THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 2004, *57A* (2), 325–356

Beyond the two-strategy model of skilled spelling: Effects of consistency, grain size, and orthographic redundancy

Conrad Perry

The University of Hong Kong, Hong Kong and Macquarie Centre for Cognitive Science, Macquarie University, Sydney, Australia

Johannes C. Ziegler

Centre National de la Recherche Scientifique and Université de Provence, France

Most models of spelling assume that people rely on two procedures when engaging in spelling: a lexical look-up procedure that retrieves spellings in their entirety, and a nonlexical procedure that constructs spellings with a set of phoneme–grapheme rules. In the present research, we investigated whether larger sized subsyllabic relationships also play a role in spelling, and how they compare to small-sized phoneme–grapheme relationships. In addition, we investigated whether purely orthographic units can explain some of the variance typically attributed to the mapping between sound and spelling. To do this, we ran five spelling experiments, two using real words and three using nonwords. Results from the experiments showed that there were independent contributions of both phoneme–grapheme and larger sized subsyllabic sound–spelling relationships, although the effect of phoneme–grapheme-sized relationships was always stronger and more reliable than larger sized subsyllabic sound–spelling relationships. Purely orthographic effects were also shown to affect word spelling, but no significant effects were found with nonword spelling. Together, the results support the hypothesis that a major constraint on spelling comes from phoneme–grapheme-sized relationships.

Two strategies that are important in skilled spelling are a lexical strategy, which relies on accessing word-specific memory, and a sublexical strategy, which relies on exploiting regularities/consistencies in the mapping between sounds and letters (e.g., phoneme–grapheme rules). This two-strategy assumption has been incorporated into a number of spelling models (e.g., Barry & Seymour, 1988; Ellis, 1984; Kreiner, 1992) and is supported by both behavioural and neuropsychological evidence (e.g., Barry & Seymour, 1988; Goodman-Schulman & Caramazza, 1987; Kreiner & Gough, 1990).

Correspondence should be addressed to Conrad Perry, Joint Laboratories for Language and Cognitive Neuroscience, The University of Hong Kong, Hong Kong. Email: conrad@hkucc.hku.hk

This work was supported by an ARC special centre grant awarded to Max Coltheart and a University Development Fund grant from the University of Hong Kong.

---

http://www.tandf.co.uk/journals/pp/02724987.html DOI:10.1080/02724980343000323

In the present article, we attempt to go beyond the two-strategy model of skilled spelling by investigating whether there is more than one sublexical strategy used in skilled spelling, and whether word-specific memory involves orthographic information below the whole-word level. More specifically, most models mentioned above assume that, in general, the sublexical strategy uses some sort of phoneme–grapheme rules (e.g., Barry & Seymour, 1988; Ellis, 1984; Kreiner, 1992; Kreiner & Gough, 1990; Perry, Ziegler, & Coltheart, 2002a). However, a number of authors have suggested that sublexical units other than phonemes, such as rimes and consonant–vowel clusters, are also important in spelling (e.g., Bruck & Treiman, 1990; Treiman, 1991; Treiman & Zukowski, 1988). It therefore seems important to examine to what extent smaller and/or larger relationships play a role in spelling. Furthermore, since the use of these relationships may be affected by more than a single factor, the second goal of the study was to investigate which factors are involved and how they affect people's spelling.

One way to investigate whether larger relationships are used when spelling is to examine the statistical distribution of vowel spellings that people give when spelling nonwords. For example, in a recent study, Treiman, Kessler, and Bick (2002) manipulated rime consistency (i.e., the number of times a rime is spelled in a particular way divided by the total number of spellings of that rime) with two groups of nonwords, one where the most common vowel spellings were in rimes of relatively high consistency (that is, in rimes that typically have the same vowel spelling), and one where the same vowel spellings were in rimes of relatively low consistency (that is, in rimes that have many vowel spellings). For instance, they examined the proportion of responses that used an -igh grapheme spelling with the nonword /zaɪt/ (*zight*), which shares a rime with many words using an -igh spelling (e.g., *might*, *tight*, *fight*), and compared it with the proportion of times the -igh spelling occurred with a nonword such as /paɪf/ (*pighf*), which does not share any rimes with words that use an -igh spelling. They found that people were influenced by rime consistency and were more likely to use vowel spellings that were the same as the most common rime spellings in the high-consistency compared to the low-consistency group. From this they concluded that the rime was an important unit in people's spelling.

