244000298

Amenta, S., & Crepaldi, D. (2012). Morphological processing as we know it: an analytical review of morphological effects in visual word identification. *Frontiers in Psychology*. doi:10.3389/fpsyg.2012.00232

Baayen, R. H., Dijkstra, T., & Schreuder, R. (1997). Singulars and plurals in Dutch: Evidence for a parallel dual-route model. *Journal of Memory and Language*, 37(1), 94–117. doi:10.1006/jmla.1997.2509

Ballardini, N., Yamashita, J. A., & Wallace, W. P. (2008). Presentation duration and false recall for semantic and phonological associates. *Consciousness and Cognition*, 17(1), 64–71. doi:10.1016/j.concog.2007.01.008

Ballou, M. R., & Sommers, M. S. (2008). Similar phenomena, different mechanisms: Semantic and phonological false memories are produced by independent mechanisms. *Memory & Cognition*, 36(8), 1450–1459. doi:10.3758/MC.36.8.1450

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

Bergman, M. W., Hudson, P. T., & Eling, P. A. (1988). How simple complex words can be: Morphological processing and word representations. *The Quarterly Journal of Experimental Psychology*, 40(1), 41–72. doi:10.1080/14640748808402282

Bozic, M., & Marslen-Wilson, W. (2010). Neurocognitive Contexts for Morphological Complexity: Dissociating Inflection and Derivation. *Language and Linguistics Compass*, 4(11), 1063–1073. doi:10.1111/j.1749-818X.2010.00254.x

Broadbent, D. E. (1967). Word-frequency effect and response bias. *Psychological Review*, 74(1), 1–15. doi:10.1037/h0024206

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, 46(3), 904–911. doi:10.3758/s13428-013-0403-5

Burani, C., & Thornton, A. M. (2003). The interplay of root, suffix and whole-word frequency in processing derived words. *Trends in Linguistics Studies and Monographs*, 151, 157–208.

Caramazza, A., Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition*, 28(3), 297–332. doi:10.1016/0010-0277(88)90017-0

Chan, K. Y., & Vitevitch, M. S. (2009). The influence of the phonological neighborhood clustering coefficient on spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 35(6), 1934.

Chateau, D., Knudsen, E. V., & Jared, D. (2002). Masked priming of prefixes and the influence of spelling – meaning consistency. *Brain and Language*, 81(1), 587–600. doi:10.1006/brln.2001.2549

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82(6), 407–428. doi:10.1037/0033-295X.82.6.407

Cortese, M. J., Khanna, M. M., White, K. K., Veljkovic, I., & Drumm, G. (2008). The activation and monitoring of memories produced by words and pseudohomophones. *Journal of Memory and Language*, 58(2), 393–414. doi:10.1016/j.jml.2007.11.003

Crepaldi, D., Rastle, K., Coltheart, M., & Nickels, L. (2010). "Fell" primes "fall", but does "bell" prime "ball"? Masked priming with irregularly-inflected primes. *Journal of Memory and Language*, 63(1), 83–99. doi:10.1016/j.jml.2010.03.002

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, 58(1), 17–22. doi:10.1037/h0046671

Diependaele, K., Grainger, J., & Sandra, D. (2012). Derivational morphology and skilled reading. *The Cambridge Handbook of Psycholinguistics*, 311–332. doi:10.1017/CBO9781139029377.021

Dodson, C. S., & Schacter, D. L. (2001). "If I had said it I would have remembered it: Reducing false memories with a distinctiveness heuristic. *Psychonomic Bulletin & Review*, 8(1), 155–161. doi:10.3758/BF03196152

Feldman, L. B. (2000). Are morphological effects distinguishable from the effects of shared meaning and shared form? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(6), 1431.

Feldman, L. B., & Larabee, J. (2001). Morphological facilitation following prefixed but not suffixed primes: Lexical architecture or modality-specific processes? *Journal of Experimental Psychology: Human Perception and Performance*, 27(3), 680–692.

Forster, K. I. (1976). Accessing the mental lexicon. In R. J. Wales & E. Walker (Eds.), *New approaches to language mechanisms* (Vol. 30, pp. 256–287). New York: North-Holland.

