The effect of orthography on the recognition of pronunciation variants.

Jeanne Charoy

Department of Psychology, Stony Brook University

Arthur G. Samuel

Department of Psychology, Stony Brook University

Basque Center on Cognition Brain and Language, Donostia-San Sebastian 20009 Spain

IKERBASQUE, Basque Foundation for Science, Bilbao 48011 Spain

Correspondence concerning this article should be addressed to Jeanne Charoy, Department of Psychology, Stony Brook University, Stony Brook NY 11794

Contact: jeanne.charoy@stonybrook.edu

Experiment 1 and 2 were presented in May 2018 at the International Meeting of the Psychonomic Society in Amsterdam, the Netherlands.

Data and analyses can be found at https://github.com/jeanne-charoy/Charoy-Samuel2018_data

Abstract

In conversational speech, it is very common for words' segments to be reduced or deleted. However, previous research has consistently shown that during spoken word recognition, listeners prefer words' canonical pronunciation over their reduced pronunciations (e.g., *pretty* pronounced [prɪti] vs. [prɪɾi]), even when the latter are far more frequent. This surprising effect violates most current accounts of spoken word recognition. The current study tests the possibility that words' orthography may be one factor driving the advantage for canonical pronunciations during spoken word recognition. Participants learned new words presented in their reduced pronunciation (e.g. [trɒti]), paired with one of three spelling possibilities: (1) no accompanying spelling, (2) a spelling consistent with the reduced pronunciation (a reduced spelling, e.g., "*troddy*"), or (3) a spelling consistent with the canonical pronunciation (a canonical spelling, e.g., "*trotty*"). When listeners were presented with the new words' canonical forms for the first time, they erroneously accepted them at a higher rate if the words had been learned with a canonical spelling. These results remained robust after a delay period of 48 hours, and after additional learning trials. Our findings suggest that orthography plays an important role in the recognition of spoken words and that it is a significant factor driving the canonical pronunciation advantage observed previously.

Keywords: orthography; phonological variation; spoken word recognition

Introduction

"*Hear them down in Soho square, Dropping "h's" everywhere. Speaking English anyway they like!*" (Lerner & Loewe, 1956). This is how Henry Higgins, the phonetics professor in My Fair Lady, complains about his peers' English pronunciation. But whether one speaks in Higgins' Standard English, Eliza Doolittle's Cockney accent or General American, when it comes to conversational speech, words are not often produced in their full, canonical form. In fact, a corpus analysis of conversational American English (the Variation in Conversation corpus, Pitt et al., 2003) revealed that over 60% of words deviated in some way from their canonical form (Johnson, 2004). For example, Johnson (2004) found that in one instance, the word *apparently,* which has nine segments in its canonical form ([əp^{h}ɝrəntli]), was produced with only four segments ([p^{h}ɝɾɪ]). Even more dramatically, Johnson found that the two-syllable word *because* can be realized with zero syllables ([k^{h}z]) in some contexts (e.g., the phrase *because if,* [k^{h}zɪf]).

Conversational speech is characterized by such reductions of vowels and consonants. For example, a common word-internal variation in northern dialects of American English is the flapping of the consonant /t/ when it is positioned between a stressed and an unstressed vowel (e.g., in words like *butter*, *pretty*, or *letter*). Another common case is the consonant cluster /nt/ which, in this same intervocalic position, can be reduced to a nasalized flap (e.g., in words like *center*, *gentle,* or *winter*). The speech signal is notoriously variable, and these changes only increase its lack of invariance. How does the perceptual system map two acoustically different signals, such as [sɜntər] and [sɜɾ̃ər], to the same underlying word, in this case *center*? In other words, how are phonological variants processed and represented?

Extensive research has been devoted to this question, and several theories have been proposed that make different predictions about how listeners process phonological variants. Ranbom and Connine (2007) evaluated five mechanisms for recognition of phonological variants in spoken language including underspecification (e.g., Lahiri & Marslen-Wilson, 1991), phonological inference (e.g., Gaskell & Marslen-Wilson, 1996), feature parsing (Gow, 2003), tolerance (e.g., Connine, Titone, Deelman, & Blasko, 1997) and their own frequency-based account. Ranbom and Connine compared each accounts' predictions to listeners' behavior when processing the nasal flapping of intervocalic /nt/ clusters (e.g., in words like *gentle* or *center*). They focused specifically on the hypothesis that representations in lexical memory are influenced by how frequently listeners are exposed to each phonological variant. Indeed, while one phoneme can be pronounced in several different ways (e.g., as a flap, as a glottal stop, etc.), there tend to be "dominant" variants that occur far more often than others (e.g., Patterson & Connine, 2001). This frequency-based hypothesis is grounded in research that has demonstrated a tight coupling between exposure frequency and pronunciation variant processing (e.g., Pitt, Dilley, & Tat, 2011; Sumner & Samuel, 2009).

Ranbom and Connine (2007) first used a corpus analysis of conversational speech to measure the frequency of two regular variants of the /nt/ cluster -- the canonical, fully pronounced form and the reduced nasal flap. This analysis revealed the reduced form to be far more frequent in conversational speech (about 80% of the observed /nt/ words were produced with a nasal flap) than the canonical form (about 15% were fully pronounced). Next, Ranbom and Connine used a lexical decision task to observe how a word's phonological form influences listeners' speech processing. By their frequency account of speech processing, listeners should recognize the more frequent phonological variants (in this case the nasal flap) more quickly and more accurately. But contrary to this hypothesis, Ranbom and Connine found that listeners recognized words' canonical form more quickly and accurately than their more frequent reduced form.

These results also contradict underspecification and phonological inference accounts, as both predict no differences in processing between canonical and reduced forms; but the pattern could be accommodated with a tolerance-based account. Ranbom and Connine used a second lexical decision task to compare performance on trials with regular reductions (e.g., a nasal flap in *gentle*) versus artificial ones (e.g., *whisper* reduced to *whisser*). They found that artificially reduced nonwords like *whisser* failed to activate their real word lexical representations, whereas regular reductions (e.g., *gennle*) did. This result is not consistent with a tolerance account which predicts that all reductions should be equal. Overall, Ranbom and Connine's findings demonstrate that the theories of variant processing that have been proposed cannot easily account for listeners' preferences for canonical pronunciations in lexical decision tasks.

The advantage for the canonical form has also been observed in priming paradigms. Sumner and Samuel (2005) investigated the effect of regular phonological variation on word recognition over both the short and the long term, using semantic and form priming tasks. Their study focused on three regular variants of the final /t/ consonant in American English: the canonical full [t] form, the glottal stop form [ʔ], and the coarticulated and glottal stop form [ʔt]. As with the intervocalic /nt/ cluster, the canonical form is relatively infrequent in American English. Sumner and Samuel tested whether the three variants were equally efficient in priming semantically related words in the short term (500ms ISI and 100ms ISI). Under these conditions, all three regular phonological variations primed semantically related words efficiently. In other words, immediate word recognition was not hindered by phonologically legal variation in pronunciation. As in Ranbom and Connine's (2007) study, these results are not consistent with exemplar or experienced-based accounts of variant processing, which predict stronger lexical activation for more frequent forms.

Sumner and Samuel (2005) also investigated how variants are accessed and stored in the long term. Using tasks that tap both implicit memory and explicit memory, they found a facilitation effect over the long term only for the canonical form; the other two legal variants were not as effective in activating the stored representation. The authors argued that the canonical form best matches the lexical memory representation, despite its being heard relatively infrequently. In a subsequent study, Sumner and Samuel (2009) reported further evidence that only the canonical form is fully represented in memory, with one exception: Participants who frequently produced a reduced form due to their dialect seemed to have both the canonical form and that reduced form represented in memory.

The studies by Ranbom and Connine (2007) and by Sumner and Samuel (2005, 2009) demonstrate an advantage for the canonical form of consonants during word recognition, even when this form is infrequent. In fact, the "canonical advantage" has been reported in a good number of other studies, across a range of different tasks (e.g. Ernestus, Baayen, & Schreuder, 2002; Janse, Nooteboom, & Quené, 2007; Pitt, 2009a; Pitt, 2009b; Racine & Grosjean, 2005; Tucker & Warner, 2007; Van der Ven, Tucker & Ernestus, 2011). For example, Pitt (2009a) looked at intervocalic nasal flapping in American English (e.g., words like *center*) using a battery of tasks including the Ganong paradigm, phonemic restoration and a learning paradigm. Across these different tasks, he reported a robust advantage for the canonical form during lexical activation.

However, recent studies have questioned whether there truly is a benefit for the canonical pronunciation (Bürki, Viebahn, Racine, Mabut & Spinelli, 2018; McGowan & Sumner, 2014; Sumner, 2013; Sumner, Kurumada, Gafter & Casillas, 2013; Viebahn & Luce, 2018). Specifically, Bürki and colleagues suggest that the canonical advantage can be explained with exposure frequency, while Sumner and colleagues argue that phonetic mismatches in the stimuli can explain why such an advantage was observed. Both findings may help further define the boundaries of the canonical advantage, but as of yet, they do not fully explain past results such as the ones listed above. We will address both accounts more thoroughly in the General Discussion.

While many studies report a canonical advantage, few address why it might occur. Why should listeners prefer the canonical form when it is not the most frequently produced variant? After finding results that were inconsistent with all existent theories, Ranbom and Connine (2007) speculated that this advantage could be due to the canonical pronunciation's correspondence with the orthographic forms of words. Indeed, spelling is usually a closer match to the canonical pronunciation than the reduced pronunciation (although orthography can be changed to match the reduced pronunciation, such as the informal *wanna* for *want to*). Because all the adults who participated in previous studies were literate, they had orthographic representations for words in memory. If listeners bring to bear information from both the written and spoken domains to deal with the variability in speech, the canonical form would be supported by the orthographic representations. Note that this would reduce the relative influence of spoken production frequency of phonological variants.

There is quite a bit of research investigating the relationship between phonological and orthographic representations (in domains unrelated to the canonical form advantage) that has demonstrated a link between the two. Early studies reported that knowledge about words' spelling seemed to influence speech perception (e.g., Jakimik, Cole, & Rudnicky, 1985; Morais, Cary, Alegria, & Bettelson, 1979; Seidenberg & Tanenhaus, 1979; Taft & Hambly, 1985; Taft, 2006). Seidenberg and Tanenhaus (1979) found that rhyming words that had matching spellings were recognized faster than non-matching ones (*pie* is a better prime for *tie* than for *rye*), and Morais et al. (1979) showed that illiterate adults had difficulty making meta-linguistic judgments about phonemes. Taft (2006) asked speakers of Australian English to make pseudohomophone judgments about pairs such as word "corn" and non-word "cawn". Because Australian English is non-rhotic, these words have the same pronunciation and should be treated as homophones. Yet, such pairs were judged to be homophonous 54% of the time on average, compared to 79% for control non-r pairs (e.g. word "soak" and non-word "soke"). Taft (2006) concluded that this is evidence that orthography influences phonological representations.

Another robust set of evidence comes from studies of orthographic inconsistencies. Orthographic inconsistencies occur when one phoneme can be mapped onto different graphemes (e.g., in French, [o] can be written as *o*, *au*, *eau*, *ot* ...), or when one grapheme can be mapped onto different phonemes (e.g., in English, [ough] is pronounced differently in *although*, *cough*, and *through*). Listeners are slower to make lexical decisions about words that are orthographically inconsistent (e.g., *leaf* which could potentially have been spelled *leef* or *lief*) compared to orthographically consistent words (e.g., French: Perre & Ziegler, 2008; Ziegler & Ferrand, 1998; English: Stone, Vanhoy, & VanOrden, 1997; Chinese: Chen, Chao, Chang, Hsu, & Lee, 2016; Portuguese: Ventura, Morais, Pattamadilok, & Kolinsky, 2004).

While these findings clearly demonstrate that spelling influences spoken word recognition, they do not directly address its effect on pronunciation variants. A recent study comparing readers and pre-readers brings more direct support for the hypothesis that the canonical advantage might be driven by words' written form (Racine, Bürki, & Spinelli, 2014; also see Coridun, Ernestus & Bosch, 2015, for a related test using second language learners). Racine and colleagues compared how literate and pre-literate children processed the schwa-deletion variant in French. This variant is optional in some words (e.g., *renard*) and mandatory in others (e.g., *bracelet*, although there is variation across dialects of French). However, in both cases, the vowel schwa is always encoded in the spelling with the letter 'e'. Consistent with previous research that found an advantage for the canonical pronunciation during spoken word recognition, French listeners have been found to recognize the canonical "with schwa" pronunciation faster, regardless of how frequently it is produced (e.g., Racine & Grosjean, 2005). To assess whether this is due to the canonical pronunciation's consistency with the words' orthographic forms, Racine et al. (2014) tested French readers (aged 9 to 10 years old) and pre-readers (aged 5 to 6 years old) with a word monitoring task.

The listeners heard the canonical and reduced pronunciations of both obligatory schwa-deletion words (e.g., *bracelet* pronounced [bʁasəlɛ]) and optional schwa-deletion words (e.g., *renard* pronounced [ʁənaʁ] or [ʁnaʁ]). For words that are frequently produced in both forms, both readers and pre-readers were faster at recognizing the canonical form. However, for words that are never produced with the schwa, pre-readers were slower to recognize the canonical form compared to the reduced form, whereas readers recognized both forms equally fast. This difference in processing speed between readers and pre-readers suggests that readers were influenced by orthographic information during spoken word recognition. Because the "with schwa" canonical pronunciation was consistent with the words' spelling, readers accepted that pronunciation readily even though it does not occur in speech. Although comparing literate and illiterate populations is a clever test of orthographic influences on spoken word recognition, there may be differences (e.g., in terms of cognitive development or linguistic experience) between 6-year-olds and 9-year-olds that could be producing the difference in performance. In addition, the schwa case is similar to the canonical advantage effect but differs in an important way: The canonical pronunciation is also the more common pronunciation.

<u>The Current Study</u>

As reviewed above, there are hints in the literature that support Ranbom and Connine's (2007) speculation that the canonical advantage in speech processing may be driven by orthography. The current study provides a direct empirical test of this hypothesis, based on adult native English listeners' recognition of English words. We focused on two relatively well-studied American English cases: intervocalic occurrences of the consonant /t/ and the consonant cluster /nt/. When /t/ or /nt/ occur between a stressed and an unstressed vowel they are usually reduced to a flap or a nasal-flap, respectively. These two cases provide an appropriate testbed because previous studies have shown that 1) their most frequently produced form is *not* the canonical form, and 2) listeners prefer the canonical form during spoken word recognition. A notable difference between the two cases is that flapping is almost mandatory for intervocalic /t/ in American English but nasal-flapping of the /nt/ cluster is optional. The nasal-flap is the more frequent phonological variant, but the /nt/ canonical form is acceptable.

We used a new-word learning paradigm to test the effect of spelling on phonological variant processing. Using new words reduces interference from listeners' previous linguistic experience, although the new words were designed to follow English phonotactics. An American English speaker recorded a set of new words that contained either an intervocalic /t/ consonant or /nt/ cluster. Each new word was recorded once in its canonical form (i.e., with a fully pronounced [t] or [nt]) and once in its reduced form (i.e., with [ɾ] or [ɾ̃]). The learning paradigm we used to teach listeners these new words was similar to the one used successfully by Leach and Samuel (2007) and by Samuel and Larraza (2015). Listeners learned to associate new words with pictures of unusual objects, as if the new words were the names of the objects. In our study, an orthographic manipulation was included so that listeners would also learn the spelling of some of the new words (Samuel & Larraza included a related manipulation).

Listeners were randomly assigned to one of three groups that differed in the spellings they were taught. One group (hereafter, Canonical_None) learned spellings consistent with the canonical pronunciation for half of the new words and received no spellings for the other half. The second group (Reduced_None) learned spellings consistent with the reduced pronunciation for half of the new words and received the other half with no spelling. The final group (Canonical_Reduced) learned half of the words with spellings consistent with the canonical pronunciation and the other half with spellings consistent with the reduced pronunciation.

Aside from these differences in the orthographic information that was provided, the conditions for the three groups were identical. Listeners learned the same new words and learned to pair them with the same pictures. Critically, all the new words were *always* presented with the reduced pronunciation (i.e., a t-flap or a nasal-flap) during the learning phase. Because the only difference among words was the orthographic information that listeners were given, these three groups allowed us to observe whether orthography influences how listeners recognize different phonological variants of the same words.

