Language, Cognition and Neuroscience, 2014
Vol. 00, No. 00, 1–11, http://dx.doi.org/10.1080/23273798.2014.952315

Routledge
Taylor & Francis Group

Influences on spelling: evidence from homophones

Rebecca Treiman[a]*, Mark S. Seidenberg[b] and Brett Kessler[c]

[a]*Washington University in St. Louis, Campus Box 1125, One Brookings Drive, St. Louis, MO 63130-4899, USA;* [b]*University of Wisconsin-Madison, WJ Brogden Hall, 1202 West Johnson Street, Madison, WI 53706-1696, USA;* [c]*Washington University in St. Louis, Campus Box 1125, One Brookings Drive, St. Louis, MO 63130-4899, USA*

(*Received 21 January 2014; accepted 26 July 2014*)

Three experiments used homophones as a test case to examine the roles of phonology and morphology in the spelling process. We introduced university students to novel meanings of spoken forms, for example, presenting /fid/ as a rare word for a type of furniture. We asked whether participants avoided spelling the new word as ‹feed›, instead using alternatives such as ‹fead›. Although participants produced some alternative spellings, they used spellings that resulted in homophones, i.e. spelled /i/ as ‹ee›, more often in items like /fid/ than in control items like /fip/. Participants were more likely to use novel spellings for homophones when given a choice between a novel spelling and an alternative than when asked to produce their own spellings. A major influence on spelling production thus appears to be the lesser effort that is required to use a familiar whole-word orthographic form compared to that needed for assembling a novel spelling.

Keywords: spelling; homophones; morphology; morphemes; phonemes

The spelling of a word in an alphabetic writing system reflects the word's phonological form. When trying to spell an unfamiliar word, people can often construct a plausible spelling based on the phonemes that the word contains. Much of the relatively small body of research that has been done on spelling has focused on how people construct spellings from phonological forms. For example, researchers have examined how people spell phonemes that have more than one possible spelling, asking about the factors that influence their choices (e.g. Barry & Seymour, 1988; Treiman, Kessler, & Bick, 2002).

A consideration of the nature of English and some other alphabetic writing systems suggests that people may be sensitive to morphology and not just phonology when selecting among spelling alternatives (e.g. Chomsky & Halle, 1968). Consider homophones: morphemes that have the same string of phonemes but different meanings. In some cases, as with the /bæt/ that refers either to a flying animal or to a stick that is used to hit a ball, the two morphemes are spelled alike. The fact that the same spelling is used for a given phonological form, regardless of its meaning, is consistent with the idea that spelling represents phonology. In other cases, as with *site* and *sight*, each meaning of a homophone has its own spelling. Although patterns at the level of phonology become more complex when a phoneme is spelled differently in the members of a homophone pair, the regularity of the writing system at the level of morphemes increases, in that each morpheme would have one and only one spelling. Many English homophones are spelled alike, but there are enough homophones with different spellings to suggest that morphology plays a role in the structure of English writing system and that this characteristic benefits its users (Venezky, 1999). Indeed, in the historical development of English, it has been argued that spellers sometimes used different spellings for homophones intentionally, in order to show the distinction in meaning between them (Vachek, 1971).

In the experiments reported here, we asked whether present-day spellers of English would elect to spell novel homophones differently. If so, this would suggest that they are sensitive to morphology and not just phonology. The only previous study to have touched on this issue is that of Baker (1980), who asked English-speaking university students whether various words were spelled in a rational manner. Participants who judged a spelling to be less than perfectly rational were asked to provide a different and more rational spelling. When participants stated that a spelling of a homophone should be changed, they showed some tendency to avoid spellings that mapped onto other words. Baker interpreted this finding to suggest that English-speaking adults prefer to spell homophones differently. However, a participant may have suggested that ‹rain› should be spelled as ‹rane› rather than ‹reign› because the mapping between /e/ and ‹a› followed by final ‹a› is more common than the mapping between /e/ and ‹eig›, not because ‹reign› spells another word. Baker's study did not control items to rule out such an explanation.

We designed the present experiments to provide more conclusive evidence on whether university students spell homophones differently from existing words and, in so

*Corresponding author. Email: rtreiman@wustl.edu

doing, to elucidate the roles of phonology and morphology in spelling. We did so by introducing participants to new words that were homophones of known words. Experiments 1 and 2 used the same set of items in spelling production tasks. For example, a participant might be told that /'wɪntɚ/ is a rare English word that means 'a person who is eager to learn the latest news and gossip'. This novel morpheme is homophonous with /'wɪntɚ/ the season, but its meaning is quite different. As another example, a participant might be told that /'fɪkəl/ means 'scorn, derision, or contempt'. We asked whether participants would spell these novel items differently from the familiar homophones, perhaps as ‹whinter› and ‹phickle›. Control items such as /'fɪntɚ/ and /'wɪkəl/ were included in order to provide a baseline measure of participants' tendency to use ‹w›, ‹f›, ‹inter› and ‹ickle› to spell /w/, /f/, /'ɪntɚ/ and /'ɪkəl/, respectively. This confers a degree of control that was not present in the study by Baker (1980). If participants prefer to use different spellings for the two meanings of a homophone, as Baker suggested, then they should be less likely to spell the homophone items /'wɪntɚ/ and /'fɪkəl/ as ‹winter› and ‹fickle› than to spell the control items /'fɪntɚ/ and /'wɪkəl/ as ‹finter› and ‹wickle›. Participants should produce some spellings of control items that use alternative letter sequences, such as ‹wiccal› for /'wɪkəl/, but spellings that contain plausible alternatives such as these should be more frequent for homophone items than control items. In Experiment 1, the homophone and control items were intermixed with real words. We told participants that all of the items they were asked to spell were English words but that some of them were very rare.