Fowler, C. A., Napps, S. E., & Feldman, L. (1985). Relations among regular and irregular morphologically related words in the lexicon as revealed by repetition priming. *Memory & Cognition*, 13(3), 241–255. doi:10.3758/BF03197687

Gallo, D. A. (2006). *Associative illusions of memory: False memory research in DRM and related tasks*. Psychology Press.

Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & Cognition*, 38(7), 833–848. doi:10.3758/MC.38.7.833

Garoff-Eaton, R. J., Kensinger, E. A., & Schacter, D. L. (2007). The neural correlates of conceptual and perceptual false recognition. *Learning & Memory*, 14(10), 684–692. doi:10.1101/lm.695707

Giraudo, H., & Grainger, J. (2003). On the role of derivational affixes in recognizing complex words: Evidence from masked priming. In R. H. Baayen & R. Schreuder (Eds.), *Morphological Structure in Language Processing* (Vol. 151, pp. 209–232). New York: Mouton de Gruyter. doi:10.1515/9783110910186.209

Glanzer, M., & Bowles, N. (1976). An analysis the word-frequency effect in recognition memory. *Journal of Experimental Psychology-Human Learning and Memory*, 2(1), 21–31. doi:10.1037/0278-7393.2.1.21

Goh, W. D., Suárez, L., Yap, M. J., & Tan, S. H. (2009). Distributional analyses in auditory lexical decision: Neighborhood density and word-frequency effects. *Psychonomic Bulletin & Review*, 16(5), 882–887. doi:10.3758/PBR.16.5.882

Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language*, 28(5), 501–518. doi:10.1016/0749-596X(89)90009-0

Hicks, J. L., & Marsh, R. L. (1999). Attempts to reduce the incidence of false recall with source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(5), 1195–1209.

Hyönä, J., Laine, M., & Niemi, J. (1995). Effects of a word's morphological complexity on readers' eye fixation patterns. *Studies in Visual Information Processing*, 6, 445–452. doi:10.1016/S0926-907X(05)80037-6

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59(4), 434–446. doi:10.1016/j.jml.2007.11.007

Jarvella, R. J., & Wennstedt, O. (1993). Recognition of partial regularity in words and sentences. *Scandinavian Journal of Psychology*, 34(1), 76–85. doi:10.1111/j.1467-9450.1993.tb01102.x

Johnson, M. K. (2006). Memory and reality. *American Psychologist*, 61(8), 760–776. doi:10.1037/0003-066X.61.8.760

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, 114(1), 3. doi:10.1037/0033-2909.114.1.3

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review*, 88(1), 67. doi:10.1037/0033-295X.88.1.67

Kempley, S. T., & Morton, J. (1982). The effects of priming with regularly and irregularly related words in auditory word recognition. *British Journal of Psychology*, 73(4), 441–454. doi:10.1111/j.2044-8295.1982.tb01826.x

Kroll, N. E., Knight, R. T., Metcalfe, J., Wolf, E. S., & Tulving, E. (1996). Cohesion failure as a source of memory illusions. *Journal of Memory and Language*, 35(2), 176–196. doi:10.1006/jmla.1996.0010

Laine, M., & Koivisto, M. (1998). Lexical access to inflected words as measured by lateralized visual lexical decision. *Psychological Research*, 61(3), 220–229. doi:10.1007/s004260050027

Leinonen, A., Grönholm-Nyman, P., Järvenpää, M., Söderholm, C., Lappi, O., Laine, M., & Krause, C. M. (2009). Neurocognitive processing of auditorily and visually presented inflected words and pseudowords: evidence from a morphologically rich language. *Brain Research*, 1275, 54–66. doi:10.1016/j.brainres.2009.03.057

Luce, P. A., Goldinger, S. D., Auer, E. T., & Vitevitch, M. S. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception & Psychophysics*, 62(3), 615–625. doi:10.3758/BF03212113

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, 19(1), 1–36. doi:10.1097/00003446-199802000-00001

Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science*, 31(1), 133–156. doi:10.1080/03640210709336987

Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities. *PLoS ONE*, 7(8), e43230. doi:10.1371/journal.pone.0043230

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, 25(1), 71–102. doi:10.1016/0010-0277(87)90005-9

Marslen-Wilson, W. D., Ford, M., Older, L., & Zhou, X. (1996). The combinatorial lexicon: Priming derivational affixes. In *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society: July 12–15, 1996, University of California, San Diego* (Vol. 18, pp. 223–227). Psychology Press.