After the associations of the new words to their respective pictures were well-learned, participants were given a picture-name matching task. Each of the pictures was presented together with either the (reduced) word form it had repeatedly been associated with during training, or with a (never-before heard) canonical version of the associated word. For example, for the item that was always heard as "senno" during training, the corresponding picture would be presented either with auditory "senno", or with auditory "sento". The critical question is how likely the participants are to accept auditory "sento" as the name of the picture, as a function of whether that item had been learned with no orthography, with the spelling *senno*, or with the spelling *sento* (keeping in mind that the trained auditory form was always "senno"). If listeners only rely on the phonological information provided during training, they should not accept the canonical pronunciation at a high rate because they have never heard it before. However, if orthography influences spoken word recognition, listeners should accept the canonical pronunciation at a higher rate for those words learned with the canonical spelling (e.g., *sento*). High acceptance rates for the canonical form of words cannot be due to having both reduced and canonical phonological representations in memory unless the canonical representations arose from orthography.

Experiment 1: Does orthography drive the canonical pronunciation advantage in spoken word recognition?

The main goal of Experiment 1 was to test whether orthography influences the recognition of spoken words shortly after listeners had learned the words. The experiment included three phases: a presentation phase, a learning phase, and a picture matching task. Participants were randomly assigned to one of the three groups described above: Canonical_None, Reduced_None or Canonical_Reduced. They learned new words while seeing pictures of unfamiliar objects, so that each word would be treated as the name of one object. The objects were chosen such that it was very unlikely participants would have existing labels for them. Immediately after the learning phase, participants were tested with a picture-name matching task. The experiment was run in a single session that lasted approximately one hour.

Method

Participants. 77 undergraduate students from Stony Brook University participated in Experiment 1. They were all 18 years old or older and were native speakers of American English.

Stimuli. The stimuli consisted of sets of new and existing words, and pictures of objects (some of which were presented with written words).

New words. A set of 32 fictitious English words was created. Half of the words contained an intervocalic /t/ (e.g., "trotty" pronounced [trɒti]), the other half an intervocalic /nt/ (e.g., "sento" pronounced [sɛnto]). In all cases, the vowel preceding the critical phoneme was stressed and the one following was unstressed. 16 words were bisyllabic and 16 were trisyllabic (see Appendix A for a full list). A male native speaker of American English recorded each item twice, once in its canonical form (i.e., with [t] or [nt]) and once in its reduced form (i.e., with [ɾ] or [ɾ̃]. On average, canonical productions were longer (*M* = 726ms, *SD* = 176ms) than reduced productions (*M* = 700ms, *SD* = 175ms). A linear mixed effect model with number of syllables (two or three) and pronunciation type (canonical or reduced) as fixed factors and word pair as a random factor (e.g. [sɛnto] and [sɛɾ̃o] were treated as a pair) revealed that, as expected, reduced pronunciations were significantly shorter (β = -25.59, *SE* = 6.91, *df* = 31, *t* = -3.70, $p < 0.001$).

Control words. Control words and nonwords were recorded by the same speaker. 32 tokens were real English words and 32 were nonword versions of the words created by changing a single phonetic feature. The real words were chosen such that they contained one phoneme among the pairs /b/-/p/, /s/-/ʃ/ or /m/-/n/. To create the nonword counterpart for each word, we replaced one phoneme with its paired partner (e.g., "priest" and "briest"; "snail" and "smail"*;* "basketball" and "bashketball"). These small differences in the nonwords were intended to be subtle so that listeners would need to pay attention to notice them. The idea was to make it clear to subjects that even single-feature differences should not be accepted in the picture-name matching task (see below).

Pictures. For each control word we selected a color picture of the object it referred to. For example, a picture of a basketball was paired with the word *basketball*. The same picture was paired with the nonword *bashketball* during the final picture-matching task. The new words were paired with pictures of unusual objects that listeners were not likely to have seen, or to have labels for (see Leach & Samuel, 2007, for some examples). There were 64 pictures, 32 of familiar and 32 of unusual objects.

Spellings. For each new word a written form was created. 32 were canonical spellings that were consistent with the canonical pronunciation (e.g., *trotty* for [trɒti]), and 32 were reduced spellings that were consistent with the reduced pronunciation (e.g., *troddy* for [trɒɾi]). Appendix A presents the full list of items.

Apparatus and procedure. Groups of up to three participants were tested in sound-attenuated booths. Participants sat in front of a computer monitor on which visual stimuli were displayed. They wore high quality headphones and used a labelled button-pad to respond. Figure 1 illustrates the sequence of tasks that were run.

Presentation phase. Participants first saw each of the 64 pictures of objects in a random order. The objects were accompanied by the auditory presentation of their “correct” names: familiar objects were presented with their correctly pronounced names (e.g., *basketball*, rather than *bashketball*) and unfamiliar objects with the reduced-form pronunciation of their names (e.g., *troddy*). This phase lasted approximately two minutes.

Learning task. During this phase participants learned to pair each new word with its associated unfamiliar object. This task was broken into three blocks lasting about 15 minutes each, with rest periods in between. *During the learning phase, participants only heard the reduced form of the new words (i.e., /t/ and /nt/ were always flapped).* On each trial, two unfamiliar objects were displayed side by side on a computer monitor, and the name of one of the two pictures was played over headphones; word onset co-occurred with picture onset. Participants were instructed to select the picture they believed matched the word by pressing one of two buttons. For example, if they believed the right picture matched the word, they pressed the right button. After they pressed a button, or after a five second timeout, participants received feedback. The correct picture (i.e., the one matching the word) stayed on the screen and the incorrect one disappeared.

Depending on which group participants were assigned to, the feedback could also include the printed form of the word. Participants in two of the groups were given spellings for half of the new words (i.e., 16 items), and no spellings for the other half: The Canonical_None group only learned spellings that were consistent with the canonical pronunciation of the words (e.g., *sento* for [sɛ̃ɾo], whereas the Reduced_None group only learned spellings consistent with the reduced pronunciation of the words (e.g., *senno* for [sɛ̃ɾo]). The third group, Canonical_Reduced, received a spelling for all 32 new words. Half of the words were learned with a spelling that matched the canonical pronunciation, and the other half with a spelling that matched the reduced pronunciation. In total, during the learning phase, each word-picture pair was presented 20 times. Previous studies using a similar learning paradigm have shown that 20 repetitions will be more than enough for listeners to accurately match new words with their referents (e.g., Leach & Samuel, 2007; Samuel & Larraza, 2015). This phase lasted approximately 45 minutes.

Picture matching task. During this task, participants saw only one picture on each trial. The picture was presented simultaneously with an auditory word. Participants were instructed to indicate whether they believed the word matched the picture by pressing one of two buttons labelled “yes” and “no”. After participants pressed a button, or after a three second timeout, the next trial began; no feedback was presented. The newly learned words were presented once in their reduced form (the form that had always been presented during training) and once in their canonical form. This was the first time participants were exposed to the canonical pronunciation of the new words. The control words were presented once in their correct form (e.g., “basketball”) and once in their incorrect, nonword version (e.g., “bashketball”). The nonwords differed from real words by a single feature, requiring the participants to notice small deviations from the correct form. Accuracy for control words and their nonword counterparts gave a measure of how engaged in the experiment participants were. These trials were included to make sure that the participants knew that even a single-feature deviation should receive a “no” response. In addition, the task instructions emphasized this point: To ensure that participants would not accept “close enough” pronunciations as correct, they were informed that they would hear small deviations in the word forms they had learned and were instructed to only accept a name for a picture if it matched exactly what they had learned. Given these instructions and context, participants should reject the canonical versions (e.g., auditory “trotty”, after having learned auditory “troddy”), just as they should reject “bashketball” as the name for a picture of a basketball.

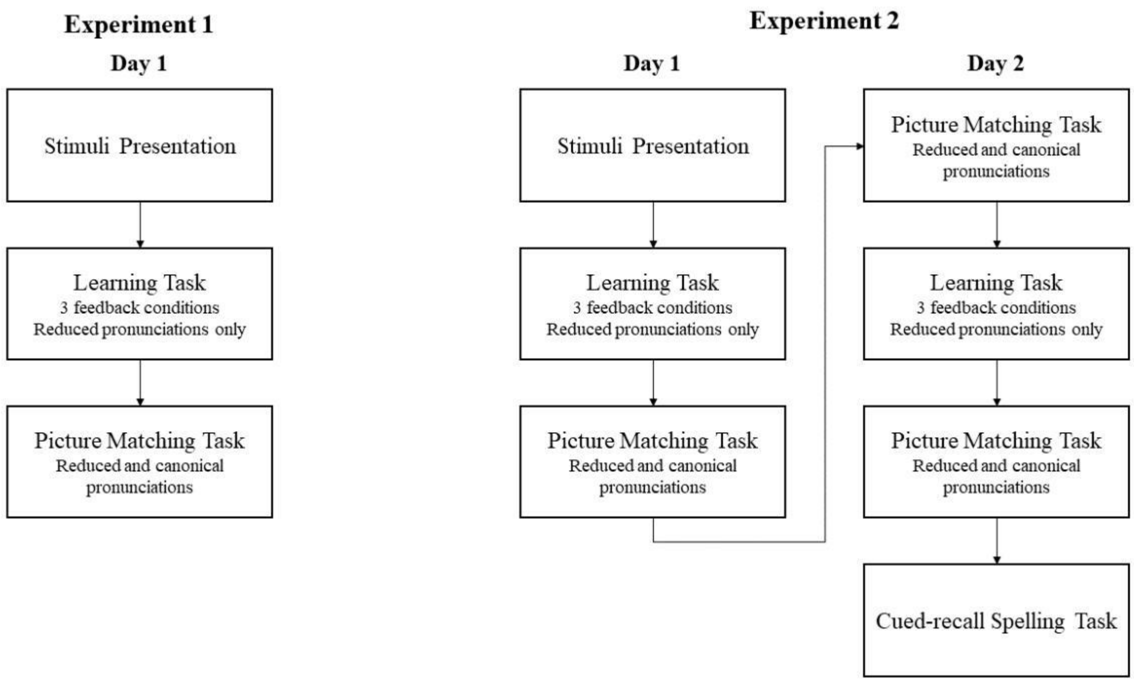


Figure 1. Schematic illustration of the organization of the different tasks in Experiment 1 and Experiment 2

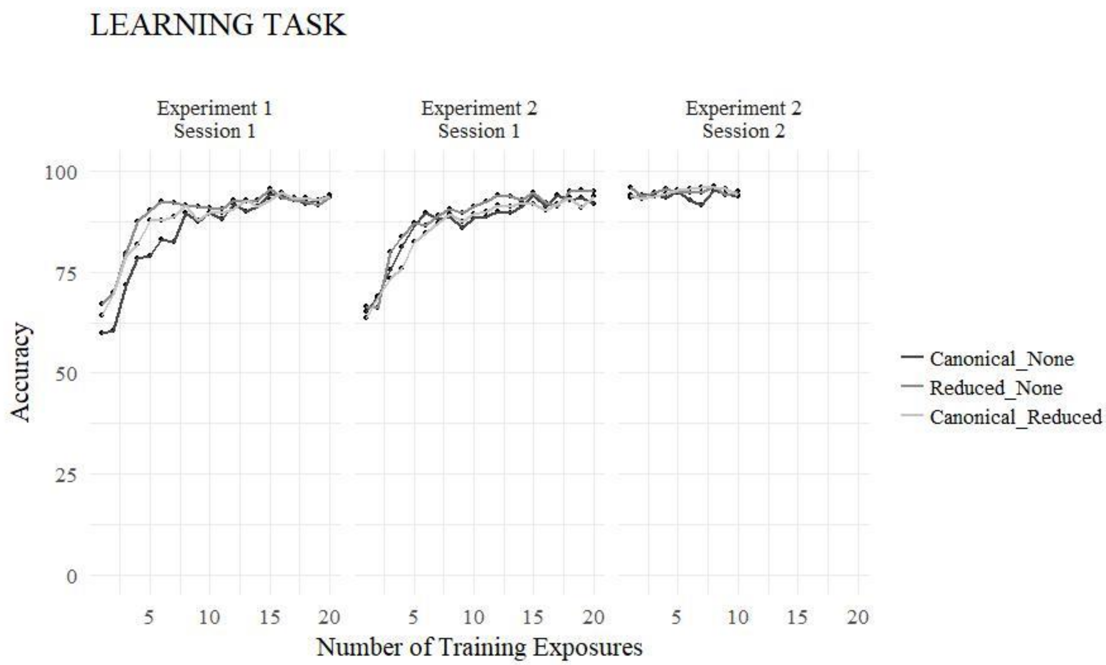
Results

In all tasks, our primary focus was on accuracy. Accuracy data were analyzed with generalized linear mixed models with a binomial link function (e.g., Baayen, Davidson & Bates, 2008), using the R statistical software package (R Core Team, 2015). Specifically, we used the *glmer* function of the R package *lme4* for our analyses (Bates, Mächler, Bolker, & Walker, 2015). All mixed effects models included the maximal random effect structure justified by the data (as determined by a stepwise selection procedure) while also avoiding overfitting (Bates, Kliegl, Vasishth, & Baayen, 2015). All models included at least by-subject and by-item random intercepts. Main effects were evaluated with log likelihood tests that measured how much a model fit increased when a new factor was added. All models' specifications can be found in Appendix F.

We also measured reaction times, recorded from word onset. However, our study was not designed to compare reaction times between canonical and reduced pronunciations of the items. In the picture matching task, the canonical pronunciation of the newly learned words called for a “no” answer, while the variant pronunciation called for a “yes”. Performance for the reduced pronunciations was almost perfect; participants had learned these forms during the learning task and recognized them very well. In contrast, performance for the canonical pronunciations averaged 50% overall. Reaction times in erroneous trials tend to be slower, and they also tend to be slower for “no” responses compared to “yes”. Finally, the reduced pronunciations were shorter than their canonical counterparts. These differences prevent the reaction time data from being interpreted in a meaningful way. However, interested readers can find means and analyses in Appendix B.

The data from 16 participants were excluded because they performed very poorly during the last block of the learning phase and/or on the control item trials during testing. Our bases for exclusion were 1) performance below two standard deviations from the group mean during the learning phase and/or 2) performance below two standard deviations from the group mean on the familiar words trials during the picture matching task. In addition, one subject was excluded because of experimenter error. Responses made in less than 250ms (< 1% of all observations) were removed from the data set, as they are likely too fast to reflect participants’ processing of the words.

Learning Task. As Figure 2 shows, average accuracy rapidly increased with the number of



word-object pair repetitions in all groups. As in previous studies using a similar learning paradigm (e.g., Samuel & Larraza, 2015), participants were near ceiling after approximately 10 repetitions. Therefore, we can be confident that after 20 repetitions, participants knew the pairings well. More detailed analyses of the learning data can be found in Appendix D.

Figure 2. Average accuracy of each group during the learning of the word-picture pairs as a function of the number of exposures to each word-picture pair: 20 for Experiment 1 and for Session 1 of Experiment 2, and an additional 10 for Session 2 of Experiment 2.

Picture Matching Task. During the critical trials of the picture matching task, participants were exposed to the canonical pronunciation of the newly learned words for the first time. Performance on this task informed us about whether listeners erroneously accept canonical pronunciations as "old" more often if the original words had been learned with canonical orthography rather than reduced orthography or no orthography. Control trials included the reduced pronunciation of newly learned words, as well as correct and incorrect pronunciations of English words (e.g., “basketball” and “bashketball”). As expected, participants were very accurate on the control words trials, with 96% correct report overall in accepting the correct pronunciation of English words, and 90% correct report overall in rejecting their incorrect pronunciation. Also matching expectations, participants performed well on the new-word trials when they were presented in their (trained) reduced pronunciation, with 95% accuracy overall (see Figure 3). Participants’ accuracy on these trials was not moderated by the spelling they had learned with the words (93% accuracy for words with no accompanying spelling, 97% for words accompanied with a reduced spelling, and 96% for words accompanied with a canonical spelling).