We assigned half of the participants in Experiment 1 to a condition in which they were asked to explain why they spelled each homophone and control item the way they did. If morphology plays an important role in the spelling process, participants in this self-report condition should often say that they tried to avoid producing the same spellings for words with different meanings. The other half of the participants were assigned to a condition in which they were not asked to report on why they spelled words the way they did. By comparing the spellings produced in the two conditions, we can determine whether increasing the metalinguistic focus by adding self-reports changes the nature of the spellings that people produce. As another between-subjects manipulation, half of the participants were asked to construct a sentence exemplifying the meaning of each item before producing a spelling and the other half were not. Constructing a sentence, we reasoned, would help to ensure that participants fully processed the new definitions. It might increase their tendency to spell the novel homophone differently from its familiar mate.

Experiment 1

Method

Stimuli

The experimental items were designed in groups of four. A list of all 14 quadruplets appears in Appendix 1, and one sample quadruplet is shown in Table 1. The items in each quadruplet had one of two onsets (initial consonants or consonant clusters). In the sample quadruplet shown in Table 1, onset 1 was /w/ and onset 2 was /f/. There were two possible remainders (vowels and following segments). These both started with same vowel, and some were monosyllabic and others disyllabic. In the example quadruplet of Table 1, remainder 1 was /'ɪntɚ/ and remainder 2 was /'ɪkəl/. The combination of onset 1 and remainder 1 yielded a word, here /'wɪntɚ/ ‹winter›. The combination of onset 2 and remainder 2 also yielded a word, /'fɪkəl/ ‹fickle›. Neither of the other combinations formed a word. (One item, /sɛk/, proved to be an uncommon word known by some participants, but exclusion of this item did not alter the pattern of results.) The phonemes in the onsets and remainders had several potential spellings. For example, /w/ could be spelled as ‹wh› or ‹w. Of interest was how often participants used the spellings that occur in the corresponding real words, which we call the critical spellings, and whether they used the critical spellings more often for homophone items than control items. Table 1 shows the critical spelling of each item in the sample quadruplet as well as other potential spellings of the items. Because the homophone and control items in each quadruplet contained the same phonemes, any difference in the rate of critical spellings for homophone and control items could not be attributed to differences in the number of possible spellings of the phonemes and phoneme groups in the two types of items or in the relative frequencies of those spellings given the phonemes. For example, although ‹wh› is a less common spelling of /w/

Table 1. Sample quadruplet of items from Experiments 1 and 2.

Item	Item type	Onset	Remainder	Critical spelling	Examples of other potential spellings
/'wɪntɚ/	Homophone	/w/	/'ɪntɚ/	‹winter›	‹whinter›, ‹winnter›
/'wɪkəl/	Control	/w/	/'ɪkəl/	‹wickle›	‹whickle›, ‹wiccal›
/'fɪntɚ/	Control	/f/	/'ɪntɚ/	‹finter›	‹phinter›, ‹finnter›
/'fɪkəl/	Homophone	/f/	/'ɪkəl/	‹fickle›	‹phickle›, ‹ficcal›

than ‹w› is, this should affect the spelling of homophone and control items in the same way. Both remainders began with the same vowel, which means that any influence of the following vowel on the spelling of the onset or vice versa should affect the spelling of the homophone and control items in the same way. The critical spellings of the homophone and control items in each group had the identical digrams. The mean trigram frequencies of the critical spellings of the homophone and control items, calculated from a sample of approximately one trillion words (Brants & Franz, 2006), did not differ significantly (p = .44).

Seven of the 14 quadruplets were assigned to Set A and seven to Set B, and each participant spelled items from one set. This procedure kept the experiment at a manageable length while ensuring that the total number of items was relatively large. To devise plausible definitions for the experimental items, we chose 56 rare nouns (e.g. *quidnunc*, a person who is eager to learn the latest news and gossip). A pilot study with three individuals from the same population as the study participants showed that none was familiar with any of the rare words. We used the definitions of these rare nouns in the experiment, dividing them into two equal sets.

The experiment included 56 filler words with their correct definitions. Some were difficult to spell, such as *penicillin*. Others were potentially unfamiliar to some participants, such as *masticate*. We used the same filler words for all participants.

Procedure

Each participant was assigned to one of the two sets of experimental items and one of the two sets of definitions. An equal number of participants had each of the four possible combinations. For each participant, the definitions were randomly assigned to the experimental items. The 28 experimental items for each participant were intermixed with the 56 filler words. The resulting list was quasi-randomly ordered for each participant such that at least six items intervened between two experimental items from the same quadruplet.

Half of the participants were assigned to a self-report condition in which they were asked after spelling each experimental item why they spelled it the way they did. To avoid undue focus on the unfamiliar items, these participants were also asked to explain their spellings of 11 randomly chosen filler words. The other half of the participants were not asked to explain any of their spellings. Half of the participants in each of the self-report and standard conditions were asked to invent and say aloud a sentence for each item that exemplified the given definition. The other half were not asked to do this.

The experimenter explained that the participant would hear English words, some of which were extremely rare. The experimenter pronounced each item, read the definition and said the item again. The participant repeated the item. Participants in the sentence condition were asked to produce a sentence using the item and then to spell it. The definition of each item was printed on an answer sheet, together with a line on which to spell it. Participants in the self-report condition were asked, immediately after spelling the selected items, why they spelled the item the way they did. After finishing the spelling task, the participants were asked to describe their observations about the items and were then asked whether they believed that all of the items were real English words with the definitions that were provided. At the end of the experiment, the experimenter disclosed that some items were not words with the given definitions. As a part of this debriefing, she gave the participant a list showing the items for which this was the case and the words that actually had those definitions.

Participants

The participants were 48 individuals from the Washington University subject pool, which consists primarily of undergraduate students. The mean age of the participants in this and the subsequent experiments was 21 years.

Results

We asked whether participants spelled both the onset and remainder of an experimental item as in the corresponding real words from the same quadruplet of items, producing what we call a critical spelling. If participants avoid producing the same spellings for words with different meanings, as Baker (1980) suggested, then they should produce fewer critical spellings for homophone items than control items. The results showed the opposite pattern. For homophones, as Table 2 shows, 66% of responses were the critical spelling. For control items, 42% of responses were the critical spelling. That is, participants often used the spelling that was associated with a familiar phonological sequence even though the sequence was presented with a new meaning. This was true even though alternative spellings of the phonemes (e.g. ‹wh› for /w/) were available to participants, as shown by the fact that they used alternative spellings over half the time on control

Table 2. Mean proportion critical spellings of homophone and control items in Experiments 1, 2, and 3.