A problem with studies that look only at one mapping at a time, such as the rime–body mapping (the body is the orthographic equivalent of the rime, e.g., the -ack in *black*) is that we do not know how that mapping would compare quantitatively to another mapping. For example, Perry, Ziegler, and Coltheart (2002b) found that phoneme–grapheme contingency (i.e., the number of times a phoneme is spelled with a particular grapheme divided by the total number of spellings of that phoneme, e.g., /ɜː/→ur = .35 (99/286), since -ur occurs 99 times, whereas there are 286 spellings of the vowel /ɜː/) played an important role in a spelling study where rime consistency was held constant. In particular, they found that even if a nonword shared a rime that could be spelt in two different ways a similar number of times (e.g., the nonword /θɜːn/, which uses the /ɜːn/ rime, shares the rime with four words spelt with a -urn spelling, e.g., *burn*, and five with a -ern spelling, e.g., *fern*), people would typically give a spelling with the most common phoneme–grapheme relationship. In the nonword /θɜːn/ case, this meant that participants predominately used an -ur spelling (*thurn*), which is the most common grapheme spelling of /ɜː/, compared to an -er spelling (*thern*), which is the second most common spelling of /ɜː/. Similarly, Barry and Seymour (1988) also observed a preference for phoneme–grapheme correspondences when other influences were controlled. Does this mean that small-size mappings are therefore more important than the large-size mappings investigated by Treiman and colleagues (Treiman et al., 2002)? To answer this question, we need to compare the contribution of small-size mappings (e.g., phoneme–grapheme relationships) with that of large-size mappings (rime–body relationships) within a single spelling study. Teasing apart the effects of different grain-size mappings was one of the major goals of the present study.

This enterprise is illustrated in Figure 1. Here, the sublexical procedure of the two-strategy model is further broken down into various grain sizes, including syllables, rimes, and phonemes. One of the problems we wish to address is to what extent people exploit consistency at the rime/body level compared to consistency at the phoneme/grapheme level when they engage in skilled spelling. Because the focus of the present article is on monosyllabic stimuli, we do not discuss or investigate the influence of a syllable or morpheme level on spelling (but see, e.g., Fischer, Shankweiler, & Liberman, 1985; Holmes & Caruthers, 1998). Note also that the focus of this study is on spelling to dictation. A model of spelling that acquires its input from picture naming (e.g., Bonin, Peeremen, & Fayol, 2001) might make different predictions, since it would not have an initial auditory input stage.

The second problem we wish to address is related to the lexical strategy and the nature of orthographic information that is used. In particular, it is often assumed that the lexical strategy is analogous to looking up the spelling of a word in a dictionary. The question we are interested in is whether people use purely orthographic information below the whole-word level to constrain their spellings. For example, when spelling, one can imagine that people exploit the orthographic redundancy of a language—that is, the fact that some sequences of letters occur more often together than others. This issue differs from the mapping issue discussed above. While the mapping issue is concerned with finding out which grain sizes play an important role in the mapping *between* phonology and orthography, the orthographic redundancy issue is concerned with whether commonly occurring patterns *within* the orthographic system itself have a measurable effect on spelling.