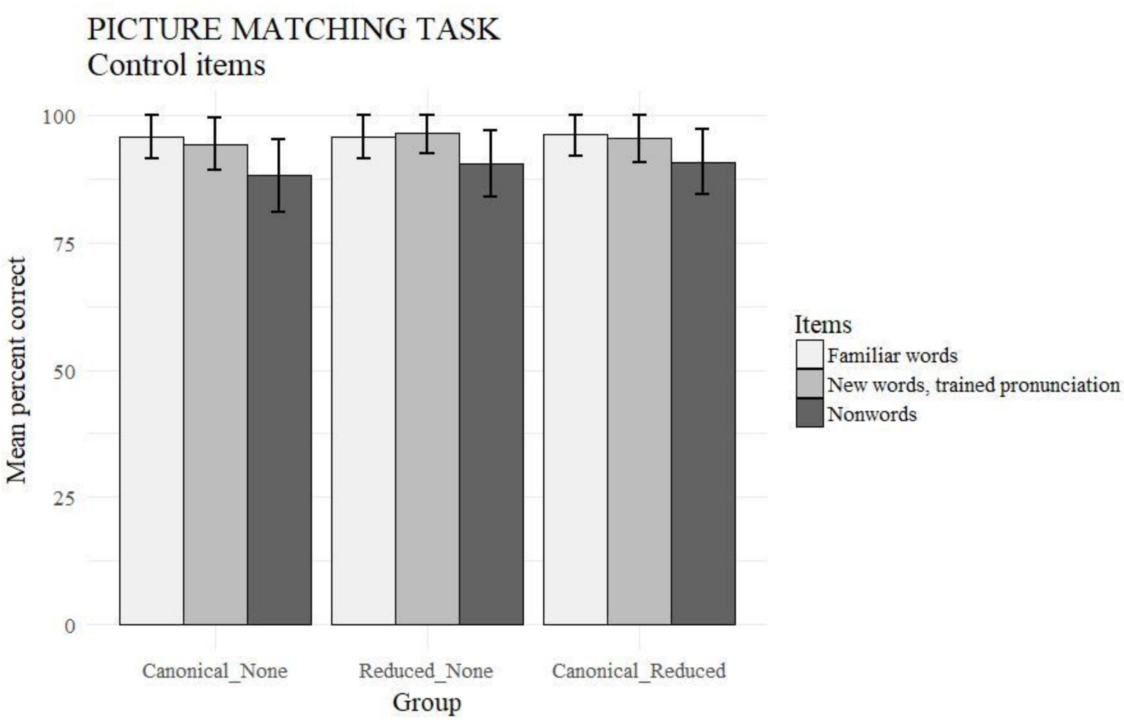


Figure 3. Accuracy of each group of participants on the picture matching task of Experiment 1. “FAMILIAR WORDS” were correctly pronounced common English words. “NEW WORDS” were the novel words presented in their trained pronunciation. The “NONWORDS” were derived from the familiar English words but differed by one segment and were thus improperly pronounced.

The critical trials were ones in which new words were presented in their canonical form (i.e., with a full [t] or [nt] sound). On these trials, participants were correct if they *rejected* the canonical form as a correct name for an object because they only heard reduced forms during training. Because we specifically instructed participants to pay attention to small deviations in pronunciation and to only accept a word if it exactly matched what they had learned during the learning phase, erroneously accepting the canonical form would suggest that it is represented in lexical memory. If such representations are promoted through experience with words' spellings, the false acceptance rate should vary as a function of the spelling learned with each word.

We used generalized linear mixed effect models to assess whether there was a main effect of spelling. Spelling was entered as a fixed factor, with three levels that reflected which spelling participants had learned for a given word: 1) Canonical, for spellings that were consistent with the canonical pronunciation, 2) Reduced, for spellings that were consistent with the reduced pronunciation, and 3) None, for words that were not trained with an accompanying spelling. Accuracy was the dichotomous dependent variable (1 for correct; 0 for incorrect). The reference level for the intercept was set to the Canonical spelling. For this analysis, we used data from all three experimental groups.

Figure 4 shows the average accuracy in rejecting canonical pronunciations as a function of the spelling information that had been provided. Participants' accuracy for new words learned with a canonical spelling was very poor (M = 38%) – the canonical pronunciations were usually accepted as matching the picture, despite their phonetic mismatch with the words that had been learned. Accuracy for words learned with no accompanying orthography was noticeably better (M = 63%), as was performance for words learned with an orthography that matched the trained (reduced) pronunciation (M = 71%). There was a robust effect of Spelling ($\chi^2 = 30.9$ $df = 2$, $p < 0.0001$). The accuracy for words learned with a canonical spelling was significantly worse than for words learned with no spelling (β = -1.260, *SE*=0.222, *Wald's z* = - 5.68, $p < 0.0001$), and than for words learned with a reduced spelling (β =-1.594, *SE* = 0.341, *Wald's z* = - 4.67, p <0.0001). Accuracy for words learned with no spelling was numerically but not significantly worse than accuracy for words learned with a reduced spelling (β = -0.333, *SE* = 0.216, *Wald's z*-1.54, p = 0.1230).

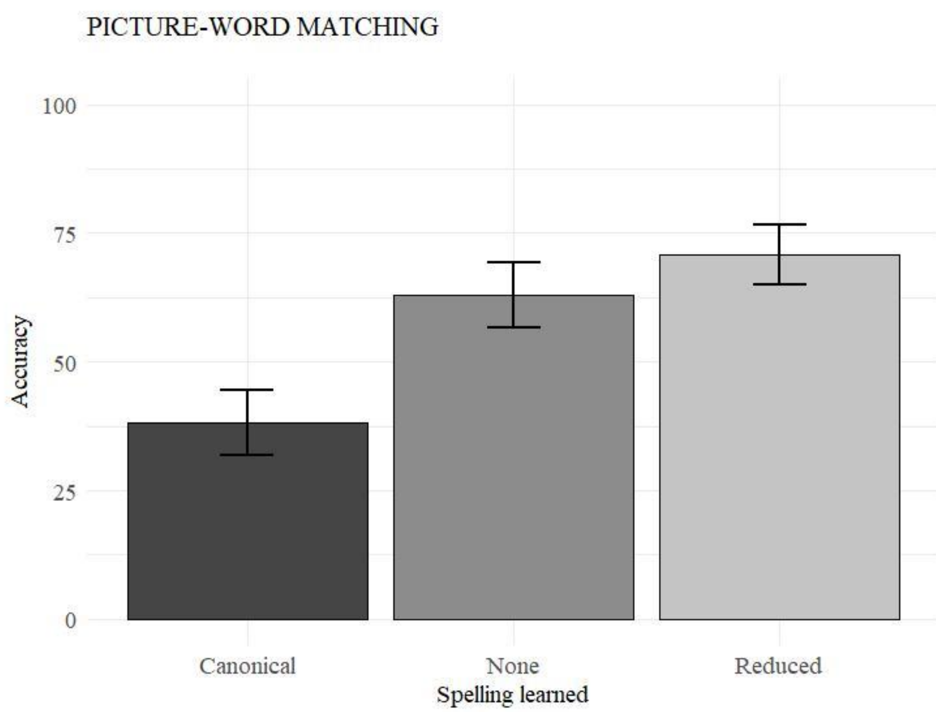
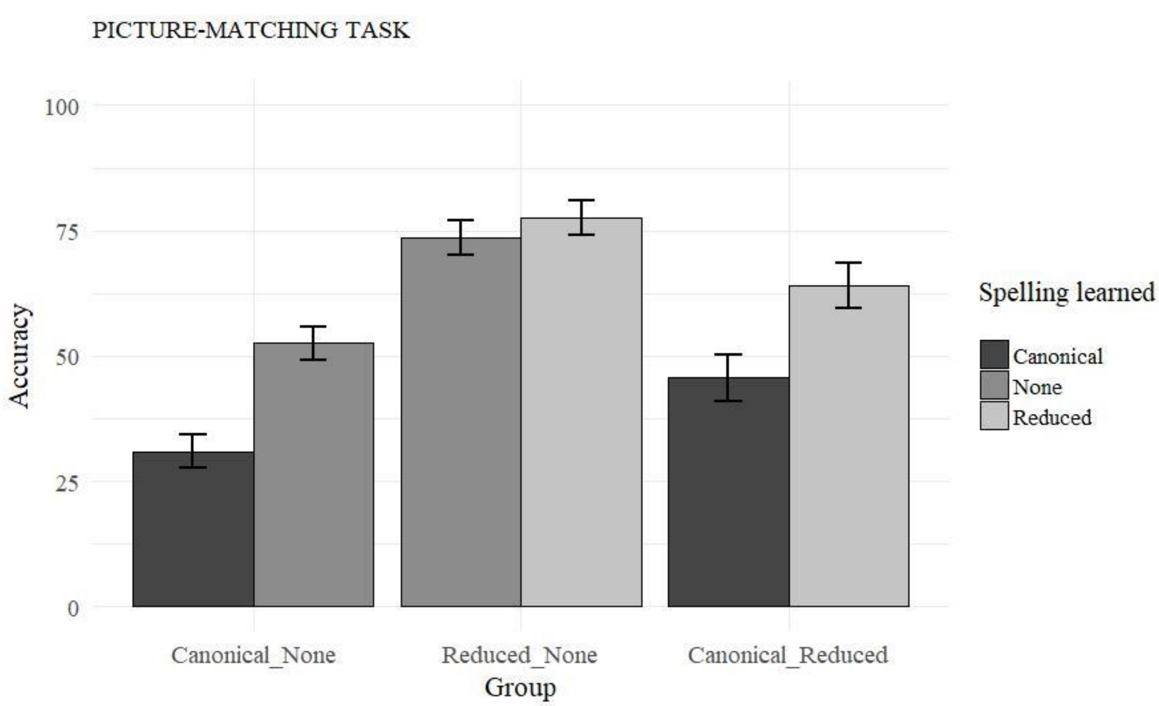


Figure 4. Accuracy for the critical trials of Experiment 1's picture matching task as a function of the new words' spellings, averaged across the three participant groups. Being accurate on these trials meant rejecting the canonical pronunciation (since it had never been heard before)

Learning a spelling consistent with the reduced form did not completely abolish acceptance of the canonical form: Listeners still accepted the latter approximately 30% of the time. This presumably is a result of a lifetime of experience with American English, with the reduced form's dominant frequency of occurrence for the intervocalic /t/ consonant and /nt/ clusters. Despite this inherent preference, the significant main effect of Spelling demonstrates that spoken word recognition is strongly influenced by words' orthographic form.

Do listeners accept the canonical pronunciation of a word more when they learned a canonical spelling for that word, compared to a reduced spelling?

Participants in the Canonical_Reduced group learned half of the new words with a spelling consistent with the canonical pronunciation and half with a spelling consistent with the reduced pronunciation. This group provides a direct within-subject test of the orthographic hypothesis. Again, our dependent variable was Accuracy (1 for correct, 0 for incorrect), and the independent variable was Spelling (Canonical versus Reduced). As Figure 5 shows, accuracy on the critical trials was poorer for words learned with a canonical spelling (*M* = 46%) compared to words learned with a reduced spelling (*M* = 64%). There was a main effect of Spelling ($\chi^2 = 5.53$, $df = 1$, $p = 0.019$), with more accurate performance for words learned with a variant spelling rather than canonical ($\beta = 1.076$, $SE = 0.446$, *Wald's z* = 2.41, $p = 0.016$). Listeners accepted the canonical pronunciation of the newly learned words more when the words' spelling matched the canonical pronunciation than



when it did not.

Figure 5. Correct rejection rate for canonical test words on Experiment 1's picture matching task, as a function of group (Canonical_None, Reduced_None and Canonical_Reduced) and the type of spelling learned for each new word.

Do listeners generalize their knowledge about sound-to-spelling rules to spoken words for which they have no orthographic experience?

The preceding analyses confirm our hypothesis that knowing an orthographic form for a word can drive the canonical advantage observed in previous studies. Next, we examine whether new words learned with no spelling can inherit the orthographic properties of the other new words they were learned with and that did receive a spelling. Participants in the Canonical_None and Reduced_None groups learned half of the words with no accompanying spellings. We compared performance between these two groups for words that had been learned with no accompanying orthography. As Figure 5 shows, participants who had learned reduced spellings for half of the words rejected the canonical pronunciation of the words learned without orthography more often (M = 74%) than participants who had learned canonical spellings for half of the words (M = 52%). This difference in rejection rate between the two groups was significant, with better accuracy on no-spelling words for the Reduced_None group (β = 1.239, SE = 0.545, *Wald's z* = 2.27, p = 0.023). This result indicates that listeners generalized the phoneme-grapheme mapping they were given for some new words (e.g., the intervocalic t-flap is spelled with "tt") to the words being learned in the same context, even in the absence of orthographic information for those words. In turn, this impacted participants' acceptance rate of the canonical pronunciation of the new words on the picture matching task.

Does orthographic inconsistency modulate the effect of orthography on spoken word recognition?

A potentially important difference between two of the groups (Canonical_None and Reduced_None) and the third one (Canonical_Reduced) is that the "rule" for mapping a spelling to a sound was simpler for the first two groups. For participants in the Canonical_None and Reduced_None groups, when the critical flap and nasal flap phonemes were mapped to a spelling, it was always the same spelling. For example, when participants in the Canonical_None group were given a spelling for the [ɾ] sound, it was always "tt". Participants in the Canonical_Reduced group, in contrast, were given a more complex rule because one sound could be mapped to two different spellings. Half of the time, the [ɾ] sound was written as "dd" while the other half of the time it was written as "tt".

We have already seen that overall, accuracy for words learned with a canonical spelling was poorer than for words learned with a reduced spelling. However, this difference was larger when comparing performance between the Canonical_None group (*M* = 31% accuracy for words learned with a canonical spelling) and the Reduced_None group (*M* = 78% accuracy for words learned with a reduced spelling) than when looking at these same words within the Canonical_Reduced group (*M* = 46% accuracy for canonical, *M* = 64% accuracy for reduced), as shown in Figure 4. We used a best-fit model with by-subject and by-item random intercepts to evaluate the interaction between the grapheme-phoneme mapping complexity and the spelling learned for each word. Fixed factors were Spelling, Learning Environment (complex or simple) and their interaction. For the Canonical_None and Reduced_None groups, Learning Environment was coded as *simple* (one-to-one mapping between phoneme and grapheme). For the Canonical_Reduced group it was coded as *complex* (one-to-many mapping between phoneme and grapheme). We used a log-likelihood test to compare this model to a model that did not contain the interaction term. The interaction between learning environment and the effect of spelling was significant ($\chi^2 = 17$, $df = 1$, $p < 0.0001$), confirming that the effect of spelling was stronger for participants who had learned a simple phoneme-to-grapheme mapping (e.g., [ɾ] is always spelled “dd”) compared to those who had learned a more complicated rule.

Discussion

Experiment 1 tested whether orthography could be driving the advantage for canonical pronunciations during spoken word recognition, an advantage that has been found even when the canonical form is relatively rare. We found that when listeners learned words that included a reduced variant with a spelling that favored the canonical form, they were more likely to accept the canonical form as the word they had learned. This result suggests that listeners developed a phonological representation consistent with the word’s spelling. Note, however, that recognition of the reduced forms was neither hurt nor enhanced by the spellings learned (see Figure 3), suggesting that orthographical knowledge does not impair recognition of these forms.

There is, in fact, prior evidence that reading involves phonological recoding (see Leinenger, 2014, for a review), so that reading a word with a “t” could lead to implicitly hearing the same word with a full [t] (see Mitterer & Reinisch, 2015, for such a suggestion). Interestingly, listeners not only showed evidence of modifying their phonological representations when orthography was given for a word being learned, they also generalized to words that were learned without any explicit orthographic information. Presumably, they inferred the orthographic form of these words from their experience with the words that were given spellings. During word recognition the modified phonological representations were activated, causing listeners to accept the canonical pronunciation as a correct form at significantly higher rates.

We also found that the effect of orthography was stronger when the mapping between a phoneme and its spelling was simple. For example, if the t-flap was consistently spelled with “tt”, listeners were more likely to accept the full [t] variant than if it was sometimes spelled with “tt’ and other times spelled as “dd”. Note that overall, regardless of the spelling they had learned, listeners were not very good at correctly rejecting the canonical pronunciation. There are at least two reasons why this might have occurred. First, previous research has demonstrated that learners may only encode information relevant to the task at hand (e.g., Sulpizio & McQueen, 2011, 2012). Our listeners might not have paid attention to acoustic details that were not useful at the time: During the learning task in our experiment, the words were relatively easy to distinguish. Listeners may not have encoded enough information about the words during learning to be able to properly distinguish the two forms consistently. Second, listeners’ tendency to accept the canonical pronunciation could be driven by their long experience with the phonotactic and orthographic patterns of American English. They may have inferred that a flap sound should be represented with a “t” or “nt” (even when given a different spelling), leading to their substantial tolerance for canonical pronunciations. The effect of our experimental manipulation was overlaid on this existing experience.