Experiment	Homophone items (SD)	Control items (SD)
1	.66 (.17)	.42 (.14)
2	.64 (.26)	.50 (.12)
3, production task	.60 (.18)	.27 (.12)
3, choice task	.27 (.18)	.46 (.14)

items. Of the 48 participants, 44 produced more critical spellings for homophone items than for control items, two produced the same number of critical spellings for both and two showed a difference in the opposite direction.

Statistical support for the conclusions came from a multilevel analysis of performance that used as fixed factors: the main effects of item type (homophone vs. control), self-report condition (self-report vs. standard) and sentence condition (sentence vs. no sentence), as well as the interactions among these factors. The model included separate random intercepts for each item quadruplet and each participant, random slopes for each quadruplet as a function of item type, self-report condition and sentence condition, and random slopes for each participant as a function of item type. The dependent variable was whether the participant produced the critical spelling. This and the other multilevel analyses that we report were performed using the software package lme4 (Bates, Maechler, & Bolker, 2011), selecting a generalised mixed-effects model with a logit link function. The only significant effect was the main effect of item type (β = 1.55, SE = .28, z = 5.49, $p < .001$). The proportion of critical spellings was significantly higher for homophone items than for control items.

One potential explanation for the finding that participants often produced spellings such as ‹winter› for /'wɪntɚ/ with a new meaning is that the experiment included a number of real words. Producing a real word on one trial might have carried over to producing a word on the next trial. However, a model that included the factor of preceding trial type – whether the preceding trial was a word with the correct definition – and the interaction of preceding trial type with item type did not account for significantly more variance than a basic model that included only the factor of item type (p = .62).

The participants varied greatly in their proportion of correct spellings of filler words (mean = .68; range = .23–.98). To determine whether the results on the experimental items varied as a function of spelling ability, we compared a model that included proportion correct on words and the interaction of proportion correct on words with item type to the basic model described above. Proportion correct was centred before inclusion in the model. The more complex model did not account for significantly more variance than the basic model (p = .42).

Two participants, when asked at the end of the experiment to describe their observations about the items, mentioned thinking that some of the items might not be real English words with the definitions that were provided. One additional participant mentioned this during the experiment itself. When directly questioned about this matter, 24 additional participants expressed some doubts. Although some of the stated doubts involved correct definitions, we counted all 27 of the participants just mentioned as having doubts about the items. The 21 participants who expressed confidence that all the items were real English words showed virtually the same results as the group as a whole. A model that included the factors of participant belief (participant thought that some of the items were not real words with the given definitions vs. participant did not question the items), item type and their interaction did not fit the data significantly better than the basic model (p = .97).

Many of the justifications that participants in the self-report condition provided for their spellings were vague, as in ‘That’s just the way it sounded’. However, participants sometimes mentioned the corresponding word when asked why they spelled a homophonic non-word the way they did. In 127 of these cases, participants said that they spelled the new item like the known word. For example, one participant who heard /'wɪlo/ defined as ‘a musical instrument related to a guitar’ explained that she spelled it as ‹willow›, as in willow tree. Explanations that involved an attempt to avoid identical spellings for words with different meanings, such as ‘I didn’t think that it should be spelled like *culture*, so I added a vowel’ (to explain the spelling ‹coulture› for /'kʌltʃɚ/ presented with the meaning of ‘a metal tag or sheath at the end of a lace used for tying, as of a shoelace’) occurred only 40 times.

Discussion

The participants in Experiment 1 sometimes generated novel spellings such as ‹fead› for /fid/ when told that it was a rare word meaning ‘an upholstered settee for two persons’. However, such spellings were surprisingly uncommon, less common than anticipated given the hypothesis (Baker, 1980) that experienced spellers prefer to spell the two meanings of a homophone differently. About two-thirds of the time, the participants in our spelling production task used the spellings with which they were familiar, in this example ‹feed›. Participants did not do this just because the /f/–‹f› and /i/–‹ee› correspondences are common in English, for they used these correspondences at a lower rate when they spelled control items such as /fip/ and /dʒid/ than when they spelled homophone items such as /fid/ and /dʒip/. The pattern of results was the same for the best spellers in our sample and for those who were below average for this highly select group. The pattern of results was also the same whether or not participants were asked to produce a sentence exemplifying the novel meaning of the item before spelling it. Moreover, participants were more likely to report that they spelled a novel item like its homophonous mate than to say that they chose a different spelling in order to visually distinguish the two meanings.

Asking participants to explain the reasons for their spellings did not appear to change the nature of the spellings that they produced. A number of previous studies of spellings have solicited self-reports, assuming

that they provide insights into the processes that people use without substantively changing the nature of those processes (e.g. Sénéchal, Basque, & Leclaire, 2006). However, we know of no study that compared a self-report condition to a standard spelling condition within the same experiment. Our results suggest that asking people to explain why they spell words as they do, although it increases the metalinguistic focus, does not change the spellings that people produce in any substantial way.

Why did our participants show a different pattern of results than expected on the basis of Baker's (1980) findings? We suggested one potential explanation earlier: the absence of appropriate control items in Baker's study. Another potential explanation is that we did not ask people to make explicit judgments about how words should be spelled in English, whereas Baker did. That is, Baker's task may have required a higher degree of metalinguistic awareness than even the self-report condition of Experiment 1. We tested this explanation in Experiment 2 by asking participants to think about the situation in which a new word comes into English and a spelling must be decided upon. We asked participants to consider which way of spelling each non-word would be best, were the non-word to come into wide use with the new definition. Half of the participants in Experiment 2 served in a self-report condition in which they were asked to explain their reasons for each judgement, further increasing the metalinguistic requirements of the task. The other half of the participants were not asked to provide such explanations.

Experiment 2

Method

Stimuli

The stimuli were the same as in Experiment 1.

Procedure

Each participant was assigned to one of the two sets of experimental items and one of the two sets of definitions. An equal number of participants had each of the four possible combinations. For each participant, the definitions were randomly assigned to the experimental items. The experimental items were presented in a different quasi-random order for each participant such that at least six items intervened between two experimental items from the same quadruplet. The 56 filler words were presented separately, in a different random order for each participant.