As a measure for orthographic redundancy, we use body neighbours (BodyN)—that is, the number of words that share an orthographic pattern that corresponds to the body unit (all letters after the initial consonant cluster; i.e., *black* (-ack) has the body neighbours *tack*, *sack*, *lack*, etc.). As far as orthographic measures are concerned, this seems like the most

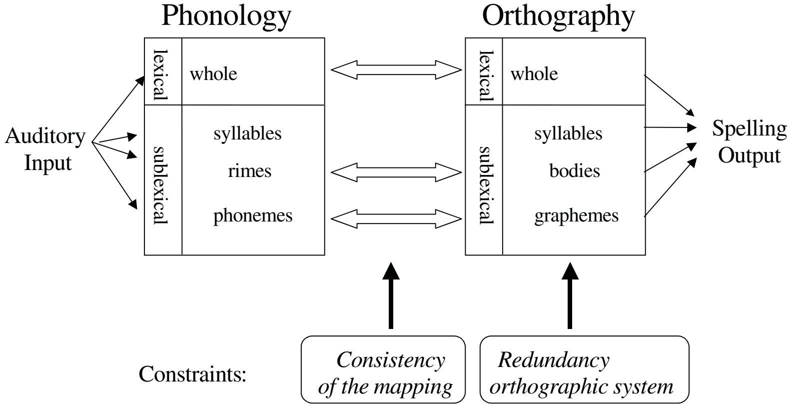

Figure 1. Schematic diagram of the different constraints and levels of subsyllabic information within the standard two-strategy model of skilled spelling.

straightforward choice to begin with because the body unit has been found to affect people's performance in studies of reading (e.g., Forster & Taft, 1994; Montant & Ziegler, 2001; Ziegler & Perry, 1998; Ziegler, Perry, Jacobs, & Braun, 2001). Also, if orthographic processing is dependent on phonological processing (e.g., Share, 1995), then the orthographic body may have special status over other potential larger units.

An additional theoretical reason for examining BodyN is that subsyllabic effects may arise at more places than only a sound–spelling translation. In particular, there may be subsyllabic effects that arise at an orthographic lexical level (Ziegler & Perry, 1998; Ziegler et al., 2001). For example, in the case of spelling a homophone that has two or more potential spellings, the frequency of subsyllabic orthographic units might influence which homophone spelling is chosen over and above purely whole-word orthographic frequency. Similarly, sublexical sound–spelling relations at the phoneme or rime level might influence word spelling over and above purely lexical information (e.g., Barry & Seymour, 1988; Kreiner, 1992; Perry et al., 2002a; Rapp, Epstein, & Tainturier, 2002). In addition, rather than base their entire spelling on purely sound–spelling or lexical information, people may use other sources of information, such as which patterns of letters are commonly occurring. The idea here is that people may not like to use non-existing or low-frequency letter sequences, such as -ighf. Thus, if there is a trade-off between different forms of information (i.e., sound–spelling vs. purely orthographic), purely orthographic information might have some effect.

Looking at rime consistency (sound–spelling information) and BodyN (orthographic information) within a single study is also important to fully address the role of rimes in skilled spelling. This is because rime effects could come from either the mapping between phonology and orthography or purely orthographic processes. That is, they may result from the consistency of the rime–body mapping or the orthographic salience of body units over and above the effects of sound–spelling consistency. In summary, the goal of the following experiments was to examine (1) what the sound–spelling relationships are that people use when spelling, and (2) whether purely orthographic information below the whole-word level constrains skilled spelling.

EXPERIMENT 1

In Experiment 1, we investigated the relative contribution of phoneme–grapheme contingency (a small-sized sound–spelling effect), rime consistency (a large-sized subsyllabic sound–spelling effect), and BodyN (an orthographic effect) by having people spell homophones. The existence of sound–spelling effects at either grain size would suggest that assembled orthography has been automatically generated and has interfered with people's spelling (e.g., Barry & Seymour, 1988; Kreiner, 1992; Perry et al. 2002a; Rapp et al., 2002). Furthermore, the existence of BodyN effects would show that the salience of common orthographic units has some effect on people's homophone choice, over and above purely whole-word frequency differences.