Overall, the results of Experiment 1 provide evidence that orthographic knowledge could be driving the preference for canonical pronunciations observed in previous studies. The results suggest that listeners created phonological representations that were influenced by orthographic information. This is consistent with a study by Bakker, Takashima, van Hell, Janzen, and McQueen (2014) that showed that phonological representations can arise from a word’s written form and vice versa.

Experiment 2 provides a replication of the conditions of Experiment 1, to determine whether the effects observed there are reliable. In addition, it examines how the learned representations change over a two-day delay: Are these short-term episodic effects that disappear after a two-day delay, are they stable, or do they show evidence of strengthening over time through a sleep-based consolidation process, as has been observed for some types of lexical information (e.g., Bakker et al., 2014; Davis & Gaskell, 2009; Dumay & Gaskell, 2007; Gaskell & Dumay, 2003)?

Experiment 2: Does the effect of orthography change after a two-day delay?

The results of Experiment 1 suggest that the phonological representations for the newly learned words were influenced by orthographic information when it was available. Given the conventions of English orthography, this effect can bias perception towards the canonical pronunciation. As this was the first demonstration of a link between orthographic information and the canonical form's advantage, it is important to determine if this connection is reliable. Thus, in Experiment 2, we first replicated the training and testing conditions of Experiment 1, with new participants. We then had those participants return to the lab two days later and had them do the picture-matching task again. The purpose of this delayed test was to learn more about the representations that had been formed on the first day of training.

The delayed test was designed to reveal which of three potential outcomes would occur. One possibility is that the influence of the orthographic information is based on short-term episodic representations. If so, then after a two-day delay, we should find little or no influence of the orthography experienced on day 1. The second potential outcome is that stable lexical representations were developed during training, which should lead us to find essentially the same pattern after a two-day delay as found on day 1. The final possibility is that lexical representations are formed during the first session, but that two days and two nights later, these representations will show even stronger effects of the orthography.

The third possibility stems from previous research showing that newly learned words are not immediately fully lexicalized (e.g., Bakker et al., 2014; Davis & Gaskell, 2009; Dumay & Gaskell, 2007; Gaskell & Dumay, 2003). These studies have shown that while newly learned words can be correctly identified after a number of learning trials, the new words do not engage in lexical competition with existing words before a period of sleep (see Leach & Samuel, 2007, for a discussion of the distinction between lexical configuration and lexical engagement). Although their study did not involve the canonical advantage, Bakker and colleagues (2014) investigated sleep consolidation under conditions that share certain features with the current study. Their participants learned new words either in auditory form or in printed form, and in both cases the authors found consolidation effects: The newly learned words showed stronger competition effects after sleep. After replicating the basic consolidation effect in each modality, Bakker et al. investigated cross-modality effects: Would words learned in one modality compete with existing words when tested in the other modality? They found that cross-modal lexicalization took place after a longer delay than within-modality testing, with an even longer delay needed for words learned from print to compete with words learned auditorily. These results, as well as previous sleep consolidation findings (e.g., Gaskell & Dumay, 2003), raise the possibility that we will find a stronger effect of the orthographic information on a test two days after training than on a test conducted right after the training period.

In sum, we had participants do the same tasks as in Experiment 1 during their first session. They returned two days later and performed the picture matching task first so that we could determine whether the orthographic effect diminished (implicating episodic influences), remained stable, or increased (via consolidation). We then gave them additional training trials, followed by a final picture matching task, to see whether the orthographic influence would increase with additional training. Finally, participants completed a cued-recall spelling memory test to assess how well they had learned the spelling of the new words. Figure 1 shows the sequence of tests in Experiment 2. Each session took approximately one hour.

Method

Participants. 76 undergraduate students from Stony Brook University participated in this study. They were all 18 years old or older and were native American English speakers.

Stimuli. The stimuli were identical to those used in Experiment 1.

Apparatus and procedure. The same equipment was used as in Experiment 1. The first day of Experiment 2 (hereafter Session 1) was a replication of Experiment 1, following the same procedures. As in Experiment 1, there were three groups of participants: Canonical_None, Reduced_None, and Canonical_Reduced. Participants returned to the lab two days later, allowing for two nights of sleep between tests. They were first given the picture matching task (hereafter Session 2). This was followed by three additional blocks of training. Each block lasted about seven minutes, and each word was repeated ten times across the task (versus the 20 presentations during Session 1). Participants then completed the picture matching task again, to assess the effect of the additional training (hereafter Session 3). Finally, participants were given a cued recall spelling test, to assess how well they had learned the spellings. They saw the picture of each unfamiliar object for which they had learned a spelling, as well as the first two letters of the object's "name" and were instructed to complete the words' spelling as best as they could. The results of the cued-recall spelling task can be found in Appendix E.

Results

16 participants were excluded from further analysis due to poor performance, using the same criteria as in Experiment 1. All subsequent analyses were based on this data set, using the same statistical procedures as in Experiment 1. The reaction time analyses can be found in Appendix C. The learning task results were similar to Experiment 1 (see Figure 2 and Appendix D).

Accuracy on familiar English word trials was very high for both their correct (M = 97%) and nonword versions (M = 90%), across all three picture matching tasks (see Fig. 6). Accuracy for the new words when heard in their trained (reduced) pronunciation was also quite good (M = 94%). As in Experiment 1, participants' accuracy on the learning task reached ceiling after about ten repetitions during Session 1 and remained very high during Session 2 and 3 (see Figure 2).

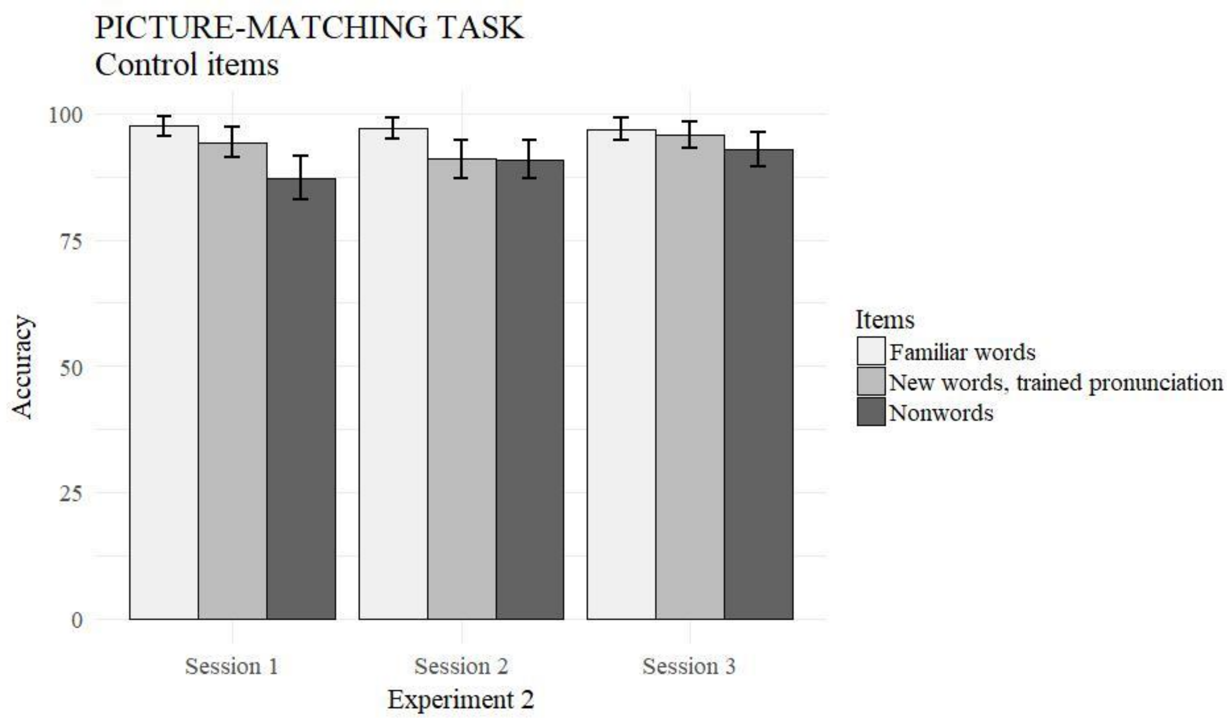


Figure 6. *Accuracy on control items for all participants in Experiment 2 (n = 60) and for each picture matching task. "Session 1" was the picture matching task at the end of Day 1. "Session 2" was the picture matching task at the beginning of Day2. "Session 3" was the picture matching task at the end of Day 2, taken after participants were given additional learning trials.*

Session 1: Do we replicate the effects observed in Experiment 1?

As Session 1 was a replication of the first experiment, we conducted the same analyses.

Is the main effect of spelling replicated? As before, we used generalized linear mixed models to analyze the main effect of spelling. Figure 7 shows the average performance on the picture matching task for canonical form trials, forms that should have been rejected because all the training was with reduced forms. The pattern in Experiment 2 closely matches what was found in Experiment 1: 43% average accuracy for items learned with a canonical spelling, 62% for those learned with no spelling and 67% for those learned with a variant spelling (see Figure 4 to compare). A log likelihood test revealed that the main effect of spelling was significant (χ^2 = 24.2, *df* = 2, *p* < 0.0001). Listeners were more likely to erroneously accept the canonical pronunciation of the new words when they had learned a spelling consistent with that pronunciation compared to a reduced spelling (β = -1.338, *SE* = 0.307, *Wald's z* = -4.37, *p* < 0.0001) or no spelling at all (β = -0.973, *SE* = 0.200, *Wald's z* = -4.86, *p* < 0.0001). As in Experiment 1, the difference between words learned with a reduced spelling and no spelling did not reach significance (β = -0.365, *SE* = 0.201, *Wald's z* = -1.82, *p* = 0.0686). Thus, the core result replicates well.

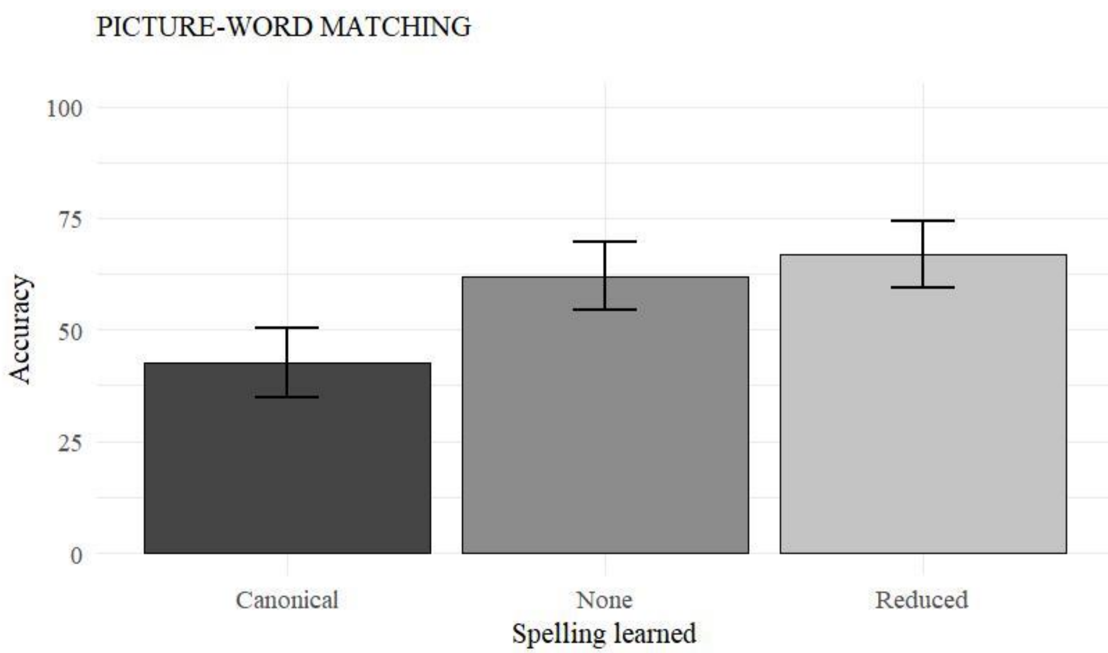
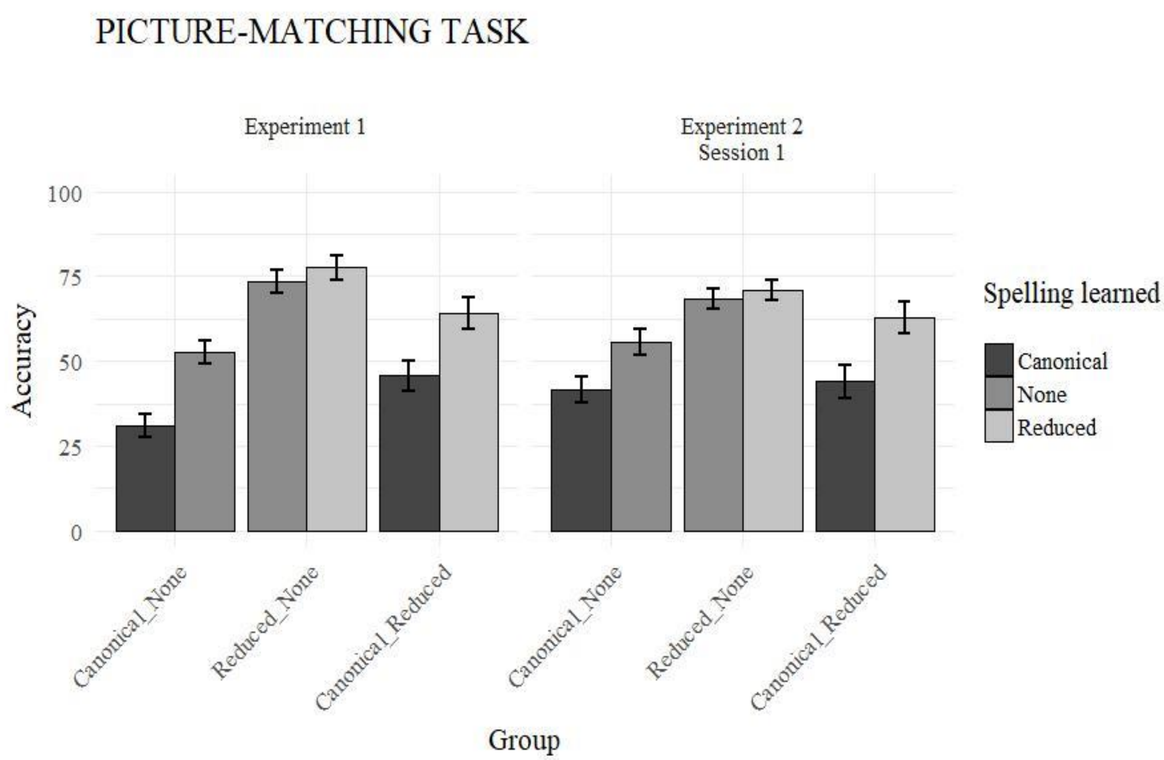


Figure 7. Accuracy for the critical trials of Experiment 2's first picture matching task as a function of the new words' spellings, averaged across the three participant groups. Being accurate on these trials meant rejecting the canonical pronunciation (since it had never been heard before)

Is the within-subject effect of spelling replicated? As in Experiment 1, when participants in the Canonical_Reduced group had learned a spelling consistent with the canonical pronunciation, they failed to reject the canonical form during spoken word recognition and were less accurate (*M* =

44%) than if they had seen spellings consistent with the reduced form (*M* = 63%), (β = -1.029, *SE* = 0.379, *Wald's z* = - 2.72, *p* = 0.0066).

Is the generalization effect found in Experiment 1 replicated? In Experiment 1 we found that the "no spelling" words inherited the properties of the "with spelling" words. The same pattern was found in Experiment 2, as shown in Figure 8: On the picture matching test, participants who had learned half of the words with a canonical spelling were less accurate (*M* = 56%) on the words that had been learned with "no spelling" than participants who had learned half of the words with a reduced spelling (*M* = 68%). However, while the patterns across Experiments 1 and 2 are similar, in the first test of Experiment 2, this difference did not reach significance (β = -0.714, SD = 0.501,



Wald's z = -1.43, *p* = 0.15).