Marslen-Wilson, W. D., Hare, M., & Older, L. (1993). Inflectional morphology and phonological regularity in the English mental lexicon. In *Proceedings of the 15th annual meeting of the Cognitive Science Society* (pp. 693–698).

Marslen-Wilson, W. D., Tyler, L. K., Waksler, R., & Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological Review*, 101(1), 3–33. doi:10.1037/0033-295X.101.1.3

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, 10(1), 29–63. doi:10.1016/0010-0285(78)90018-X

Marslen-Wilson, W. D., & Zhou, X. (1999). Abstractness, allomorphy, and lexical architecture. *Language and Cognitive Processes*, 14(4), 321–352. doi:10.1080/016909699386257

McDermott, K. B., & Watson, J. M. (2001). The rise and fall of false recall: The impact of presentation duration. *Journal of Memory and Language*, 45(1), 160–176. doi:10.1006/jmla.2000.2771

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, 76(2), 165–178. doi:10.1037/h0027366

Niemi, J., Laine, M., & Tuominen, J. (1994). Cognitive morphology in Finnish: Foundations of a new model. *Language and Cognitive Processes*, 9(3), 423–446. doi:10.1080/01690969408402126

Norman, K. A., & Schacter, D. L. (1997). False recognition in younger and older adults: Exploring the characteristics of illusory memories. *Memory & Cognition*, 25(6), 838–848. doi:10.3758/BF03211328

Nusbaum, H. C., Pisoni, D. B., & Davis, C. K. (1984). Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words. *Research on Speech Perception Progress Report*, 10(10), 357–376.

Pastizzo, M. J., & Feldman, L. B. (2002). Does prime modality influence morphological processing? *Brain and Language*, 81(1), 28–41. doi:10.1006/brln.2001.2504

Pastizzo, M. J., & Feldman, L. B. (2009). Multiple dimensions of relatedness among words: Conjoint effects of form and meaning in word recognition. *The Mental Lexicon*, 4(1), 1–25. doi:10.1075/ml.4.1.01pas

Paul Boersma, & David Weenink. (2014). Praat: doing phonetics by computer. (Version 5.4.04). Retrieved from www.praat.org

Pérez-Mata, M. N., Read, J. D., & Diges, M. (2002). Effects of divided attention and word concreteness on correct recall and false memory reports. *Memory*, 10(3), 161–177. doi:10.1080/09658210143000308

Perlak, D., Feldman, L. B., & Jarema, G. (2008). Defining regularity: Does degree of phonological and orthographic similarity among Polish relatives influence morphological processing? *The Mental Lexicon*, 3(2), 239–258. doi:10.1075/ml.3.2.04per

Rastle, K., Davis, M. H., Marslen-Wilson, W. D., & Tyler, L. K. (2000). Morphological and semantic effects in visual word recognition: A time-course study. *Language and Cognitive Processes*, 15(4–5), 507–537. doi:10.1080/01690960050119689

Rastle, K., Davis, M. H., & New, B. (2004). The broth in my brother's brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin & Review*, 11(6), 1090–1098. doi:10.3758/BF03196742

Reid, A., & Marslen-Wilson, W. D. (2000). Organising principles in lexical representation: Evidence from Polish. In *Proceedings of the 22nd Annual Conference of the Cognitive Science Society* (pp. 387–392).