Homophones provide a great possibility to look at spelling variability because they have by definition more than one spelling (e.g., *might* and *mite*). Because our variables of interest were highly correlated, one way to investigate the relative and unique contribution of each variable is through the use of a large number of items and regression techniques (for similar approaches in the study of reading, see e.g., Spieler & Balota, 1997; Treiman, Mullennix, Bijeljac-Babic,

& Richmond-Welty, 1995; Ziegler, Rey, & Jacobs, 1998). We therefore obtained spelling responses for all 632 monosyllabic homophones that have two spellings in English according to the CELEX database.[1]

The logic of the study is as follows: On the auditory presentation of a homophone, a person will inevitably have to use one spelling of the homophone (e.g., *mite* vs. *might*). This choice will be primarily influenced by the frequency of each alternative[2]—that is, in most cases, people will be more likely to use a high-frequency alternative than a low-frequency alternative (thus they will use *might* over *mite*, since *might* has a frequency higher than of *mite*). Once differences in word frequency between the two homophone alternatives are partialled out, however, one can investigate what other variables are responsible for a homophone being spelled one way or the other. In this case, this can be done by calculating measures to do with the variables of interest in a systematic way and then correlating those measures with the probability that people spell one homophone compared to another.

For example, the homophone /maɪt/ has two spellings, *mite* and *might*. The phoneme–grapheme contingency values for /aɪ/→i-e (e.g., *mite*, *wine*) and /aɪ/→igh (i.e., *might, light*) differ, however, with /aɪ/→i-e having a higher phoneme–grapheme contingency value than /aɪ/→igh (i.e., -i-e is more commonly used to spell /aɪ/ than is -igh; see Perry et al., 2002a, for a full analysis). This means that the difference in the contingency value of each homophone pair (that is, the contingency value of one item in the pair minus the contingency value of the other) can be correlated with its spelling probability (that is, how likely participants were to spell a particular homophone answer) whilst partialling out frequency differences between the items in each pair. Thus, in this example, if participants had spelt /maɪt/ as *mite* 80% of the time and /maɪt/ as *might* 20% of the time, the criterion variable in a regression analysis would have two data points, .8 for *mite* and .2 for *might*.[3] The predictor variable would be the phoneme–grapheme contingency of /aɪ/→i-e minus the phoneme–grapheme contingency of /aɪ/→igh. In addition, difference in frequency would also need to be generated by taking the log word frequency of *mite* and subtracting the log word frequency of *might*, so that the effect of whole-word frequency could be partialled out from the initial values generated.

The same logic can be used to investigate the effects of rime consistency and BodyN. To calculate the value of the BodyN predictor variable, we subtract the BodyN value of *might* (i.e., the number of times -ight occurs) from the BodyN value of -ite (i.e., the number of times -ite occurs). To calculate the value of the rime consistency predictor variable, we subtract the rime consistency value of *might* from the rime consistency value of *mite*. In all cases, the criterion variable is the same—that is, the proportion of times *mite* is given compared to *might*. Take again the example of the homophone pair *mite–might*. The *-ite* body is less common than the *-ight* body. Thus, people should give the spelling *mite* less often than *might*, according to a

[1]This database was also used in the study by Perry et al. (2002a).

[2]Of course, in writing, the choice of a homophone spelling will be primarily determined by its meaning. However, in a laboratory situation, words can be presented in isolation, with no context provided. This allows us to study the influence of purely phonological and orthographic factors that are not contaminated by the meaning of a word.

[3]Note that taking the ratio of responses for one homophone compared to the other or using the difference in the proportion of times one homophone is given compared to the other leads to identical values in the correlation. We used difference scores instead of ratios in order to stay consistent with the analysis performed in Perry et al. (2002a).

rime consistency measure and also according to a BodyN measure (the opposite of a phoneme–grapheme contingency measure, since –i–e is more common than -igh).

Note that whilst rime consistency and BodyN are highly correlated, they are not completely so. Thus the effects of these two variables may differ, depending on the strength of each of the predictors. This allows an orthographic frequency effect (BodyN) to be compared with a sound–spelling effect (rime consistency) in a meaningful way. Thus, for example, both the /aɪs/ and /ɑːb/ rimes are only ever spelt one way in real words (-ice and -arb). They therefore have an identical rime consistency value (1.0). Alternatively, there are 14 words that use the /aɪs/ body (e.g., *rice*, *lice*, *mice*), but only two that use the /ɑːb/ body (e.g., *barb*, *garb*). Thus, in terms of BN and rime consistency measures, the values are quite different.