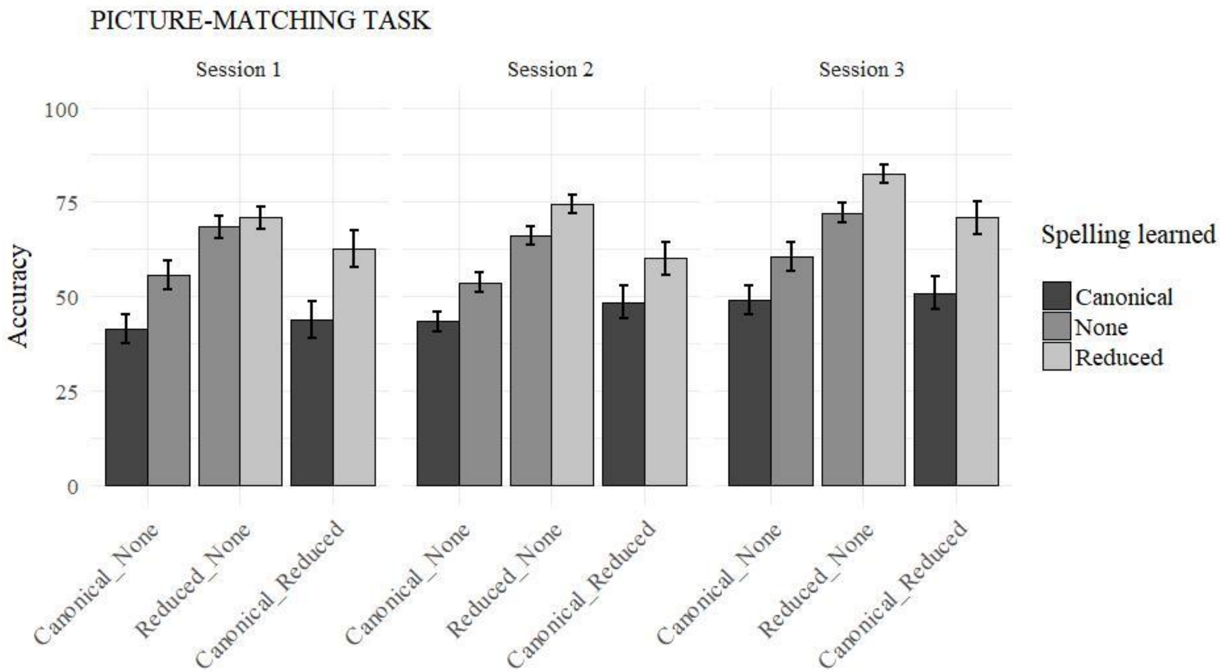
Figure 8. Correct rejection rates for canonical words on the picture matching task for Experiment 1 (left) and the first session of Experiment 2 (right), as a function of the spelling learned for each word.

Was the interaction between the effect of spelling and the effect of learning environment replicated? In Experiment 1 we found a significant interaction between learning environment and orthography. As with the generalization test, although the numerical pattern is similar in Experiment 2 Session 1, the interaction did not reach significance (χ^2 = 1.79, *df* = 1, *p* = 0.18).

Does the orthographic effect change after two days? As we noted, the results from Session 1 provide a strong replication of the core result of Experiment 1: Orthographic information can drive the preference for the canonical form. The second goal of Experiment 2 was to see whether the effect of orthography on processing phonological variants changes after two days and two nights. Recall that we outlined three possible outcomes: The effect could fade over two days, it could remain essentially the same, or it could strengthen if it is subject to the kind of consolidation process that has been found for some aspects of lexicalization (e.g., Bakker et al., 2014; Dumay & Gaskell, 2007; Gaskell & Dumay, 2003).

As before, to evaluate the effect of orthography on canonical form recognition, we measured participants' responses to trials of the picture matching task when they were presented with the canonical pronunciation of the new words. Accuracy on these trials is *higher* if participants *reject* the canonical pronunciation, since it is not the form they heard during the learning task. The following analyses compare performance on Session 1 (shown on the left side of Figure 9) and Session 2's picture-matching tasks (shown in the middle of Figure 9).

We used generalized mixed models to compare participants' performance on these two picture matching tasks. The fixed effects were Spelling, Group and Session (two levels: Session 1 and Session 2). We tested for a main effect of Session by evaluating whether adding this factor increased model fit, using a likelihood ratio test. We found no main effect of Session (χ^2 = 0.23, *df* =



1, *p* = 0.63), indicating that there were no significant differences in performance on the two picture-matching tasks. As Figure 9 shows, the effect of orthography was very stable after two days.

Figure 9. Correct rejection rates for canonical words on the three picture matching tasks of Experiment 2, for each experimental group. "Session 1" was the picture matching task at the end of Day 1. "Session2" was the picture matching task at the beginning of Day 2. "Session3" was the picture matching task at the end of Day 2, taken after participants were given additional learning trials

As this stability implies, Session 2 replicated the main effect of Spelling (χ^2 = 7.18, *df* = 2, *p* = 0.028), with a significant difference in accuracy between words learned with a canonical spelling versus no spelling (*β* = -0.462, *SE* = 0.199, *Wald's z* = -2.32, *p* = 0.0201), and between words learned with a canonical spelling versus a reduced spelling (*β* = -0.813, *SE* = 0.301, Wald's z = -2.70, *p* = 0.0069). Participants were marginally more likely to correctly reject the canonical pronunciation for words learned with a reduced spelling compared to no spelling (β = 0.351 *SE* = 0.199, *Wald's z* = -1.76, *p* = 0.079)

We assessed the effect of the spelling learned on accuracy during the picture-matching task within the Canonical_Reduced group, providing a within-subject test of our hypothesis. The difference in accuracy for words learned with a canonical spelling (*M* = 49%) and reduced spelling (*M* = 60%) was significant (β = 0.656, *SE* = 0.330, *Wald's z* = 1.98, p = 0.047), replicating Session 1 and Experiment 1. Next, we compared performance on the no spelling trials for the Canonical_None versus the Reduced_None groups. On average, participants in the Reduced_None group were more likely to reject the canonical pronunciation (*M* = 66%) than participants in the Canonical_None group (*M* = 54%) for words that weren't learned with any spellings, but this difference did not reach significance (β = - 0.634, *SE* = 0.470, *Wald's z* = - 1.35, p = 0.18). Finally, we tested the interaction between the complexity of the grapheme-to-phoneme mapping and the effect of orthography on spoken word recognition. Recall that this interaction was significant in Experiment 1 but did not reach significance in Session 1 of Experiment 2. On the Session 2 picture matching test, the interaction was significant (χ^2 = 5.8, *df* = 1, *p* = 0.016): When the mapping rule was complex (the Canonical_Reduced group), the effect of orthography was not as strong ($M_{canonical}$ = 49%; $M_{reduced}$ = 60%) as when the mapping rule was simple (the Canonical_None group - $M_{canonical}$ = 43%; Reduced_None group -$M_{reduced}$ =74%).

Effect of additional exposure?

In Experiment 2, after the participants did a picture-matching task to assess the effect of the two-day delay, they were given three more blocks of training (ten more repetitions of each novel word-unfamiliar object pair). This additional exposure was then followed by a final picture-matching task (Session 3). Our analysis tests whether the additional training changed the effect of orthography. The right side of Figure 9 shows the relevant results.

Giving participants some additional exposure improved their accuracy overall on the last picture matching task (β = 0.492, *SE* = 0.142, *Wald's z* = 3.47, *p* = 0.001). Participants were more likely to correctly reject the canonical pronunciation for words learned with a reduced spelling (Session 2 *M* = 67%; Session 3 *M* = 77%), words learned with no spelling (Session 2 *M* = 60%, Session 3 *M* = 66%) and words learned with a canonical spelling (Session 2 *M* = 46%, Session 3 *M* = 50%). Importantly, this increase in accuracy did not change the overall pattern of results observed in Experiment 1 and the first two picture-matching tasks in Experiment 2. The main effect of Spelling in this third Picture Matching task was significant (χ^2 = 16.6, *df* = 2, *p* = 0.0002). Participants' accuracy was worse for words learned with a canonical spelling compared to no spelling (β = -0.738, *SE* = 0.234, *Wald's z* = -3.16, *p* = 0.0016) and to a reduced spelling (β = -1.498, *SE* = 0.360, *Wald's z* = -4.16, $p < 0.0001$). In addition, participants were more accurate for words learned with a reduced spelling compared to no spelling (β = 0.760, *SE* = 0.231, *Wald's z* = 3.29, *p* = 0.001).

Similarly, the effect of orthography within the Canonical_Reduced group was replicated, with participants more likely to erroneously accept the canonical pronunciation of words when they had learned a canonical spelling for them (β = - 1.106, *SE* = 0.409, *Wald's z* = - 2.70, *p* = 0.0069). The difference in accuracy for words learned with no spelling trials in the Reduced_None (*M* = 72%) and Canonical_None (*M* = 61%) groups did not reach significance (β = 1.193, *SE* = 0.796, *Wald's z* = 1.50, *p* = 0.13), although the pattern is similar to that in the previous tests. Finally, the interaction between the complexity of the grapheme-to-phoneme mapping and the effect of orthography on spoken word recognition was replicated in this third Picture Matching task (χ^2 = 4.09, *df* = 1, *p* = 0.043).

Cross experiments analysis.

Across Experiments 1 and 2, the pattern of performance was very similar, and the core result was robustly significant in both experiments. For two of the secondary questions, while the patterns were very similar across experiments, significant effects in Experiment 1 did not reach significance in (parts of) Experiment 2. In particular, the generalization effect did not reach significance, and the interaction between the effect of learning environment and the effect of orthography failed to reach significance in the first session of Experiment 2.

To test whether there were genuine differences between Experiments, we conducted cross experiment analyses. The reference level for the Session variable was set to Session 0 (i.e., Experiment 1) for all models. First, we looked at the robustness of the generalization effect by comparing the performance of the Canonical_None and Reduced_None groups on the critical trials of the Picture Matching tasks for words learned without a spelling. The model included Group, Session (four levels: Session 0, 1, 2 and 3) and their interaction as fixed factors. Accuracy for words learned without spellings was higher for the Reduced_None group (β =1.238, *SE* = 0.362, *Wald's z* = 2.38, *p* = 0.017); this is the generalization effect. The interaction between Group and Session was not significant (χ^2 = 1.99, *df* = 3, *p* = 0.57); there is no evidence that the generalization effect varies significantly across experiments and sessions.

Next, we investigated the interaction between learning environment (simple versus complex) and the effect of orthography across all sessions. We compared performance of the three groups on the critical trials of the Picture Matching for words that had received a spelling (either reduced or canonical). The model included Spelling (reduced or canonical), Learning Environment (simple or complex), Session, and their two-way and three-way interaction terms. Note that to allow this model to converge, we adjusted the model's integer scalar setting (nAGQ argument was set to 0) which resulted in a faster, but slightly less accurate form of parameter estimation (Bates et al., 2015).

Under these conditions, the two-way interaction between Spelling and Learning Environment was significant: Accuracy was worse when the learning environment was complex (i.e., when critical phonemes could be spelled in different ways- β = -1.582, *SE* = 0.397, *Wald's z* = -3.99, $p <$ 0.001). This is consistent with the interaction effect found in Experiment 1 and parts of Experiment 2: The effect of orthography was stronger when the mapping between phonemes and graphemes was simple. This pattern did not depend on the Session/Experiment: The three-way interaction term was not significant ($\chi^2 = 4.5$, $df = 3$, $p = 0.21$).

Discussion

Experiment 2 had two central goals: to test how reliable the findings of Experiment 1 were, and to determine whether the orthographic influence found in Experiment 1 changes as a function of time and/or sleep. The critical result in Experiment 1 was a demonstration that orthography produces a substantial effect on the processing of phonological variants, an effect that could underlie listeners' preference for canonical pronunciations. The results of Experiment 2 provide clear answers to our two key questions. First, although some effects did not reach significance in Experiment 2, the pattern of results we found in Experiment 1 was replicated, and the core effect of orthography was statistically reliable in both Experiments. Second, within Experiment 2, there was no noticeable change in the pattern two days after words had been learned. In addition to the patterns being numerically similar across experiments, we found no evidence for a difference across experiments with our cross-experiment analyses, which suggests that the generalization effect and the learning-condition complexity effects were robust across sessions.

The fact that the effect of orthography persisted after a two-day delay rules out the possibility that it was driven by ephemeral episodic traces rather than encoded information about the words. If this had been the case, accuracy on the first picture matching task of Session 2 should have been about the same for all three groups. Instead, the initial picture matching results in Session 2 replicated the substantial influence of the orthographic information that had been learned. Note that this close replication also demonstrates that the orthographic influence did not increase after two nights of sleep. Research on new word acquisition has shown that sleep-based consolidation strengthens lexical representations (e.g., Bakker et al. 2014; Gaskell & Dumay, 2003). The absence of consolidation effects here could indicate that the kind of information driving the orthography effect does not require consolidation.

Alternatively, the conditions in Experiment 2 might not have been conducive to consolidation. The delay between the first and second picture matching tasks (48 hours) may have been too short to allow for consolidation of this type of information to occur. Although many studies have shown consolidation effects within 24 hours, there is some evidence that cross-modal lexicalization may require a longer consolidation period than intra-modal lexicalization. Bakker et al. (2014) tested whether a consolidated phonological representation can be established for a word learned only in its written form (i.e., print-to-speech lexicalization). They found that when listeners were trained and tested with new words in the same modality (for example, all in speech), consolidation was observed 24 hours after training. However, there was no evidence for print-to-speech lexicalization at this same time point, suggesting that a phonological representation of the printed words had not yet been consolidated. Bakker et al. did find evidence for the emergence of a phonological representation for words learned in written form when the test took place a week after the original training.

General Discussion

The current study was motivated by the surprising result, found in previous spoken word recognition research, that listeners generally prefer canonical pronunciations, even when these are not frequently encountered in speech. This "canonical advantage" has been replicated in multiple studies using different paradigms and cannot be explained by existing accounts of phonological variant processing based on underspecification, feature parsing, tolerance or frequency. Given the failure of these hypotheses, Ranbom and Connine (2007) suggested that the canonical advantage might arise because of listeners' experience with orthography. They speculated that listeners develop a phonological representation that contains a canonical sound because such a sound matches the word's written form. Previous studies in second language learning (e.g., Escudero, Simon, & Mulak, 2014; Escudero & Wanrooij, 2010) and reading acquisition (e.g., Racine et al., 2014) reported findings consistent with this hypothesis.

The current study provides a direct test of whether orthography influences how adult listeners process phonological variants in their native language. We used a word learning paradigm in which all participants learned the same new words, with the same reduced pronunciations, but paired with different spellings. New words could either be paired with a spelling consistent with their canonical pronunciation (e.g., [sɛɾo] paired with the spelling *sento*), their reduced pronunciation (e.g., [sɛɾ̃o] paired with the spelling *senno*), or without any spelling. Across two experiments using similar procedures, we found that listeners were consistently more likely to accept pronunciations promoted by the words' spellings. Critically, listeners were more likely to accept canonical pronunciations of the new words if they had been paired with canonical spellings than if they had been learned with spellings matching the reduced form, or with no spelling information at all. These robust findings provide a clear answer to the question that motivated our study: Orthography plays a significant role in driving the canonical pronunciation advantage observed in previous research, consistent with

Ranbom and Connine's (2007) speculation. Our findings suggest that listeners built phonological representations from the words' printed forms, and that these representations guided their spoken word recognition.

Ironically, our clear demonstration of a mechanism that supports the canonical advantage comes in the context of a small flurry of papers that question its reality. These recent findings come mostly from Bürki, Viebahn and their colleagues on one hand, and Sumner and her colleagues on the other hand. We will describe each of these clusters of studies in turn and explain why we think that they merely place boundary conditions on a canonical advantage, rather than eliminating it as a phenomenon to explain.