Reinitz, M. T. (2001). Illusions of Memory. *Comments on Theoretical Biology*, 6, 411–430.
Reinitz, M. T., & Demb, J. B. (1994). Implicit and explicit memory for compound words. *Memory & Cognition*, 22(6), 687–694. doi:10.3758/BF03209253
Roediger, H. L., Balota, D. A., & Watson, J. M. (2001). Spreading activation and arousal of false memories. *The Nature of Remembering: Essays in Honor of Robert G. Crowder*, 95–115.
Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(4), 803–814.
Schacter, D. L., Verfaellie, M., & Anes, M. D. (1997). Illusory memories in amnesic patients: conceptual and perceptual false recognition. *Neuropsychology*, 11(3), 331. doi:10.1037/0894-4105.11.3.331
Scharinger, M., Reetz, H., & Lahiri, A. (2009). Levels of regularity in inflected word form processing. *The Mental Lexicon*, 4(1), 77–114. doi:10.1075/ml.4.1.04sch
Solomon, R. L., & Postman, L. (1952). Frequency of usage as a determinant of recognition thresholds for words. *Journal of Experimental Psychology*, 43(3), 195–201. doi:10.1037/h0054636
Sommers, M. S., & Lewis, B. P. (1999). Who really lives next door: Creating false memories with phonological neighbors. *Journal of Memory and Language*, 40(1), 83–108. doi:10.1006/jmla.1998.2614
Stanners, R. F., Neiser, J. J., Hernon, W. P., & Hall, R. (1979). Memory representation for morphologically related words. *Journal of Verbal Learning and Verbal Behavior*, 18(4), 399–412. doi:10.1016/S0022-5371(79)90219-6
Stockall, L., Stringfellow, A., & Marantz, A. (2004). The precise time course of lexical activation: MEG measurements of the effects of frequency, probability, and density in lexical decision. *Brain and Language*, 90(1), 88–94. doi:10.1016/S0093-934X(03)00422-X
Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior*, 14(6), 638–647. doi:10.1016/S0022-5371(75)80051-X
Underwood, B. J., Kapelak, S. M., & Malmi, R. A. (1976). Integration of discrete verbal units in recognition memory. *Journal of Experimental Psychology: Human Learning and Memory*, 2(3), 293.
Vitevitch, M. S. (2007). The spread of the phonological neighborhood influences spoken word recognition. *Memory & Cognition*, 35(1), 166–175. doi:10.3758/BF03195952
Wallace, W. P., Shaffer, T. R., Amberg, M. D., & Silvers, V. L. (2001). Divided attention and prerecognition processing of spoken words and nonwords. *Memory & Cognition*, 29(8), 1102–1110. doi:10.3758/BF03206379
Wallace, W. P., Stewart, M. T., & Malone, C. P. (1995). Recognition memory errors produced by implicit activation of word candidates during the processing of spoken words. *Journal of Memory and Language*, 34(4), 417–439. doi:10.1006/jmla.1995.1019
Wallace, W. P., Stewart, M. T., Shaffer, T. R., & Wilson, J. A. (1998). Are false recognitions influenced by prerecognition processing? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(2), 299.
Wallace, W. P., Stewart, M. T., Sherman, H. L., & Mellor, M. D. (1995). False positives in recognition memory produced by cohort activation. *Cognition*, 55(1), 85–113. doi:10.1016/0010-0277(94)00646-3
Watson, J. M., Balota, D. A., & Roediger, H. L. (2003). Creating false memories with hybrid lists of semantic and phonological associates: Over-additive false memories produced by converging associative networks. *Journal of Memory and Language*, 49(1), 95–118. doi:10.1016/S0749-596X(03)00019-6

Westbury, C., Buchanan, L., & Brown, N. R. (2002). Sounds of the neighborhood: False memories and the structure of the phonological lexicon. *Journal of Memory and Language*, 46(3), 622–651. doi:10.1006/jmla.2001.2821

Whaley, C. P. (1978). Word-nonword classification time. *Journal of Verbal Learning and Verbal Behavior*, 17(2), 143–154. doi:10.1016/S0022-5371(78)90110-X

Wurm, L. H. (2000). Auditory processing of polymorphemic pseudowords. *Journal of Memory and Language*, 42(2), 255–271. doi:10.1006/jmla.1999.2678

Appendix

Experimental stimuli. Study words are in plain text. Critical items are in italics.