Method

Participants

A total of 21 people participated in the study. The majority were students at Macquarie University.

Items and procedure

The entire list of 733 monosyllabic homophones in English (including those with more than two spellings) were read aloud to the participants. Participants were asked to write the "first spelling that comes to mind" for each word. The speed at which the list was read out was fast enough such that subjects did not have time to pause between words. The homophones, which were read to small groups of subjects in a single session, were presented in random order. Occasionally, subjects would interrupt the experimenter when they did not hear a word properly. When this happened, the word was simply read out again.

Phoneme–grapheme contingency

To simplify matters, only vowel phoneme–grapheme contingency was used in this study (for an analysis of consonant contingency, see Perry et al., 2002a). This was done because the greatest level of sound–spelling inconsistency in English lies in the vowel. Phoneme–grapheme contingency for vowels was calculated for each phoneme–grapheme relationship by counting the number of times the relationship occurs and dividing that number by the total number of times the phoneme occurs. In terms of segmentations used, we used the computations provided by Perry et al. (2002a), who calculated phoneme–grapheme contingency for all phoneme–grapheme relationships from every word in the monosyllabic CELEX database.

Rime consistency and BodyN

Rime consistency was calculated for each rime–body relationship by counting the number of times the relationship occurs and dividing that number by the total number of times the rime occurs. BodyN was calculated for each homophone by counting the number of words that shared the same orthographic body. Orthographic bodies of homophones were segmented from onsets by taking all letters after the initial consonants. The only exception to this was when a -u occurred after the consonant –q (as in *quack*). In these cases, the -u was considered to be part of the onset.

TABLE 1
Summary of partial correlations for all homophones used in Experiment 1 and for the subset of homophones that differed with respect to vowel contingency

	Comparison	*df*	*partial r*	*p*
Overall	BodyN	580	.12	< .005
	Rime consistency	580	-.016	*ns*
	Vowel contingency	580	.055	*ns*
Subset	BodyN	469	.092	< .05
	Rime consistency	469	-.090	= .051
	Vowel contingency	469	.15	< .005

Note: BodyN = difference in body neighbours.

Results

For the sake of simplicity, in the present analysis, we used only the data from 632 homophones, namely those that had two spellings in the CELEX database. The results can be broken into two main sections: the analysis of the overall database and the analysis of a subset of homophones that differed with regard to vowel contingency. The results of all the correlations reported below appear in Table 1.

In terms of the overall analysis, all homophone pairs were selected that had different values on their phoneme–grapheme contingency, BodyN, or rime consistency measures. Thus, for example, the homophone pair *retch* and *wretch* was excluded because all of the measures examining vowel phoneme–grapheme contingency, BodyN, and rime consistency were the same for both homophone spellings (only the onset differed). That criterion left 583 items in the database. The effect of each individual variable was then examined by correlating that variable (as calculated by subtracting one value from the other and thus obtaining the difference) with the percentage of times participants spelled one homophone over the other, whilst partialling out the other three variables and difference in log frequency (the log orthographic CELEX frequency of one homophone spelling minus the log CELEX orthographic frequency of the other). Thus, when vowel phoneme–grapheme contingency was examined, difference in rime consistency, difference in BodyN, and difference in log frequency were partialled out. Similarly, when rime consistency was examined, difference in BodyN, difference in phoneme–grapheme contingency, and difference in log frequency were partialled out. The only variable that was significant with this procedure was difference in BodyN. Neither difference in rime consistency nor difference in phoneme–grapheme contingency remained significant.