The experimenter told the participant that she would present a series of new words that might come into English and would ask the participant to think about the best way to spell each one. Participants were told that the items could generally be spelled in more than one way and that they should choose the best spelling. The experimenter pronounced each item, read the definition and pronounced the item again. The experimenter asked the participant to repeat the item and then write down the best way to spell it. The participants in the self-report condition were asked, immediately after they responded to each item, why their proposed spelling was better than other potential spellings. Participants were then told that they would be asked to spell a list of existing English words. The experimenter said each filler word, together with its definition. The participant was asked to repeat each word and then spell it. The definitions that were provided to the participants were written on the answer sheets for both the experimental and filler spelling tasks. Next, participants were asked in writing whether it is better, when a new word enters English that has the same pronunciation as an existing word but a different meaning, to spell the new word like the existing word or differently. Participants were asked to explain their reasons and also to give reasons why another person might support the opposing view. At the end of the experiment, participants were debriefed as in Experiment 1.

Participants

The participants were 24 individuals from the Washington University subject pool.

Results

Participants produced more critical spellings of homophone items than control items, as Table 2 shows. A multilevel analysis using item type (homophone vs. control), condition (self-report vs. no self-report) and their interaction as fixed factors, random intercepts for participants and item quadruplets, random slopes for each participant as a function of item type, and random slopes for each item quadruplet as a function of item type and condition showed a significant effect of item type (β = 1.07, SE = .37, $z = 2.86$, $p = .004$) and no other significant effects. Of the 24 participants, 18 produced more critical spellings for homophone items than control items and one showed a tie. In the self-report condition, there were 94 cases in which a participant stated that he or she spelled a homophonic item like the corresponding word and 31 cases in which the participant stated that the new item should be spelled differently.

The mean proportion of correct responses on the real words, .68, was identical to that found in Experiment 1. A model that included the proportion of words that the participant spelled correctly and its interaction with item type did not account for significantly more variance than a model that included item type only ($p = .75$).

When asked on the post-test whether it is better to spell a new word differently from or the same as an existing word that happens to be pronounced in the same way, 19 of the 24 participants stated that it is better to spell the new word differently. These participants often cited the

confusion that could occur in reading if the two meanings were not differentiated in spelling. All participants were able to provide an argument for the opposing view, however, often along the lines that it could be difficult for spellers to learn and remember a second spelling for the same phonological form. Participants were more likely to state that it is better to spell a new word differently from its homophone mate in the post-test, in which they were asked about their beliefs about spelling in general, than to actually produce such spellings in the experimental task ($p < .001$ by Fisher's exact test).

Discussion

Experiment 2 was designed to test whether increasing the metalinguistic focus of a spelling production task would cause participants to produce a high rate of novel spellings for homophones. It did not. The participants in Experiment 2, like those in Experiment 1, spelled a novel meaning of a phonological form in the same way as the existing meaning about two-thirds of the time. For example, when asked about the best way to spell /fid/ were it to come into English meaning 'an upholstered settee for two persons', participants often wrote ‹feed›. They tended to use the familiar whole-word spelling that was linked to the phonological form, even though their performance on the control items showed that they did not use ‹f› for /f/ and ‹ee› for /i/ in all situations. Thus, asking participants to think about how words should be spelled, as Baker (1980) did, did not cause them to them to produce a high rate of novel spellings for homophones. We thus attribute the findings of Baker's experiment to a lack of appropriate control items, as mentioned earlier.

The participants in Experiment 2 showed a different pattern of results at the end of the experiment when asked more abstractly how homophones should be spelled. Now, most participants stated that members of a homophone pair should be spelled differently, often citing the ambiguity that could result if the meanings were not distinguished orthographically. In this, the university students in our study were like the scholars who have suggested that spelling distinctions such as ‹tale› versus ‹tail› should be retained if the English writing system were reformed (e.g. Wijk, 1961) and that visual differentiation of homophones is an essential and positive feature of the writing system (e.g. Venezky, 1999). The results thus suggest that there is some pressure favouring different spellings of homophones that did not emerge in spelling production, perhaps because it competed with another force.

One force that may work against using a new spelling for a homophone in a spelling production task may be a tendency to use familiar whole-word spellings that are stored in memory. Constructing and producing a new spelling may take more effort than reusing a familiar one, and this may cause people to favour known spellings over novel spellings. If so, we may find a different pattern of results when we do not require participants to construct a novel spelling. In Experiment 3, therefore, we compared a spelling production task to a task in which participants chose between two options on each trial. One option was the critical spelling, which for homophone items was the spelling of the corresponding real word. The second option was a phonologically plausible alternative. For example, participants selected between ‹feed› and ‹fead› as spellings of /fid/ meaning 'an upholstered settee for two persons'. In this choice task, where both spellings are visible, participants can select ‹fead› without having to generate this novel orthographic form themselves.

Half of the participants in both the production and choice tasks of Experiment 3 received instructions like those of Experiment 2. Specifically, these participants were asked to think about the situation in which a new word comes into English and were asked to write down (in the production task) or circle (in the choice task) the best way to spell each item, were it to enter the language with the specified definition. The other half of the participants were told that the items were rare English words and were asked to produce a spelling for each item (production task) or to select the correct spelling (choice task).

We used a different set of items in Experiment 3 than in Experiments 1 and 2 in an attempt to determine whether the results of the production task replicate with a new set of items and to address some potential concerns about the items. One concern is that, although all of the onsets in Experiments 1 and 2 had more than one spelling among the words of English, some of the remainders did not. For example, /k/ has alternative spellings including ‹ck›, ‹c›, ‹k› and ‹cc›, but only ‹ck› occurs in words ending with /'ɪkəl/, the remainder that was used in Experiments 1 and 2. The remainders that were used in Experiment 3 all had more than one alternative spelling. For example, /aɪt/ is spelled as ‹ight› in some words, such as ‹fight›, and as ‹ite› in other words, such as ‹trite›. Given the familiarity of the ‹ight› sequence, it is possible that participants who heard /traɪt/ with the meaning 'felonious taking of personal property, a robbery' would avoid the known spelling ‹trite› and produce ‹tright›. The experimental items of Experiment 3 were all monosyllables. This, too, might also make it easier for participants to override familiar whole-word spellings and generate novel ones.