Set A. Simple critical items

aid
ace
ache
add (ad)
age
aim
aired
ale (ail)
eyed (I'd, ide)
odd

list
fist
hissed
laced
last
lift
lint
lisp
lost
lust

trade
braid (brayed)
frayed
trace
trail
train
trait
trays
tread
trod

fade
bade
faith
fed
feign
food
ford
jade
made (maid)
phase (faze)

rust
dust
gust
must (mussed)
roast
roost
roughed
roust
runt
rushed

Set A. Complex critical items

cried
bride
creed
crime

lied
died (dyed)
hide
lad

poked
choked
joked
parked

crowd
crud
crude (crewed)
dried
fried
pride (pried)

led (lead)
lice
life
lime
loud
ride

perked
poached
pocked
post
soaked
yolked (yoked)

hacked
backed
fact
hashed
hatched
hiked
hocked
hooked
tact (tacked)
whacked (wacked)

paid
laid
pad
page
pail (pale)
pain
paired
pod
raid
wade (weighed)

Set B. Simple critical items

haste
baste (based)
cased
chased (chaste)
heist
hissed
hoist
host
taste
waist (waste)

raft
daft
laughed
racked (wracked)
rant
reffed
rift (riffed)
roughed
shaft
wrapped (rapped, rapt)

rise
lies
rhyme
rice
right (write, rite)
ripe
roars
rouse
size (sighs)
thighs

phase (faze)
faith
fake
fame
fees
fizz
foes
haze
maze
pays

rest
chest
nest
raced
rent
retched
rust
west
wrecked
zest

Set B. Complex critical items

grows
crows
froze
glows
graze (grays, greys)
groan
gross
grove
prose (pros)
throws (throes)

pieced
beast
feast
least (leased)
passed (past)
paste (paced)
peaked (peeked)
peeped
post
yeast

tipped
ripped
shipped
taped
tilt
tint
topped
typed
whipped
zipped

lost
bossed
cost
laced
last
lest
list
loft
lust
tossed

ties
buys
guys (guise)
pies
tease (teas, tees)
tight
time
toys
type
vise (vies)

Set C. Simple critical items

cheese
bees
chairs
cheap
cheek
chief
choose
chose
keys
knees

glade
blade
flayed
glad
glaze
glide
glued
grade (grayed)
played
slayed

lust
bust (bussed)
cussed
dust
gust
last
lost
lucked
must (mussed)
rust

fact
faked
fast
forked
hacked
lacked

jest (gest)
gent
gist
guessed (guest)
joust
just

packed (pact)
sacked
tacked (tact)
whacked (wacked)

messed
pest
vest
zest

Set C. Complex critical items

laced
chased (chaste)
faced
haste
least (leased)
lest
list
paced (paste)
raced
taste

ticked
kicked
licked
nicked
talked (tocked)
tilt
tint
tipped
torqued
tucked

wrecked
checked
pecked
racked (wracked)
raked
reeked
rent
rest
retched
sect

lies
dies (dyes)
guys (guise)
laws
light
lime
live
lose
vise (vies)
wise

tried
bride
dried
fried
trade
tread
tribe
tripe
trite
trod

Set D. Simple critical items

bride
bide
braid (brayed)
bread
breed
brewed (brood)
bribe
brine
dried
pride (pried)

fast
cast
faced
fact
fest (fessed)
first
fist
forced
mast (massed)
vast

wrist
cyst
gist
kissed
missed (mist)
raced
rest
rift (riffed)
ripped
risk

ease	*pest*
as	best
each	passed (past)
eat	pecked
eel	pelt
eve	pierced
eyes	post
heirs (airs)	pressed
ooze	test
owes	vest

Set D. Complex critical items

blows	*laid*	*tucked*
blares	lad	chucked
blaze	lake	ducked (duct)
bloat	lame	mucked
bloke	lane	sucked
blues	lard	tact (tacked)
close (clothes)	load	talked
flows	made (maid)	torqued
glows	paid	touched
slows	wade (weighed)	toughed (tuft)
died (dyed)	*played*	
dad	blade	
deed	flayed	
dice	place	
did	plague	
dime	plane (plain)	
dive	plate	
guide	plead	
lied	plowed (ploughed)	
tied (tide)	prayed	

Corresponding address

Anne Pycha
Department of Linguistics
University of Wisconsin-Milwaukee
P.O. Box 413
Milwaukee, Wisconsin 53201-0413
USA
pycha@uwm.edu