Since the previous comparison is likely to underestimate the size of the phoneme–grapheme contingency correlation, as it includes many items that had the same vowel phoneme–grapheme value for both homophones (e.g., *arc* vs. *ark*), we repeated the same analysis on the subset of items that had different vowel phoneme–grapheme contingency scores. When this was done, both difference in phoneme–grapheme contingency and difference in BodyN correlated significantly with the data. In contrast, difference in rime consistency produced an almost significant negative correlation (i.e., the higher the difference in consistency, the less likely the more consistent spelling was used). Thus, it appeared that whilst difference in BodyN helped predict people's vowel spellings, difference in rime consistency did not.

Overall, the results suggest that phoneme–grapheme contingency played a role in people's homophone spelling. This is because, in the more powerful analysis (i.e., the subset analysis), the vowel contingency measure significantly correlated with the data set even once the contribution of larger units was partialled out, although this correlation was quite small. This is in full agreement with the findings reported by Perry et al. (2002b). Difference in BodyN also appeared to be a better predictor of people's homophone spelling than did difference in rime consistency, although again this correlation was quite small. This is because BodyN still significantly correlated with people's responses in the overall analysis even when all other variables were partialled out. That was not true for rime consistency. The partial correlation with that variable did not even approach significance in the overall analysis and in fact appeared slightly negative. Since we have no theoretical reason for the slightly negative correlation, we take it to be noise. In addition, difference in BodyN helped predict vowel spellings over and above vowel contingency. Thus, difference in BodyN correlated significantly with people's homophone spellings over and above vowel contingency, even though difference in rime consistency did not.

Note that even though the correlations are quite small, this is not surprising because the present correlations are partial correlations that were obtained after having factored out the influence of powerful predictors such as word frequency. Thus, the small correlations are not a problem of insufficient power with the measures we used, but are simply a consequence of looking at subtle effects once bigger effects are factored out.[4]

EXPERIMENT 2

Whilst the previous experiment suggested that there was some correlation between vowel phoneme–grapheme contingency, BodyN, and people's homophone spelling, the strength of the correlations was not strong. This means that the results are potentially open to the interpretation that they were caused by factors not of interest in the task, such as potential statistical anomalies that occur from using covariates with highly intercorrelated variables.

To avoid the problem of having weak correlations, we used an alternative method for examining whether BodyN, rime consistency, and phoneme–grapheme contingency affect people's homophone spelling. In particular, we examined each of the three factors in a factorial design. To do this, we chose exemplars from the database that were extreme (high or low) on BodyN, phoneme–grapheme contingency, or rime consistency measures, and we examined performance differences between high and low exemplars in each of those groups. The idea was to see whether significant differences in participants' performance would exist in a more traditional analysis where extraneous variables are controlled rather than partialled out, and where extreme group manipulations are examined.

One method of choosing exemplars for a high and low group for each of the three variables of interest would be to simply choose exemplars by hand, as is typically done. However, this

[4]Such small correlations are typical for many theoretically important variables in some common psycholinguistic tasks, such as the regularity effect in the reading aloud task (Spieler & Balota, 1997).

TABLE 2
Mean difference scores for groups used in Experiment 2

		Measure			
Comparison		*dif vowel contingency*	*dif rime consistency*	*dif BodyN*	*dif log frequency*
BodyN	Low	.13	.23	–5.93*	–.79
	High	.14	.25	12.26*	–.70
Rime consistency	Low	–.17	–.22*	4.35	.55
	High	–.14	.28*	4.4	.49
Vowel contingency	Low	–.44*	.12	4.08	.14
	High	.37*	.15	4.00	–.16

* = $p < .001$; dif = difference in.

method is slow and time consuming, particularly when balancing on a number of different factors. Furthermore, it is open to experimenter bias when choosing stimuli (Forster, 2000). We therefore automated the procedure by first choosing an extreme set of items in each condition and then allowing a computer program to find the best match to each homophone pair across conditions. Overall means across the groups were then compared to make sure they differed on the variables of interest and were controlled on all other variables. Three conditions were created with this method: a high–low difference in BodyN condition, a high–low difference in rime consistency condition, and a high–low difference in phoneme–grapheme contingency condition. Across all of the groups, all extraneous variables were balanced.[5] A summary of the items chosen in each of the three groups appears in Table 2. Individual items can be found in Appendix A.