Another change for Experiment 3 was that we tested participants' ability to spell the words on which the homophone items were based. If a participant misspelled ‹trite› when presented with its correct definition as ‹tright›, for example, use of ‹tright› versus ‹trite› when the phonological form was presented with a different definition would have a different implication than it would for a participant who knew the correct spelling.

Experiment 3

Method

Stimuli

We selected rimes that had two common spellings according to a comprehensive list of monosyllabic words whose printed forms are generally familiar to US university students (Kessler & Treiman, 2001). We chose 13 such rimes, each with a different vowel. The number of rimes was limited by the desire to avoid undue repetition of vowels in the experiment and the desire to use rimes for which the two main spelling alternatives had similar probabilities. For each rime, a homophone item was constructed by using an onset which formed a word when combined with the rime. A control item was constructed that had a different onset. For example, the homophone item involving the rime /el/ was /del/ (‹dale›) and the control item was /krel/. We constructed spelling choices for the choice task by pairing the typical spelling of the onset with each common spelling of the rime. Thus, the choices for /del/ were ‹dale› and ‹dail› and the choices for /krel/ were ‹crale› and ‹crail›. The critical spelling was the first one in each pair, the one in which the rime was spelled as in the homophone. Across the 13 rimes, the mean proportion of monosyllabic words in the Kessler and Treiman (2001) list that used the critical spelling of the rime was .37 and the mean proportion of words that used the alternative spelling was .56. The critical spellings of the homophone and control items did not differ significantly from one another in either mean digram or trigram frequencies (p = .71 and .17, respectively, using counts from Brants & Franz, 2006). The items are shown in Appendix 2. For each experimental item, we selected a definition of a rare English word that was unfamiliar to our participant population and that bore little relationship to the correct definition.

Two quasi-random orders of the 26 experimental items were constructed. In each order, one item with each rime appeared in the first half of the list and the other item with that rime appeared in the second half of the list, with at least six intervening items. The homophone item for a given rime was in the first half of the list 6 times in one of the random orders and 7 times in the other.

Participants were asked to spell the same 56 words that served as fillers in Experiment 1 as well as the 13 words on which the homophone items were based. The items on the filler and base word spelling tasks were presented with their correct definitions.

Procedure

The participants were tested in small groups. The 26 experimental items were presented first. The procedure for the spelling instructions condition was like that of the no-sentence condition of Experiment 1 with standard as opposed to self-report instructions. Specifically, participants were told that the items were rare English words with the specified definitions and were asked to produce or select the correct spelling of each word. The procedure for the spelling decision condition was like that of Experiment 2. Specifically, participants were told that the items were new words that might come into English and were asked to determine the best way to spell each item, were the word to enter the language. Participants either wrote their spelling of each experimental item on their answer sheet, the production task, or circled one of the two spellings that were provided, the choice task. The definitions were written on the answer sheets.

After spelling the 26 experimental items, participants in the spelling instructions condition completed a written questionnaire asking about their observations about the items and asking whether they believed that all the items were real English words with the definitions that were provided. After completing this questionnaire, the participants were told that the items were not actually English words. All participants were then given the base word spelling task and the filler word spelling task. The final questionnaire was the same as that of Experiment 2. It solicited participants' opinions about whether it is better, when a new word enters English that has the same pronunciation as an existing word but a different meaning, to spell the new word like the existing word or differently. Participants were then debriefed as in the earlier experiments.

Participants

The participants were 40 individuals from the Washington University subject pool. Five participants were assigned to each combination of item order, instructions and task.

Results

We performed a multilevel analysis of critical spellings using item type (homophone vs. control), task (production vs. choice), instructions (spelling vs. decision) and their interactions as fixed factors. The model included random intercepts for participants and item pairs, random slopes for participations as a function of item type, and random slopes for item pairs as a function of item type, task and instructions. In the eight instances in which a participant misspelled a word on which a homophone item was based, the participant's spellings of the experimental items that were based on that word were excluded from the analyses. There was a main effect of item type (β = 1.92, SE = .48, z = 3.98, $p < .001$), which was qualified by an interaction with task (β = 2.38, SE = .58, z = 4.14, $p < .001$). In the production task, as Table 2 shows, participants produced more critical spellings of homophone items than control items. The effect of item type was significant in a separate analysis of the production data that used the fixed factor of item type, random

intercepts for each participant and item pair, and random slopes for participants and item pairs as a function of item type (β = 2.12, SE = .49, z = 4.35, p < .001). Of the 20 participants who performed the production task, 17 produced more critical spellings of homophone items than control items. The other three participants showed the same proportion of critical spellings for homophone items and control items. The opposite pattern of results was found in the choice task. Here, as Table 2 shows, participants produced significantly fewer critical spellings for homophone items than control items (β = 1.19, SE = .42, z = 2.86, p = .004). Of the 20 participants in the choice task, 13 chose fewer critical spellings for homophone than control items and two showed the opposite effect. The remaining five participants showed a tie. The overall model also showed an interaction between task and instructions (β = .98, SE = .43, z = 2.30, p = .022). This interaction reflected the fact that participants in the production task produced more critical spellings of both homophone and control items when given decision instructions than when given spelling instructions, while participants in the choice task did not show this tendency.

We carried out another analysis to determine whether participants' use of the critical spellings for different rimes was related to the frequency of the critical spelling in the English vocabulary, using the data from Kessler and Treiman (2001). Pooling across the production and choice tasks, we observed a significant correlation for control items (r = .58, p = .037, two-tailed). For example, ‹ood› was a relatively infrequent choice for /grʌd/, consistent with the fact that /ʌd/ is less often spelled with ‹ood› than ‹ud› in the monosyllabic words of English. The correlation between rime use and rime frequency was not statistically significant for homophone items (r = .13). This latter result suggests that sub-word phonological units play a relatively small role in the spelling process for homophone items, with whole-word spellings stored in memory playing a more important role. Because only 13 rimes were used in the experiment, however, the correlational results must be interpreted with caution.