Bürki, Viebahn and their colleagues' investigations of pronunciation variants processing have focused on the case of schwa deletion in French (Bürki, Spinelli, & Gaskell, 2012; Bürki, Viebahn, Racine, Mabut & Spinelli, 2018; Viebahn, McQueen, Ernestus, Frauenfelder, & Bürki, 2018). Some French words have two variant pronunciations: the word can either be produced with a schwa (e.g., *renard*, meaning *fox*, pronounced [ʁənaʁ]) or without a schwa (e.g., *r'nard* [ʁnaʁ]). The canonical form is the one with the schwa which is always represented in the orthography (written with "e"). In a recent paper, Bürki et al. (2018) claim that the canonical advantage observed in previous studies may simply be the result of exposure frequency. Across three experiments, listeners made lexical decisions about words in their canonical schwa form or their reduced no-schwa form. The words were either presented with a determiner (Experiments 1 and 2), in a casually pronounced sentence (Experiments 2 and 3) or in a carefully pronounced sentence (Experiment 3). Bürki et al. found that infrequent schwa variants were recognized faster than infrequent no-schwa variants, but that there was no significant difference in reaction times between frequent schwa and frequent no-schwa forms.

This result suggests that the advantage for the canonical form disappears when form frequency is taken into account. However, Bürki et al. note that the absence of an advantage for the canonical form was a null result and that it "[does] not provide evidence against an intrinsic advantage for canonical variants that would exist beyond and above the influence of variant frequency" (Bürki et al., 2018, p. 507). In fact, there are several factors that limit the conclusions that can be drawn from this study and others looking at the case of schwa words in French. The schwa case may not be ideal to explore the canonical advantage, especially if the analyses only focus on reaction times. The reduced form is necessarily shorter than the canonical form, building in a reaction time disadvantage for the canonical version. Bürki et al. (2018) tried to control for this inherent issue by including word duration as a covariate in all their reaction time analyses. Another concern is that most of the no-schwa word forms used in the experiments resulted in consonant clusters that are illegal in French (e.g., "fnêtre" for "fenêtre", "fn" is not a legal cluster). Bürki et al. (2018) themselves pointed out that there are indications that schwa deletion may impair word recognition when French phonotactics are respected, but not when the deletion leads to illegal consonant clusters (Spinelli & Gros-Balthazard, 2007). Spinelli and Gros-Balthazard suggested that listeners might be restoring the schwa automatically when they encounter an illegal cluster, which would explain the absence of reduction costs in this case. Even determining whether a word is, in fact, produced with or without a schwa is not trivial (Bürki, Fougeron, Gendrot, & Frauenfelder, 2011).

Another limitation of the French schwa case for examining a canonical advantage is that the canonical schwa form is in fact more frequent overall. This differs from the cases we considered (medial t-flap and medial nasal flap in English). What makes the canonical advantage intriguing is exactly the fact that uncommon forms show the advantage. Several studies have shown that listeners recognize frequent variants better than infrequent ones, with the only exception being the canonical form (e.g. Pitt et al., 2011; Ranbom & Connine, 2007). For the schwa case it is not surprising that when frequency is factored out, the advantage for the canonical form disappears, as the two co-vary in this case. Another potentially important difference between the French schwa deletion case and the cases we focused on is that schwa deletion is commonly signaled in written text by an apostrophe (e.g., *s'maine* instead of *semaine*). Given our results, if this orthographic convention is encountered frequently, it may encourage the mental representation of the no-schwa variant, negating a possible representational advantage for the canonical schwa case.

Finally, as we noted, measuring a canonical advantage with reaction times is potentially problematic because of the inherent duration differences between the schwa and no-schwa cases. The accuracy data from Bürki et al. (2018)'s study (and in other reports from this group) actually support an advantage for canonical forms: In all experiments, about 80% of the errors occurred for no-schwa variants – non-canonical forms produced about four times as many errors as canonical forms. Taking all these concerns into account, while Bürki, Viebahn and their colleagues' findings suggest that frequency of exposure is an important factor for variant processing (consistent with other studies, e.g., Bürki & Frauenfelder, 2012; Pitt et al., 2011; Ranbom & Connine, 2007), they do not undercut the existence of the canonical advantage.

Sumner (2013; Sumner et al., 2013) has proposed an interesting explanation for the canonical advantage. She suggested that the advantage depends on the phonetic context in which the variation occurs (i.e., the whole word). In experiments observing a canonical advantage, all stimuli, including the reduced forms, are usually recorded in a careful, well-articulated manner. Sumner argues that this creates a form of mismatch for the reduced forms, since they are normally produced in a context of casual, heavily reduced speech. There is evidence that subcategorical acoustic mismatches result in perceptual costs (Marslen-Wilson & Warren, 1994). In other words, the canonical advantage observed in previous studies may be artificially created by the incongruency between a reduced phoneme and the carefully pronounced phonetic context it is embedded in; in a sense, the idea is that there is a non-canonical disadvantage, rather than a canonical advantage (Sumner et al., 2013).

Sumner (2013) reported evidence that when this incongruency is eliminated, and reduced forms are presented in a casual phonetic context (i.e., several segments in the word are reduced), they are processed just as well as the canonical forms. They are also processed better than reduced forms embedded in a careful phonetic context. She concluded that in the case of a carefully pronounced reduced form, the bottom-up signal is clear but is not consistent with top-down information, slowing down recognition. In contrast, in the case of a casually pronounced reduced form, the signal is ambiguous, inducing a stronger influence from top down processing. However, under this view, it remains unclear what word forms are represented. In Sumner's (2013) experiment, both the casual reduced form and the canonical form were recognized equally well. Does this mean that both are represented in memory, or that top down influences during casual reduced form processing restored the reduced segments? Previous studies provide evidence that only the canonical form is represented (Sumner & Samuel, 2005; 2009). If that is the case, it is not clear why canonical forms would be stored at all in cases where they are rarely encountered in speech.

Sumner's phonetic context account of the canonical advantage makes the prediction that as this context becomes more predictable (e.g., all the speech input is in a casual style), its effect on pronunciation variants' processing should be more robust. Sumner et al. (2013, Experiment 2) found results consistent with this prediction in one experiment using the medial t-flap, blocking stimuli by speech style. However, in Bürki et al.'s (2018) study of French schwa and no-schwa variants in casually or carefully pronounced sentences, there was no interaction between context and variant type (canonical versus reduced).

Viebahn and Luce (2018) also investigated how phonetic context information (e.g. a casual speaking style) changes recognition of pronunciation variants. They embedded the canonical or reduced forms of medial /nt/ cluster words (e.g. *center*) in both carefully and casually pronounced sentences. Like Bürki et al. (2018), they found no interaction between context and variant type for accuracy (there was a significant interaction between the two for reaction times, but only for slow responses). Listeners' accuracy for reduced forms was only slightly higher when they were presented in casual speech style compared to careful speech style, and this difference was not significant. However, listeners recognized the canonical forms better than reduced forms in both contexts. Interestingly, the recognition of canonical forms was independent of context, with no facilitation from a careful speech style or hindrance from a casual one.

Other studies have investigated the role of semantic or syntactic context on the recognition of pronunciation variants (e.g., Ranbom & Connine, 2007; Tuinman, Mitterer, & Cutler, 2014; Van de Ven, Tucker, & Ernestus, 2011) and found that when reduced forms are presented in a biasing context, they produce effects equivalent to the canonical forms. However, some studies show that highly reduced forms are only well recognized when presented in a heavily biasing context (Ernestus, Baayen & Schreuder, 2002; Janse & Ernestus, 2011).

Taken together, the studies we have just discussed provide interesting preliminary evidence that contextual influences may play a role in the canonical advantage. However, these initial results are a long way from explaining the broad set of findings that demonstrate an advantage for canonical forms, let alone suggesting that there is no such advantage. The widely-reported canonical advantage is puzzling because it seems to violate frequency accounts of phonological variant processing, and we know that frequency matters (e.g., Bürki et al., 2018; Connine, 2004; Pitt et al. 2011; Ranbom & Connine, 2007). It may in part be explained if exposure to a word's written form causes listeners to generate the matching phonological form. If what constitutes exposure to a spoken word includes exposure to its written form as well, the relative frequencies of canonical and reduced forms could change substantially. In our experiments, the printed and spoken forms of words were presented an almost equal number of times. For example, in Experiment 1 there were 22 exposures to a word's spoken form and 20 exposures to its written form. In a recent study looking at the relationship between orthography and spoken word recognition, Viebahn et al. (2018) used a new-word learning paradigm similar to ours (examining the French schwa case), but with a very different spoken to written form ratio for each word. In their study, listeners saw orthographic forms consistent with a canonical pronunciation 15% of the time (8 exposures) but heard the reduced phonological forms 85% of the time (46 exposures). Under these conditions, Viebahn et al. found a small effect of orthography exposure on the likelihood that speakers produced a canonical form (that matches the orthography) rather than a variant one. However, they did not find an effect of orthography on spoken word recognition. The possible effect of spoken to written exposure ratios is a promising topic for future research.

The design of our study allowed us to look for generalization effects of orthography. In particular, the Canonical_None and Reduced_None groups experienced consistent spelling patterns for half of the words being learned, with no orthographic information for the other half. An important finding from this manipulation is that listeners generalized the sound-to-spelling mappings (e.g., [ɾ] is spelled "dd") from words learned with spellings to words learned without: Words learned with no spelling inherit the orthographic properties of other words that share phonetic features. For example, when half the words were learned with a spelling consistent with the canonical pronunciation and half with no spellings at all, listeners were not only more likely to accept the canonical pronunciation for words learned with a canonical spelling but also for the words that were learned without orthographic information. That listeners generalized from "with spelling" to "without spelling" words supports the idea that the orthographic representation generates a matching phonological representation.

The generalization effect implies that direct exposure to a word's spelling is not necessary to observe orthographic effects on spoken word recognition; listeners can infer how a word would be written from their language's orthographic patterns. In fact, we found a strong bias toward accepting the canonical form, even for words learned with a reduced spelling. This presumably reflects the long-term learning of our adult literate American English participants, who are used to the mapping between a reduced variant (e.g., a flap) and a canonical form spelling ("tt", for example).

Experiment 2 demonstrated that the effect of orthography was stable: The pattern found on the picture matching task two days after the words were learned was almost identical to what we saw on such a test conducted right after the learning phase. As we noted, this rules out the possibility that the orthographic effect had resulted from the persistence of simple episodic memory traces rather than from information stored in lexical representations. We found no evidence for sleep-based consolidation of the influence of orthographic information, perhaps because this type of cross-modal transfer takes more time than the two days between the sessions.

While time did not affect the orthographic effect, the consistency of the learning environment did: Training conditions with only one orthographic mapping led to a stronger canonical advantage than less consistent ones. For example, for participants in the Reduced_None group, whenever the allophone [ɾ] appeared in a word, it was represented with "dd" in the word's spelling, whereas for participants in the Canonical_Reduced group, it was represented as either "dd" *or* "tt", depending on the word. Averaging across all four picture matching tasks in the current study, participants in the Reduced_None groups correctly rejected the canonical pronunciation 80% of the time for words learned with the (reduced) "dd" spelling, compared to 68% for the Canonical_Reduced participants.

The sensitivity of the orthographic influence to the consistency of the mapping suggests that speakers of languages with different orthographic depth may put different weight on orthographic cues and thus exhibit different degrees of orthographic influence during spoken word recognition. In fact, listeners whose first language has a transparent orthography tend to be misled when the orthography does not match the phonology in a straightforward way, while listeners of deep-orthography languages may have a weaker connection between orthography and phonology (e.g., Erdener & Burnham, 2005). Consistent with this view, Viebahn et al. (2018) suggested that the degree of consistency of the mapping between graphemes and phonemes is an important factor influencing the orthography effect. Studies on second language acquisition seem to corroborate this speculation; L2 orthography may only help with L2 spoken word recognition when L1 and L2 phoneme-grapheme mappings are consistent (e.g., Escudero, Simon, & Mulak, 2014; Escudero & Wanrooij, 2010; Ota, Hartsuiker, & Haywood, 2010).

While questions remain about the locus of the orthography effect and under which conditions it may be blocked or enhanced, the experiments in the current study have provided a clear answer to the question that motivated them: Why do listeners persistently recognize the canonical variant of phonemes better despite their reduced variants being far more frequent? Our results show that this advantage is at least partly driven by orthography: Listeners accept pronunciations that are consistent with words' spellings, even if they are never heard.

Acknowledgments

This work was supported by the National Science Foundation grant IBSS-1519908. We also acknowledge support provided by Ministerio de Ciencia E Innovacion Grant # PSI2017-82563-P, by the Basque Government through the BERC 2018-2021 program, by Economic and Social Research Council (UL) Grant #ES/R006288/1 and by Ayuda Centro de Excelencia Severo Ochoa SEV-2015-0490. We thank Marie Huffman and Donna Kat for help with this project.

References

Bakker, I., Takashima, A., van Hell, J. G., Janzen, G., & McQueen, J. M. (2014). Competition from unseen or unheard novel words: Lexical consolidation across modalities. *Journal of Memory and Language*, *73*(1), 116–130. https://doi.org/10.1016/j.jml.2014.03.002

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.1126/science.1176170

Bürki, A., Fougeron, C., Gendrot, C., & Frauenfelder, U. H. (2011). Phonetic reduction versus phonological deletion of French schwa: Some methodological issues. *Journal of Phonetics*, *39*(3), 279–288. https://doi.org/10.1016/j.wocn.2010.07.003

Bürki, A., Spinelli, E., & Gareth Gaskell, M. (2012). A written word is worth a thousand spoken words: The influence of spelling on spoken-word production. *Journal of Memory and Language*, *67*(4), 449–467. https://doi.org/10.1016/j.jml.2012.08.001

Bürki, A., Viebahn, M. C., Racine, I., Mabut, C., & Spinelli, E. (2018). Intrinsic advantage for canonical forms in spoken word recognition: myth or reality? *Language, Cognition and Neuroscience, 33*(4), 494–511. https://doi.org/10.1080/23273798.2017.1388412

Chen, W. F., Chao, P. C., Chang, Y. N., Hsu, C. H., & Lee, C. Y. (2016). Effects of orthographic consistency and homophone density on Chinese spoken word recognition. *Brain and Langua*ge, *157–158,* 51–62. https://doi.org/10.1016/j.bandl.2016.04.005

Connine, C. M. (2004). It's not what you hear but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition, *Psychonomic Bulletin & Review*, *11*(6), 1084–1089. https://doi.org/10.3758/bf03196741

Connine, C. M., Titone, D., Deelman, T., & Blasko, D. (1997). Similarity Mapping in Spoken Word Recognition. *Journal of Memory and Language*, *37*(4), 463–480. https://doi.org/10.1006%2Fjmla.1997.2535

Coridun, S., Ernestus, M., & Ten Bosch, L. (2015). Learning pronunciation variants in a second language: Orthographic effects. In *18th International Congress of Phonetic Sciences (ICPhS 2015)*. Glasgow, Scotland: University of Glasgow.

Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning: neural and behavioural evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1536), 3773-3800. https://doi.org/10.1098%2Frstb.2009.0111

Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words, *Psychological Science*, *18*(1), 35–39. https://doi.org/10.1111%2Fj.1467-9280.2007.01845.x

Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics*, *67*(2), 224-238. https://doi.org/0.3758/bf03206487

Erdener, V. D., & Burnham, D. K. (2005). The role of audiovisual speech and orthographic information in nonnative speech production. *Language Learning*, *55*(2), 191–228. https://doi.org/10.1111/j.0023-8333.2005.00303.x

Ernestus, M., Baayen, H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language*, *81*(1–3), 162–173. https://doi.org/10.1006/brln.2001.2514

Escudero, P., Simon, E., & Mulak, K. E. (2014). Learning words in a new language: Orthography doesn't always help. Bilingualism: *Language and Cognition*, *17*(2), 384– 395. https://doi.org/10.1017/S1366728913000436

Escudero, P., & Wanrooij, K. (2010). The effect of L1 orthography on non-native vowel perception. *Language and Speech*, *53*(3), 343–365. https://doi.org/10.1177/0023830910371447

Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition*, *89*(2), 105–132. https://doi.org/10.1016/S0010-0277(03)00070-2

Gaskell, M. G., & Marslen-Wilson, W. D. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 22*(1), 144–158. https://doi.org/10.1080/01690969508407097

Gow, D. W. Jr. (2003). Feature parsing: Feature cue mapping in spoken word recognition. *Perception & Psychophysics*, *65*(4), 575–590. https://doi.org/10.3758%2Fbf03194584

Jakimik, J., Cole, R. A., & Rudnicky, A. I. (1985). Sound and Spelling in Spoken Word Recognition. *Journal of Memory and Language*, *24*(2), 165–178. https://doi.org/10.1037/0096-1523.3.1.51

Janse, E., & Ernestus, M. (2011). The roles of bottom-up and top-down information in the recognition of reduced speech: Evidence from listeners with normal and impaired hearing. *Journal of Phonetics*, *39*(3), 330-343. https://doi.org/10.1016/j.wocn.2011.03.005

Janse, E., Nooteboom, S. G., & Quené, H. (2007). Coping with gradient forms of /t/-deletion and lexical ambiguity in spoken word recognition. *Language and Cognitive Processes, 22* (2). https://doi.org/10.1080/01690960500371024

Johnson, K. (2004). Massive reduction in conversational American English. In K. Yoneyama & K. Maekawa (Eds.), *Proceedings of the first session of the 10th international symposium on spontaneous speech: Data and analysis* (pp. 29-54). Tokyo: The National International Institute for Japanese Language.