Results and discussion

In the participant analysis, participants were significantly more likely to use a homophone spelling in the high phoneme–grapheme contingency group than in the low phoneme–grapheme contingency group: 55.37% vs. 41.63%; $t_1(20) = 6.44$, $p < .001$. Similarly, they gave significantly more spellings used in the high BodyN group than in the low BodyN group: 52.33% vs. 36.33%; $t_1(20) = 4.80$, $p < .001$. A similar pattern did not emerge in the rime consistency manipulation, however, and there was almost no difference between the number of responses given in the high and low groups: 59% (high) vs. 61% (low), $t < 1$. In the item analysis, we used a repeated measures comparison because the items were matched in a strictly pairwise manner by the computer program. The results show that both phoneme–grapheme contingency and difference in BodyN produced significant effects, $t_2(51) = 2.86$, $p < .01$, and $t_2(14) = 2.42$, $p < .05$, respectively. Alternatively, rime consistency failed to produce a significant effect, $t < 1$.

[5]Further details of the method are available from the first author on request.

Overall, the results of this small-scale factorial analysis went in the same direction as those of the large-scale correlational analysis of the previous experiment. That is, when extreme groups were selected that were balanced on other factors, there was a significant effect of phoneme–grapheme contingency and BodyN. No significant effect of rime consistency was found using the same method, however.

EXPERIMENT 3

The previous experiment suggested that both phoneme–grapheme contingency (i.e., small grain-size relationships) and orthographic redundancy, as measured in terms of BodyN, were predictive of people's homophone spelling performance. This result suggests that previous evidence for the role of rime consistency (i.e., large grain-size consistency) in nonword spelling (e.g., Perry et al., 2002b; Treiman et al., 2002) may have been due to differences in BodyN, rather than rime consistency.

However, the present results were obtained by analysing real-word homophone spelling, whilst previous studies used nonword spelling to estimate the influence of rime consistency. It may be that there are differences between the effects that are found for word and nonword stimuli. For instance, the BodyN effect may be stronger for words than for nonwords because words may have lexical representations that are organized in terms of orthographic body units. This hypothesis is consistent with the idea that some form of lexical body representation exists, at least in English (e.g., Forster & Taft, 1994; Montant & Ziegler, 2001; Ziegler & Perry, 1998; Ziegler et al., 2001). If this is the case, then it might be less surprising to find that BodyN effects outweigh the effects of rime consistency in the spelling of real words.

In this experiment, a nonword spelling task was used to investigate whether there was a difference between words and nonwords in terms of BodyN and rime consistency effects. For this purpose, we selected stimulus pairs that varied on either BodyN or rime consistency but shared the same vowels (see Perry et al., 2002b, for an alternative method).Take the two nonwords /fɜːd/ and /fɜːt/ as examples. Both of these have the same vowel, /ɜː/. Similarly, the orthographic body containing the most common vowel occurs at a similar frequency; *-urd* occurs three times (*curd*, *surd*, *turd*) and *–urt* occurs three times (*curt*, *hurt*, *spurt*). The two nonwords differ in terms of rime consistency, however; *-urd* has a rime consistency value of .38, whereas *-urt* has a value of .21. This is because there are more words that use alternative spellings of /ɜːt/ (e.g., *dirt*) than /ɜːd/ (e.g., *third*). Thus if rime consistency has an effect on people's spelling, it would be expected that people would use the most common vowel spelling (*-ur)* more often when spelling /fɜːd/ than when spelling /fɛːt/, due to the *-urd* body's higher consistency. BodyN can be examined in the same way, except that instead of choosing nonwords that differ on rime consistency, they are chosen to differ on BodyN instead. Take the two nonwords /bɜːb/ and /gləʊn/ as an example. The spelling that contains the most common vowel body spelling of /bɜːb/ is -urb. This spelling has BodyN of 2 (*curb*, *blurb*) and a rime consistency of .5, because there are two other words that use a different body spelling but the same rime (*herb*, *verb*). Alternatively, the body spelling with the most common vowel spelling for the nonword /gləʊn/ is -one. This body spelling occurs 14 times (e.g., *zone*, *lone*). However, if the rime consistency value of /gləʊn/

is examined then it is .48, which is very similar to the rime consistency of -urb, because there are many other words with different rime spellings (e.g., *shown*, *moan*).