Participants averaged 69% correct on the filler word spelling task, and a model that included proportion of correct responses on fillers and its interactions with item type and task did not account for significantly more variance than a basic model that included item type, task and their interaction (p = .67). We classified participants who were given spelling instructions as either accepting that all of the items were real English words with the given definitions or expressing doubt. A model that included the factor of participant belief and its interactions with item type and task did not fit the data significantly better than the basic model (p = .46).

When asked near the end of the experiment whether it is better to spell a new word differently from or the same as an existing word that is pronounced in the same way but that has a different meaning, 34 participants stated that it is better to spell the new word differently. Five participants expressed the opposite opinion, and one participant could not decide. The reasons that participants gave for their opinions were similar to those mentioned in Experiment 2, and all participants were able to give reasons for the opposing view when asked why someone else might support it.

Discussion

The results of the spelling production task were similar to those of Experiments 1 and 2. Even though an item such as /θif/ was presented as a rare word meaning 'an abnormal rattling sound in unhealthy lungs' –quite different from the familiar meaning of 'a person who steals' – participants spelled it well over half of the time with the letter string that was linked through long experience with the phonological string /θif/. Participants were more likely to produce the rime spelling ‹eef› for a control item like /plif/ than for a homophone item like /θif/. This result shows that the relatively low rate of ‹theef› spellings for /θif/ does not reflect just a low rate of use of the /i/–‹ee› or /if/–‹eef› mappings. Rather the relatively low rate of ‹theef› spellings for /θif/ appears to reflect the familiarity of the orthographic form ‹thief› and its strong link to the phonological form /θif/. The fact that we found the same pattern of results in the production task of Experiment 3 as in the production tasks of Experiments 1 and 2 suggests that the results of the earlier experiments did not reflect idiosyncratic properties of the items. Even though all the rimes in Experiment 3 had common alternative spellings, and even though all the items were monosyllabic, participants often reused familiar whole-word spellings of homophones in the production task.

The pattern of results was different in the choice task, in which participants saw two spellings of each item and were asked to pick one option. Here, participants were less likely to pick ‹thief› for the homophone item /θif/ than to pick ‹plief› for the control item /plif/. That is, participants tended to avoid the familiar spelling ‹thief› when ‹theef› was presented as an alternative. This result supports the idea that generating a novel orthographic form takes effort. That barrier was removed in the choice task, where a plausible alternative spelling was made available. Under these conditions, participants often used a different spelling for the new morpheme, in line with their stated beliefs about how spelling should work.

General discussion

The present experiments used homophones as a test case to examine the roles of phonology and morphology in the spelling process. If phonology is the most important consideration, then people should spell a new homophone

like they spell its mate. That is, they should consistently use a given orthographic form for a given phonological form. If morphology takes precedence, then people should generally spell a new homophone differently from its mate, even though this means using two different spellings for the same phonological form. Some experimental and linguistic evidence suggests that a drive towards graphic differentiation of homophones may be seen in the history of the English writing system and in the behaviour of present-day spellers (Baker, 1980; Vachek, 1971). That evidence is inconclusive, however, and so we addressed this issue in our experiments.

In our spelling production task, university students spelled a novel item that was homophonous with a familiar word in the same way that they spelled its mate close to two-thirds of the time. For example, participants who were told that /hæf/ was an English word meaning 'an alehouse' would spell it as ‹half›, like its homophonous mate, rather than as ‹haff›. Participants did this when they were asked to produce the spelling that a word should have if it were it to enter the language with a specified meaning and when they were told that the item was a rare word. They did it when they were asked to justify their spelling after having produced it and when they were not asked to do so. Thus, the results were found across a variety of experimental situations, some of which are similar to the real-life situation in which people hear and try to spell a word that is new to them.

Participants in the choice task of Experiment 3 showed a different pattern of results. When presented with both ‹half› and ‹haff›, for example, participants generally chose ‹haff› as the better spelling for /hæf/ meaning 'an alehouse'. Participants' behaviour in the choice task aligned with the belief that most expressed: that different meanings of a homophone should be spelled differently.

Models of spelling are less numerous and less well developed than models of reading, but we may view the results in terms of the computational model of spelling that is currently most well developed, that of Houghton and Zorzi (2003). This model postulates two routes: a lexical route that matches an entire phonological form to a stored spelling and a non-lexical route that learns to map phonemes to letter and groups of letters, using connectionist learning principles. The authors' descriptions of the model indicate that morphology is considered in the operation of the lexical route, meaning that homophones would not necessarily be spelled alike. This is not implemented as a part of the model, however, and so it is not possible to determine what the model would predict for items like those used here. Also, the model does not provide a ready explanation of why different results would be observed in a production task and a choice task. Single-route connectionist models of spelling (Brown & Loosemore, 1994; Olson & Caramazza, 1994) are less well developed than the Houghton and Zorzi model, and it is difficult to assess their fit to our data.

Clearly, more work is needed to develop and test models of the spelling process. Our results provide guidance in this endeavour by suggesting some general guidelines that models should follow. The findings suggest that any model of the spelling process in English must acknowledge that spellers are influenced by multiple forces. The pressures on spelling sometimes conflict with one another, and their roles may vary depending on the task.

One pressure on spelling is towards the use of a distinct spelling for each phoneme. Thus, spellers tend to have difficulty when a phoneme is spelled one way in some words and another way in other words (e.g. Kreiner, 1992; Treiman, 1984). Phonological influences of this sort have been the focus of most previous research and theory on spelling.