Kraljic, T., & Samuel, A. G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, *56*(1), 1-15. https://doi.org/10.1016/j.jml.2006.07.010

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

Lahiri, A., & Marslen-Wilson, W. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition, 38*(3), 245–294. https://doi.org/10.1016%2F0010-0277%2891%2990008-r

Leach, L., & Samuel, A. G. (2007). Lexical configuration and lexical engagement: When adults learn new words, *Cognitive Psychology, 55*(4), 306–353. https://doi.org/10.1016/j.cogpsych.2007.01.001.

Leinenger, M. (2014). Phonological coding during reading. *Psychological Bulletin, 140*(6), 1534–1555. https://doi.org/10.1037/a0037830.

Lerner, A. J., Loewe, F. (1956). Why can't the English? [Recorded by Harrison, R.]. On *My Fair Lady* [CD], New York, NY: Columbia Records.

Marslen-Wilson, W., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review*, *101*(4), 653–675. https://doi.org/10.1037/0033-295x.101.4.653

McGowan, K. B., & Sumner, M. (2014). The effect of contextual mismatches on lexical activation of phonetic variants. *The Journal of the Acoustical Society of America*, *135*(4), 2199-2199.

Mitterer, H., & Reinisch, E. (2015). Letters don't matter: No effect of orthography on the perception of conversational speech. *Journal of Memory and Language, 85*, 116–134. https://doi.org/10.1016/j.jml.2015.08.005

Morais, J., Cary, L., Alegria, J., & Bettelson, P. (1979). Does awareness of speech as a sequence of phonemes arise spontaneously? *Cognition*, *7*(4), 323–331. https://doi.org/10.1016%2F0010-0277%2879%2990020-9

Ota, M., Hartsuiker, R. J., & Haywood, S. L. (2010). Is a FAN always FUN? Phonological and orthographic effects in bilingual visual word recognition. *Language and Speech*, *53*(3), 383–403. https://doi.org/10.1177/0023830910371462

Patterson, D., & Connine, C. M. (2001). Variant frequency in flap production. *Phonetica*, *58*(4), 254-275. https://doi.org/10.1159%2F000046178

Perre, L., & Ziegler, J. C. (2008). On-line activation of orthography in spoken word recognition. *Brain Research*, *1188*(1), 132–138. https://doi.org/10.1016/j.brainres.2007.10.084

Pitt, M. A., Johnson, K., Hume, E., Kiesling, S., & Raymond, W. (2003). The ViC corpus of conversational speech. Manuscript submitted to *IEEE Transactions on Speech and Audio Processing: Special Issue on Spontaneous Speech Processing.*

Pitt, M. A. (2009a). How are pronunciation variants of spoken words recognized? A test of generalization to newly learned words, *Journal of Memory and Language, 61*(1), 19–36. https://doi.org/10.1016/j.jml.2009.02.005.How

Pitt, M. A. (2009b). The strength and time course of lexical activation of pronunciation variants. Journal of Experimental Psychology: *Human Perception and Performance, 35*(3), 896–910. https://doi.org/10.1037/a0013160.

Pitt, M. A., Dilley, L., & Tat, M. (2011). Exploring the role of exposure frequency in recognizing pronunciation variants. *Journal of Phonetics, 39*(3), 304–311. https://doi.org/10.1016/j.wocn.2010.07.004

R Core Team (2015). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Racine, I., Bürki, A., & Spinelli, E. (2014). The implication of spelling and frequency in the recognition of phonological variants: Evidence from pre-readers and readers. *Language, Cognition and Neuroscience, 29*(7), 893–898. https://doi.org/10.1080/01690965.2013.832784

Racine, I., & Grosjean, F. (2005). Le coût de l'effacement du schwa lors de la reconnaissance des mots en français. *Revue Canadienne de Psychologie Expérimentale, 59*(4), 240–254. https://doi.org/10.1037/h0088052

Ranbom, L. J., & Connine, C. M. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language*, *57*(2), 273–298. https://doi.org/10.1016/j.jml.2007.04.001

Salverda, A. P., & Tanenhaus, M. K. (2010). Tracking the time course of orthographic information in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(5), 1108–1117. https://doi.org/10.1037/a0019901.

Samuel, A. G., & Larraza, S. (2015). Does listening to non-native speech impair speech perception? *Journal of Memory and Language, 81*, 51–71. https://doi.org/10.1016/j.jml.2015.01.003

Seidenberg, M. S., & Tanenhaus, M. K. (1979). Orthographic effects on rhyme monitoring. *Journal of Experimental Psychology: Human Learning and Memory, 5*(6), 546–554. https://doi.org/10.1037/0278-7393.5.6.546

Stone, G. O., Vanhoy, M., & VanOrden, G. C. (1997). Perception is a two way street: Feedforward and feedback phonology in visual word recognition. *Journal of Memory and Language, 36*(3), 337–359. https://doi.org/10.1006/jmla.1996.2487

Sulpizio, S., & McQueen, J. M. (2011, August). When two newly-acquired words are one: New words differing in stress alone are not automatically represented differently. In *Proceedings of the Annual Conference of the International Speech Communication Association*, INTERSPEECH, 1385–1388.

Sulpizio, S., & McQueen, J. M. (2012). Italians use abstract knowledge about lexical stress during spoken-word recognition. *Journal of Memory and Language*, *66*(1), 177–193. https://doi.org/10.1016/j.jml.2011.08.001

Sumner, M. (2013). A phonetic explanation of pronunciation variant effects. *The Journal of the Acoustical Society of America*, *134*(1), 26-32. https://doi.org/10.1121/1.4807432

Sumner, M., Kurumada, C., Gafter, R., & Casillas, M. (2013, January). Phonetic variation and the recognition of words with pronunciation variants. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 35, No. 35).

Sumner, M., & Samuel, A. G. (2005). Perception and representation of regular variation: The case of final /t/. *Journal of Memory and Language, 52*(3), 330–346. https://doi.org/10.1016/j.jml.2004.11.004

Sumner, M., & Samuel, A. G. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language, 60*(4), 487–501. https://doi.org/10.1016/j.jml.2009.01.001

Taft, M. (2006). Orthographically Influenced Abstract Phonological Representation: Evidence from Non-rhotic Speakers. Journal of Psycholinguistic Research, 35(1), 67–78. doi:10.1007/s10936-005-9004-5

Taft, M., & Hambly, G. (1985). The influence of orthography on phonological representations in the lexicon. *Journal of Memory and Language*, *24*(3), 320–335. doi:10.1016/0749-596x(85)90031-2

Tuinman, A., Mitterer, H., & Cutler, A. (2013). Use of Syntax in Perceptual Compensation for Phonological Reduction. *Language and Speech, 57*(1), 68–85. https://doi.org/10.1177/0023830913479106

Tucker, B. V, & Warner, N. (2007, August). Inhibition of Processing Due to Reduction of the American English Flap. In *Proceedings of the 16th International Congress of Phonetic Sciences*, 1949–1952.

Van de Ven, M., Tucker, B. V., & Ernestus, M. (2011). Semantic context effects in the comprehension of reduced pronunciation variants. *Memory & Cognition*, *39*(7), 1301-1316. https://doi.org/10.3758/s13421-011-0103-2

Viebahn, M. C., & Luce, P. A. (2018). Increased exposure and phonetic context help listeners recognize allophonic variants. *Attention, Perception, & Psychophysics*, *80*(6), 1539–1558. https://doi.org/10.3758/s13414-018-1525-8

Viebahn, M. C., McQueen, J. M., Ernestus, M., Frauenfelder, U. H., & Bürki, A. (2018). How much does orthography influence the processing of reduced word forms? Evidence from novel-word learning about French schwa deletion. *Quarterly Journal of Experimental Psychology, 71*(11). 2378–2394. https://doi.org/10.1177/1747021817741859

Ventura, P., Morais, J., Pattamadilok, C., & Kolinsky, R. (2004). The locus of the orthographic consistency effect in auditory word recognition. *Language and Cognitive Processes*, *19*(1), 57–95. https://doi.org/10.1080/01690960344000134

Ziegler, J. C., & Ferrand, L. (1998). Orthography shapes the perception of speech: The consistency effect in auditory word recognition. *Psychonomic Bulletin & Review, 5*(4), 683–689. https://doi.org/10.3758/BF03208845

Appendix A

Intervocalic t new words		Intervocalic nt cluster new words	
Canonical spelling	Reduced spelling	Canonical Spelling	Reduced Spelling
Bytto	Byddo	Trenter	Trenner
Trotty	Troddy	Glunter	Glunner
Cluttee	Cluddee	Sento	Senno
Blittus	Bliddus	Prounter	Prounner
Ruttow	Ruddow	Smountus	Smounnus
Fluttow	Fluddow	Trounty	Trounny
Voutty	Vouddy	Clunter	Clunner
Slettow	Sleddow	Gleanty	Gleanny
Slyttenee	Slyddenee	Sounterby	Sounnerby
Pittery	Piddery	Cyntergus	Cynnergus
Trettemus	Treddemus	Vintimal	Vinnimal
Vettotus	Veddotus	Spentogy	Spennogy
Pribetty	Pribeddy	Bredento	Bredenno
Trecottus	Trecoddus	Credenter	credenner
Spearottus	Spearoddus	Suspintus	Suspinnus
Sceanetty	Sceaneddy	Lucrunty	Lucrunny

List of the new words created for the study. 16 words had an intervocalic /t/ consonant and 16 had an intervocalic /nt/ cluster. For each type, a canonical spelling and a reduced spelling were created.

Appendix B

Reaction time analysis for Experiment 1

As noted in the main text, the current study was designed to examine accuracy differences, and several features of the procedures prevent clear interpretation of response times. Nonetheless, some readers may be interested in reaction times, so we report them here (see Table 1). Reaction times were analyzed using linear mixed effect models. When necessary, denominator degrees of freedom and p-values for F-tests were computed based on Satterwaite's approximations with the Rpackage *lmerTest* (Kuznetsova, Brockhoff, & Christensen, 2014). Impossibly short RTs (< 250ms) were excluded from the data (< 1% of the data), and the experiment was set up so that RTs could not be longer than 5000ms. Following Baayen and colleagues' suggestions (Baayen, 2008; Baayen & Milin, 2010), we considered model residuals larger than 2.5 standard deviations above their mean to be outliers and removed them, independently for each model. The percentage of removed outliers was 2.5%.

The analysis showed that the main effect of spelling on the reaction times for the canonical and reduced pronunciation trials was not significant ($\chi^2 = 4.42$, $df = 2$, $p = 0.11$). However, there was a significant difference in reaction times between the critical, canonical pronunciation trials and the control, reduced pronunciation trials. Participants were faster to respond on reduced pronunciation trials than canonical pronunciation trials ($\beta = -89.8982$, $SE = 17.7471$, $df = 60.7499$, $t = -5.07$, $p < 0.0001$). But, as noted in the text, canonical pronunciations of the new words tended to be longer than their reduced pronunciation counterparts. And in fact, word duration was a significant factor influencing reaction times during these critical and control trials, with slower reaction times the longer a word was ($\beta = 0.5326$, $SE = 0.0512$, $df = 60.7251$, $t = 10.41$, $p < 0.0001$).

	Canonical Pronunciation trials		Reduced Pronunciation trials	
Spelling learned	Mean RT (ms)	% correct	Mean RT (ms)	% correct
Canonical Spelling	985	38	840	94
Reduced Spelling	979	71	849	97
No Spelling	965	63	863	95

Table 1. Mean response times and accuracy on canonical pronunciation trials and reduced pronunciation trials of the Picture Matching task as a function of spelling learned during the Learning task in Experiment 1.

The results can be divided into “acceptance RTs” for the canonical pronunciations of the new words versus the “acceptance RTs” for the reduced pronunciations. On average, participants were slower to erroneously accept the canonical pronunciations (M = 969ms, N = 817) than to correctly accept their reduced pronunciation (M = 840ms, N = 1833). By comparison, the average RT to correctly reject canonical pronunciations was 981ms (N = 1099), compared to 1083ms (N = 87) to erroneously reject reduced pronunciations. These results fit the common pattern of error responses and “no” responses being slower than correct responses. It is difficult to compare “acceptance RTs” for reduced and canonical forms in a meaningful way because canonical forms were inherently longer than the reduced ones.

Appendix C

Reaction times analysis for Experiment 2

Reaction time analyses were conducted as for Experiment 1 (see Appendix B – the percentage of removed outlier was 2.7%), with the factor Picture Matching Task added in order to examine any changes in reaction times across the three separate tasks of Experiment 2. Unlike in Experiment 1, there was a main effect of spelling on reaction times (χ^2 = 16.3, *df* = 2. *p* = 0.00029). Overall, responses for words learned with a reduced spelling were faster than for words learned without a spelling (*β* = -26.2199, *SE* = 7.8535, *df* = 1490.7472, *t* = -3.34, *p* =0.00086) or but not than words learned with a canonical spelling (*β* = -7.4092, *SE* = 7.8456, *df* = 1498.8713, *t* = -0.94, *p* = 0.345). As in Experiment 1, participants responded more quickly on reduced pronunciation trials (*β* = - 75.9346, *SE*= 16.1903, *df* = 60.8260, *t* = -4.69, *p* < 0.001). Again, word duration was a significant factor, with longer reaction times the longer a word was (*β* = 0.5091, *SE* = 0.0467, *df*= 60.8907, *t* = 10.90, *p* < 0.001). In addition, reaction times changed across the three different picture matching tasks of Experiment 2. During the second picture matching task, which took place two days after the first one, participants were slower to answer (*β* = 35.3815, *SE* = 12.3222, *df* = 58.6273, *t* =2.87, *p* = 0.00568). On the third picture matching task, which took place after some additional training, participants were now faster to answer than during the first (*β* = -39.7171, *SE* = 17.2789, *df* = 74.3692, *t* = -2.30, *p* = 0.02434) and second tasks (*β* = -75.0986, *SE* = 14.2120, *df* = 75.7852, *t* = -5.28, *p* < 0.001).

As for Experiment 1, we compared “acceptance RTs” for the canonical pronunciations of the new words to the “acceptance RTs” for the reduced pronunciations. On average, across all picture matching tasks, participants were slower to erroneously accept the canonical pronunciations (*M* = 1025 ms, *N* = 2312) than to correctly accept the reduced pronunciations (*M* = 880 ms, *N* = 5382).

Pronunciation	Spelling learned	Session 1 Mean RT (ms)	Session 2 Mean RT (ms)	Session 3 Mean RT (ms)
Canonical	Canonical	1081	1082	983
	Reduced	983	1041	978
	None	1008	1034	905
Reduced	Canonical	916	962	892
	Reduced	865	893	850
	None	865	927	848

Table 2. Mean reaction times for the canonical and reduced pronunciation of newly learned words during the three Picture Matching tasks of Experiment 2. Reaction times are reported in ms.

Appendix D

Accuracy and Reaction times analyses of the learning tasks for Experiments 1 and 2

In addition to the means and figures presented in the main text, interested readers can find the statistical analyses of the learning tasks' results here (see also Figure 10). We analyzed accuracy with generalized linear mixed models, and RTs with linear mixed models. All models included by-subject and by-item random intercepts. Fixed factors were Group and Repetition. The Repetition predictor was standardized for accuracy analysis. RTs below 250ms were excluded from all analyses. In addition, for the RTs analyses, model residuals larger than 2.5 standard deviations above their mean were removed (the percentages ranged from 2% to 2.9%). RTs were not transformed.