A second pressure on spellers, our results suggest, is a tendency to reuse spellings of whole words that have been stored in memory. This occurs, we suggest, because less effort is required to retrieve and produce a familiar whole-word orthographic form than to generate a novel spelling. Although a tendency to produce familiar whole-word spellings has not been explicitly acknowledged in the literature, it can help to explain several previous findings. These include the facts that many spelling errors are real words (Kukich, 1992) and that substitutions such as the more frequent ‹beach› for the less frequent ‹beech› are more common than the reverse (White, Abrams, Zoller, & Gibson, 2008). Our results show that the pressure to use whole-word spellings that are stored in memory often leads people to spell a novel morpheme with the letter string that they have previously associated with a known morpheme that has the same pronunciation. Although spelling homophones alike reduces effort for spellers, the resulting ambiguity can cause problems for readers (Brysbaert, Grondelaers, & Ratinckx, 2000). A similar kind of trade-off between the needs of producers and comprehenders occurs in other areas of language processing as well (Ferreira & Dell, 2000). The tendency to favour spellings of words that have been stored in memory, our results suggest, plays an especially strong role in spelling production tasks. Previous studies suggest that producing the correct spelling of a word is generally more difficult than recognising it (e.g. Fischer, Shankweiler, & Liberman, 1985), but we know of no previous study has found a qualitatively different pattern of results in a production task and a choice task as we did here.

In addition to a tendency to spell phonemes in a consistent manner and to favour stored spellings, spellers of English appear to be influenced by a third pressure: to assign a unique spelling to each morpheme. This third

pressure promotes different spellings of homophones. In production tasks, our results suggest that this tendency is to a large extent outweighed by the tendency to reuse familiar orthographic forms. In other situations, however, a tendency to spell the meanings of a homophone differently comes to the fore. According to our results, this happens when participants see two potential spellings of a homophone and must determine which one is preferable. It also happens when people think abstractly how homophones should be spelled. The idea that spelling is influenced by different pressures and that their strengths vary depending on the task can thus help to explain the surprising finding that people who spell homophones alike often espouse the belief that homophones should be spelled differently from one another.

Previous discussions of the role of morphology in spelling have tended to focus on cases in which a morpheme is spelled the same way across different forms when its pronunciation changes. For example, the fact that *magic* retains its spelling in *magician* complicates mappings at the phoneme level because /ʃ/ is spelled differently in *magician* than in *ship* and *cushion*. However, it simplifies mappings at the level of morphemes, permitting a one-to-one mapping between the letter string ‹magic› and the corresponding meaning. Many scholars (e.g. Chomsky & Halle, 1968) have drawn attention to this property of the English writing system, sometimes called morphological constancy (Bourassa & Treiman, 2008), and a number of studies have examined the extent to which people follow such morphological patterns in their spelling (e.g. Bourassa & Treiman, 2008; Kemp & Bryant, 2003; Mitchell, Kemp, & Dawson, 2011). Some findings suggest that even children are influenced by morphological considerations (see Pacton & Deacon, 2008, for a review). For example, third to sixth graders sometimes produce spelling errors such as ‹hungery› for *hungry* (Bourassa & Treiman, 2008). These children appear to spell the root word in a consistent manner, extending the principle of morphological constancy to cases in which it does not apply. However, as suggested by Pacton and Deacon (2008), someone who spells *hungry* as ‹hungery› may do so less because of a tendency to assign each morpheme a unique spelling than because of a tendency to reuse a familiar spelling, ‹hunger›. The two forces point in the same direction in such cases, and researchers are just beginning to disentangle them (e.g. Pacton, Foulin, Casalis, & Treiman, 2013).

In the experiments reported here, as in a number of previous studies (e.g. Barry & Seymour, 1988), we used performance on non-words to examine the processes involved in spelling in general. This approach is consistent with the model of Houghton and Zorzi (2003), according to which the non-lexical route that supports the spelling of non-words also plays a role in the spelling of known words. The present approach is also consistent with previous findings showing similar patterns of results in spelling tasks with words and non-words (e.g. Treiman et al., 2002). However, because the spelling of novel items might recruit some special strategies that are not normally used to spell words, it will be important in future work to look for converging evidence from studies of word spelling.

Research on the processes involved in spelling is less well developed than research on the processes involved in reading, and much work remains to be done. Using homophones, as we did here, provides one way to assess the roles of different linguistic and performance factors in spelling. With homophones, a tendency to assign each morpheme a unique spelling gives a different result than a tendency to use a familiar letter string that has been stored in memory. In this case, the present results suggest that the tendency to use a familiar orthographic form often wins out in spelling production, even in skilled spellers.

Acknowledgements

We thank Nicole Rosales and Kristina Decker for assistance with the research and preparation of the manuscript.

References

Baker, R. G. (1980). Orthographic awareness. In U. Frith (ed.), *Cognitive processes in spelling* (pp. 51–68). London: Academic Press.

Barry, C., & Seymour, P. H. K. (1988). Lexical priming and sound-to-spelling contingency effects in nonword spelling. *Quarterly Journal of Experimental Psychology*, *40A*, 5–40. doi:10.1080/14640748808402280

Bates, D., Maechler, M., & Bolker, B. (2011). lme4: Linear mixed-effects models using S4 classes. Retrieved from http://cran.r-project.org/package=lme4

Bourassa, D. C., & Treiman, R. (2008). Morphological constancy in spelling: A comparison of children with dyslexia and typically developing children. *Dyslexia*, *14*(3), 155–169. doi:10.1002/dys.368

Brants, T., & Franz, A. (2006). *Web 1T 5-gram Version 1*. Philadelphia, PA: Linguistic Data Corsortium.

Brown, G. D. A., & Loosemore, R. P. W. (1994). Computational approaches to normal and impaired spelling. In G. D. A. Brown & N. C. Ellis (eds.), *Handbook of spelling: Theory, process and intervention* (pp. 319–336). Chichester: Wiley.

Brysbaert, M., Grondelaers, S., & Ratinckx, E. (2000). Sentence reading: Do we make use of orthographic cues in homophones? *Acta Psychologica*, *105*(1), 31–56. doi:10.1016/S0001-6918(00)00047-0

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper and Row.

Ferreira, V. S., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology*, *40*, 296–340. doi:10.1016/S0001-6918(00)00047-0

Fischer, F. W., Shankweiler, D., & Liberman, I. Y. (1985). Spelling proficiency and sensitivity to word structure. *Journal of Memory and Language*, *24*, 423–441. doi:10.1016/0749-596X(85)90038-5

Houghton, G., & Zorzi, M. (2003). Normal and impaired spelling in a connectionist dual-route architecture. *Cognitive Neuropsychology*, *20*(2), 115–162. doi:10.1080/02643290242000871

Kemp, N., & Bryant, P. (2003). Do beez buzz? Rule-based and frequency-based knowledge in learning to spell plural -s. *Child Development*, *74*(1), 63–74. doi:10.1111/1467-8624.00521

Kessler, B., & Treiman, R. (2001). Relationships between sounds and letters in English monosyllables. *Journal of Memory and Language*, *44*, 592–617. doi:10.1006/jmla.2000.2745

Kreiner, D. S. (1992). Reaction time measures of spelling: Testing a two-strategy model of skilled spelling. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *4*, 765–776. doi:10.1037/0278-7393.18.4.765

Kukich, K. (1992). Technique for automatically correcting words in text. *ACM Computing Surveys*, *24*, 377–439. doi:10.1145/146370.146380

Mitchell, P., Kemp, N., & Dawson, J. (2011). Variations among adults in their use of morphemic spelling rules and word-specific knowledge when spelling. *Reading Research Quarterly*, *46*(2), 119–133. doi:10.1598/RRQ.46.2.2

Olson, A., & Caramazza, A. (1994). Representation and connectionist models: The NETspell experience. In G. D. A. Brown & M. C. Ellis (eds.), *Handbook of spelling: Theory, process and intervention* (pp. 337–363). Chichester: Wiley.

Pacton, S., & Deacon, S. H. (2008). The timing and mechanisms of children's use of morphological information in spelling: A review of evidence from English and French. *Cognitive Development*, *23*, 339–359. doi:10.1016/j.cogdev.2007.09.004

Pacton, S., Foulin, J. N., Casalis, S., & Treiman, R. (2013). Children benefit from morphological relatedness when they learn to spell new words. *Frontiers in Psychology*, *4*. doi:10.3389/fpsyg.2013.00696

Treiman, R. (1984). Individual differences among children in spelling and reading styles. *Journal of Experimental Child Psychology*, *37*, 463–477. doi:10.1016/0022-0965(84)90071-7

Treiman, R., Kessler, B., & Bick, S. (2002). Context sensitivity in the spelling of English vowels. *Journal of Memory and Language*, *47*, 448–468. doi:10.1016/S0749-596X(02)00010-4

Vachek, J. (1971). The structure of English orthography by Richard L. Venezky, Review by Josef Vachek. *Language*, *47*, 212–216. doi:10.2307/412200

Venezky, R. (1999). *The American way of spelling: The structure and origins of American English orthography*. New York: Guilford Press.

White, K. K., Abrams, L., Zoller, S. M., & Gibson, S. M. (2008). Why did I right that? Factors that influence the production of homophone substitution errors. *Quarterly Journal of Experimental Psychology*, *61*, 977–985. doi:10.1080/17470210801943978

Wijk, A. (1961). Regularized English. *Educational Research*, *3*(2), 157–159. doi:10.1080/0013188610030208

Appendix 1. Quadruplets of experimental items for Experiments 1 and 2

Homophone items are first and fourth in each quadruplet and control items are second and third. Critical spellings for each item are given in parentheses.

/'wɪlo/, /'wɪltɚ/, /'filo/, /'filtɚ/ (willow, wilter, fillow, filter)
/'sɛkənd/, /'sɛkəl/, /'hɛkənd/, /'hɛkəl/ (second, seckle, hecond, heckle)
/'krɪkət/, /krɪz/, /'frɪkət/, /frɪz/ (cricket, crizz, fricket, frizz)
/'sɛvən/, /sɛk/, /'nɛvən/, /nɛk/ (seven, seck, neven, neck)
/'sɛnət/, /'sɛvi/, /'hɛnət/, /'hɛvi/ (senate, seavy, henate, heavy)
/'kɑrtən/, /'kɑrgəl/, /'gɑrtən/, /'gɑrgəl/ (carton, cargle, garton, gargle)
/'wɪntɚ/, /'wɪkəl/, /'fɪntɚ/, /'fɪkəl/ (winter, wickle, finter, fickle)
/zum/, /zul/, /fum/, /ful/ (zoom, zool, foom, fool)
/'ʃæbi/, /'ʃærəl/, /'ɹæbi/, /'ɹærəl/ (shabby, shattle, rabby, rattle)
/dʒip/, /dʒid/, /fip/, /fid/ (jeep, jeed, feep, feed)
/dʒɛm/, /'dʒɛvɚ/, /nɛm/, /'nɛvɚ/ (gem, gever, nem, never)
/fist/, /fiv/, /wist/, /wiv/ (feast, feave, weast, weave)
/'kʌzən/, /'kʌzəl/, /'nʌzən/, /'nʌzəl/ (cousin, cuzzle, nousin, nuzzle)
/'kʌltʃɚ/, /'kʌsəl/, /'hʌltʃɚ/, /'hʌsəl/ (culture, custle, hulture, hustle)

Appendix 2. Pairs of experimental items for Experiment 3

Homophone item is first in each pair and control item is second. Alternatives in choice task are listed in parentheses after each item, with choice using critical rime spelling first.

/del/ (dale, dail), /krel/ (crale, crail)
/traɪt/ (trite, tright), /jaɪt/ (yite, yight)
/θif/ (thief, theef), /plif/ (plief, pleef)
/most/ (most, moast), /nost/ (nost, noast)
/plum/ (plume, ploom), /frum/ (frume, froom)
/fɛd/ (fed, fead), /jɛd/ (yed, yead)
/kraʊd/ (crowd, croud), /traʊd/ (trowd, troud)
/tɔt/ (taught, tought), /mɔt/ (maught, mought)
/fɝm/ (firm, ferm), /nɝm/ (nirm, nerm)
/mɪθ/ (myth, mith), /bɪθ/ (byth, bith)
/ʃʊd/ (should, shood), /ðʊd/ (thould, thood)
/blʌd/ (blood, blud), /grʌd/ (grood, grud)
/hæf/ (half, haff), /zæf/ (zalf, zaff)