In Experiment 1, there was no main effect of Group on accuracy during the learning task (χ^2 = 2.89, *df* =2. *p* = 0.24), but accuracy did increase significantly with number of repetitions (β = 0.6989, *SE* = 0.01797, *Wald's z* = 39, $p < 0.001$). The average RT decreased from 1018ms on the first block of the learning task, to 820ms on the second block and to 793ms on the last block. There was no main effect of Group on the RTs (χ^2 = 0.53, *df* = 2, *p* = 0.77), but the effect of repetition was significant such that participants answered faster the more repetitions of a word-picture pair they encountered (β = -13.937, *SE*=0.218, *df* = 36255.832, *t* = -63.9, $p < 0.001$).

In Experiment 2 there were two learning tasks, one at the beginning of Session 1 and the second at the beginning of Session 3. The first learning task had 20 repetitions of each word, while the second one only had 10. There was no main effect of Group on accuracy in either the first (χ^2 = 0.7, *df* = 2, *p* = 0.71) or the second learning task (χ^2 = 0.65, *df* = 2, *p* = 0.72). In addition, while accuracy increased with repetitions during the first learning task similarly to Experiment 1 (β = 0.7083, *SE* = 0.0181, *Wald's z* = 39.2, $p < 0.001$), it did not increase during the second task (β = 0.0452, *SE* = 0.0322, W*ald's z* = 1.4, $p = 0.16$). This suggests that by the second learning task, participants had reached a ceiling in their performance.

The RTs followed a similar pattern during the first learning task: There was no main effect of Group (χ^2 = 1.08, *df* = 2, $p = 0.58$) and the average RT decreased significantly from 1111ms in the first block to 923ms in the second block and finally to 915ms in the final one (β = -13.732, *SE* = 0.237, *df*= 36606.689, *t* = -58.0, p <0.001). During the second learning task (at the beginning of Session 3) there was no main effect of Group on RT either (χ^2 = 2.36, *df* = 2, $p = 0.31$). Average RT during this task was 922ms, and RTs increased slowly with repetitions (β =1.287, *SE* = 0.604, *df* = 18701.847, *t* = 2.13, $p = 0.033$). Again, this suggests that participants had reached ceiling by this second task.

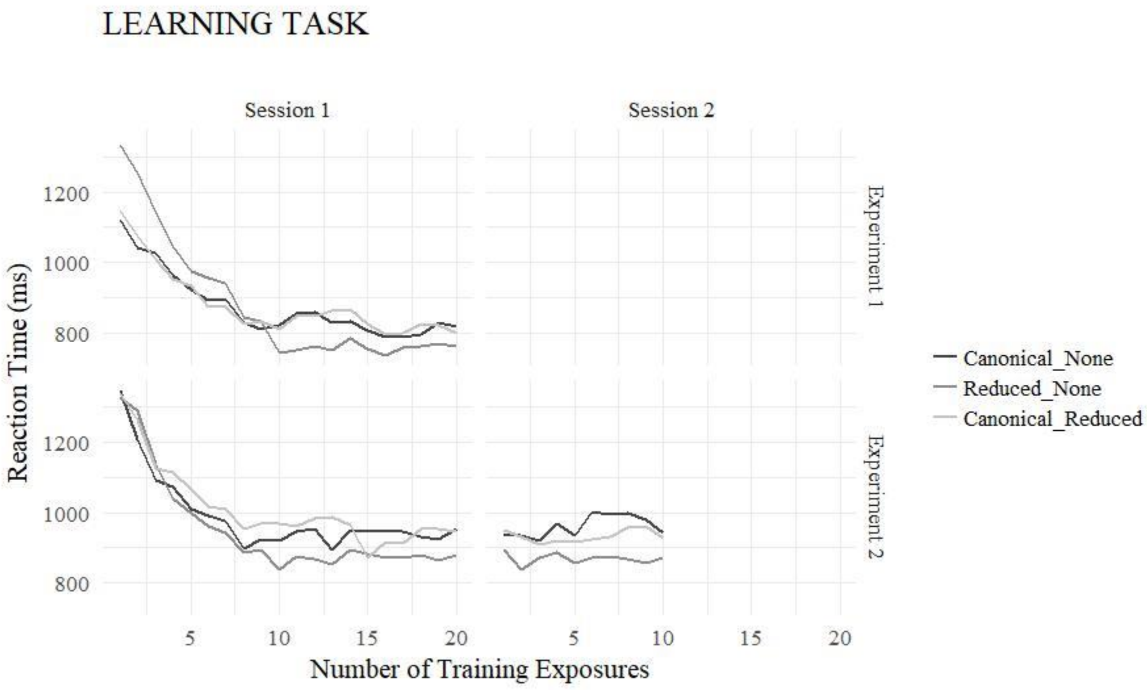


Figure 6. Reaction times during the learning tasks (includes RTs for both incorrect and correct answers) as a function of the number of exposures to each new word participants had to learn. Although the Reduced_None group appears to be faster than the other two, this difference was not significant.

Appendix E

Cued recall spelling task

Experiment 2 included a cued-recall spelling task to determine how well the participants had memorized the spellings of the new words. Participants in the Canonical_None and Reduced_None groups were exposed to 16 different word spellings, while participants in the Canonical_Reduced group were exposed to 32. As a result, the cued recall spelling task was more challenging for the latter. Participants performed better for canonical spellings than for reduced ones. The Canonical_None group correctly recalled 63% of the canonical spellings, and the Canonical_Reduced group recalled 53%. In comparison, the Reduced_None group correctly recalled 50% of the reduced spellings, and the Canonical_Reduced group 30%. The higher performance for the canonical spellings ("tt" or "nt") was significant ($\beta = 1.434$, *SE* = 0.429, *Wald's z* = 3.34, $p < 0.001$) reflecting the fact that they follow the most typical spelling patterns in English.

Within each group, we used Kendall's tau correlations to see if there was a relationship between performance on the cued-recall task and accuracy for the words learned with a spelling in the last picture matching task: Did participants who memorized the spellings better show stronger orthography effects?

For participants in the Reduced_None group (middle panel of Figure 11), performance on the last picture matching task of Experiment 2 was positively correlated with performance on the cued recall spelling task ($\tau_b(18) = 0.49$, $p = 0.004$): Participants who learned the reduced spellings better were also better at rejecting the canonical pronunciations during the last picture naming task. The relationship between performance on the two tasks was also positive for the Canonical_None group (left panel) ($\tau_b(18) = 0.41$, $p = 0.02$). For the group given twice as many spellings, with different spellings for half of the items (Canonical_Reduced), performance on the cued recall spelling task was not significantly correlated with performance on the picture matching task (words learned with a canonical spelling $\tau_b(18) = 0.074$, $p = 0.70$, and those with a reduced spelling $\tau_b(18) = 0.19$, $p = 0.30$). The weaker correlations for the Canonical_Reduced group reflect the greater difficulty of learning twice as many spelling (compared to the other two groups).

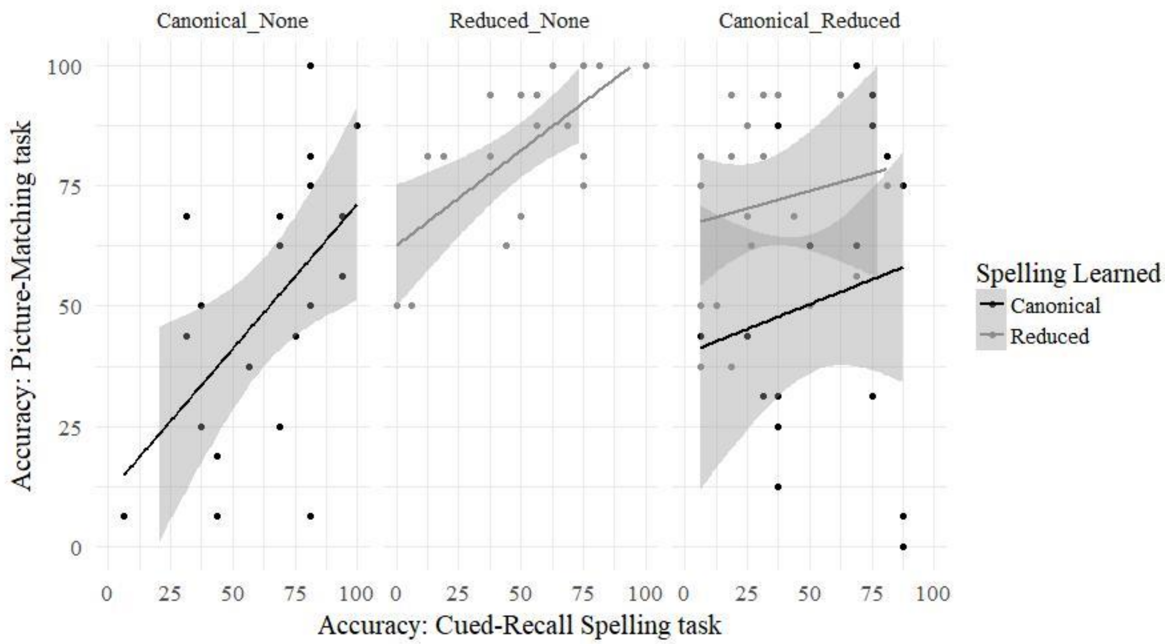


Figure 11. Correlation between accuracy on the final picture matching task of Experiment 2 and accuracy on the cued-recall spelling task that followed it.

Appendix F

Models summaries – Experiment 1 and Experiment 2

Comparing accuracy for words learned with a canonical spelling, a reduced spelling or no spelling across all groups.

	Experiment 1		Experiment 2 - replication	
	Parameter (SD)	p	Parameters (SD)	p
Subject(Intercept)	0.937 (0.968)		0.844 (0.919)	
Items (Intercept)	1.193 (1.092)		0.842(0.918)	
Intercept	-0.509	0.085	-0.391	0.14
Spelling None	1.260	<0.001	0.973	<0.001
Spelling Reduced	1.594	<0.001	1.338	<0.001

Table 3. Model specifications for the main effect of Spelling in Experiment 1 and Session 1 of Experiment 2. Session 1 of Experiment 2 is a direct replication of Experiment 1. The models included by-subject and by-item random intercepts. Spelling was entered as a fixed effect.

	Experiment 2 - consolidation		Experiment 2 – additional exposure	
	Parameter (SD)	p	Parameters (SD)	p
Subject(Intercept)	1.269 (1.13)		1.846 (1.359)	
Items (Intercept)	0.641 (0.80)		1.192 (1.092)	
Intercept	0.017	0.948	0.185	0.570
Spelling None	0.462	0.020	0.738	0.002
Spelling Reduced	0.813	0.007	1.498	<0.0001

Table 4. Model specifications for the main effect of Spelling in Session 2 and Session 3 of Experiment 2. Session 2 tested whether the effect of spelling changed after two days. Session 3 tested whether it changed after some additional exposure to the newly learned words and their spellings. The models included by-subject and by-item random intercepts. Spelling was entered as a fixed effect.

Comparing accuracy for words learned with a reduced spelling versus a canonical spelling within the Canonical_Reduced group (standard deviation in parenthesis)

	Experiment 1		Experiment 2 - replication	
	Parameter (SD)	p	Parameters (SD)	p
Subject (Intercept)	0.856 (0.925)		1.095(1.046)	
Subject (Slope)	-.0543 (0.737)		0.622(0.788)	
Item (Intercept)	1.058 (1.028)		0.604(0.777)	
Intercept	-0.228	0.522	-0.330	0.319
Spelling	1.076	0.016	1.029	0.007

Table 5. Model specifications for the main effect of Spelling within the Canonical_Reduced group for Experiment 1 and Session 1 of Experiment 2. Session 1 of Experiment 2 is a direct replication of Experiment 1. The models included by-subject and by-item random intercepts and a by-subject random slope. Spelling was entered as a fixed effect.

	Experiment 2 - consolidation		Experiment 2 – additional exposure	
	Parameter (SD)	p	Parameters (SD)	p
Subject (Intercept)	1.468 (1.212)		2.690 (1.640)	
Subject (Slope)	NA[1]		0.563 (0.750)	
Item (Intercept)	0.588 (0.767)		0.754 (0.869)	
Intercept	-0.0388	0.913	0.101	0.822
Spelling	0.6557	0.047	1.106	0.007

Table 6 Model specifications for the main effect of Spelling within the Canonical_Reduced group for Session 2 and Session 3 of Experiment 2. Session 2 tested whether the effect of spelling changes after two days. Session 3 tested whether it changed after some additional exposure to the newly learned words and their spelling. The models included by-subject and by-item random intercepts and a by-subject random slope. Spelling was entered as a fixed effect.

[1] For this data, the model did not converge with a random slope

Comparing accuracy for words learned without spellings in the Canonical_None and Reduced_None groups.

	Experiment 1		Experiment 2 replication	
	Parameter (SD)	p	Parameter (SD)	p
Subject	0.468 (0.684)		0.384(0.620)	
Item (Intercept)	1.476 (1.215)		0.993(0.997)	
Item (Slope)	0.666 (0.816)		0.919(0.959)	
Intercept	0.164	0.653	0.308(0.313)	0.33
Group	1.239	0.023	0.714(0.501)	0.15

Table 7 Model specifications for the effect of Group (Canonical_None vs. Reduced_None) on accuracy for words learned without any spellings. This table includes specifications for Experiment 1 and Session 1 of Experiment 2. Session 1 of Experiment 2 is a direct replication of Experiment 1. The models included by-subject and by-item random intercepts and a by-item random slope. Group was entered as a fixed effect.

	Experiment 2 - consolidation		Experiment 2 – additional exposure	
	Parameter (SD)	p	Parameter (SD)	p
Subject	0.848 (0.921)		2.158 (1.469)	
Item (Intercept)	0.751 (0.866)		1.062 (1.030)	
Item (Slope)	0.604 (0.777)		2.346 (1.532)	
Intercept	0.220	0.499	0.691	0.119
Group	0.633	0.178	1.193	0.134

Table 8. Model specifications for the effect of Group (Canonical_None vs. Reduced_None) on accuracy for words learned without any spellings. This table includes specifications for Session 2 and Session 3 of Experiment 2. Session 2 tested whether effects changed after two days. Session 3 tested whether they changed after some additional exposure to the newly learned words and their spelling. The models included by-subject and by-item intercepts and a by-item random slope. Group was entered as a fixed effect.

Comparing accuracy for words learned with spellings in the Canonical_None, Reduced_None and Canonical_None groups.

	Experiment 1		Experiment 2 replication	
	Parameter (SD)	p	Parameter (SD)	p
Subject (Intercept)	0.897 (0.947)		0.965(0.982)	
Item (Intercept)	1.040 (1.020)		0.660(0.813)	
Intercept	-1.708 (0.361)	0.003	-0.441(0.327)	0.18
Learning	0.853	0.017	0.100	0.78
Spelling	2.779	<0.0001	1.584	0.001
Learning * Spelling	- 1.751	<0.0001	-0.554	0.18

Table 9. Model specifications for the effect of Learning condition (i.e., simple for the Canonical_None and Reduced_None groups, vs. complex for the Canonical_Reduced group) and its interaction with the effect of Spelling. This table includes specifications for Experiment 1 and Session 1 of Experiment 2. Session 1 of Experiment 2 is a direct replication of Experiment 1. The models included by-subject and by-item random intercepts. Learning, Spelling and their interaction term were entered as fixed factors.

	Experiment 2 - consolidation		Experiment 2 – additional exposure	
	Parameter (SD)	p	Parameter (SD)	p
Subject (Intercept)	1.111 (1.054)		1.906 (1.381)	
Item (Intercept)	0.617 (0.7855)		0.952 (0.976)	
Intercept	-0.319	0.338	-0.070	
Learning	0.272	0.474	0.154	0.747
Spelling	1.77	0.0003	2.388	<0.0001
Learning * Spelling	-1.053	0.014	-1.095	0.042

Table 10. Model specifications for the effect of Learning condition (i.e., simple for the Canonical_None and Reduced_None groups, vs. complex for the Canonical_Reduced group) and its interaction with the effect of Spelling. This table includes specifications for Session 2 and Session 3 of Experiment 2. Session 2 tested whether effects changed after two days. Session 3 tested whether they changed after some additional exposure to the newly learned words and their spellings. The models included by-subject and by-item random intercepts. Learning, Spelling and their interaction term were entered as fixed factors.