It is therefore possible to examine whether rime consistency and BodyN effects exist in spelling by counting the number of times the most common vowel is given in one group high on a given measure (i.e., rime consistency or BodyN) and comparing it to the number of times the most common vowel spelling is given in another group low on that measure. If people are influenced by BodyN, then they should give the most common vowel spelling for nonwords that have many body neighbours that use the predominant vowel spelling compared to nonwords that have only few body neighbours that use the predominant vowel spelling. Alternatively, if body-sized units do not have a strong influence on spelling, then no effect should be found. Similarly, if rime consistency plays a role, then participants should use more spellings containing the most frequent vowel when relatively consistent rime–body relationships are used than when inconsistent ones are used. The design of the experiment is illustrated in Figure 2.

To make sure that the results are generalizable over task conditions, we used two experimental groups. In one of the groups, nonword fillers were used, and in the other, word fillers were used. The idea was to investigate the hypothesis that BodyN effects may trade off with rime consistency effects or sound–spelling relationships of a different size depending on whether the task involves word or nonword spelling. At least in children's reading, filler type has been found to have extremely strong effects on the type of units that are used (Brown & Deavers, 1999). The idea here is that the ratio at which small- and large-sized relationships are used might be a property of the experimental condition, rather than a more general property of spelling, and this we wanted to examine.

Finally, apart from using the experiment to examine the effect of larger sized relationships on nonword spelling, it is also possible to use the experiment to examine smaller

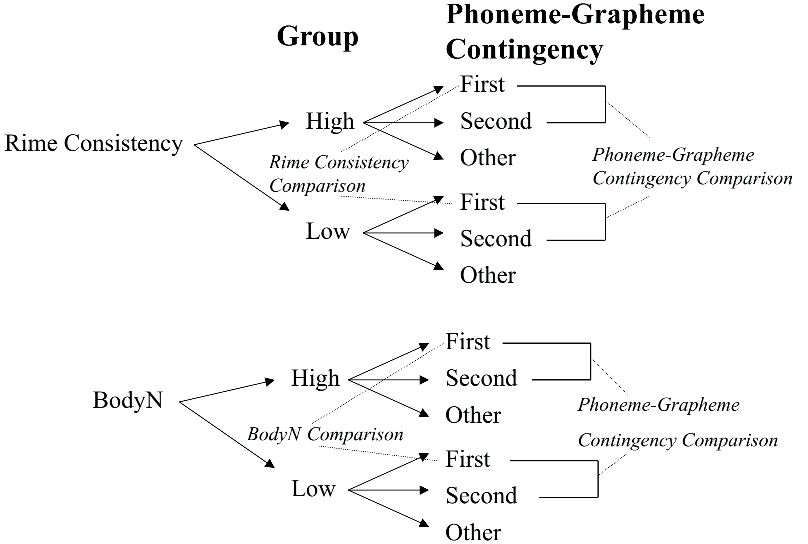

Figure 2. Design of Experiment 2.

phoneme–grapheme-sized relationships. To do this, we examined the frequency at which the most common versus the second most common phoneme–grapheme relationships were used (Barry & Seymour, 1988). Thus, for instance, for the nonword /zaɪt/ we compared how many times the spelling –i-e was used (the most common grapheme spelling for /aɪ/), with the number of times the -igh grapheme was used (the second most common grapheme spelling of /aɪ/). If the phoneme–grapheme relationship is important in people's nonword spelling, then people should, in general, use the most common vowels when spelling the nonwords.[6]

Method

Participants

A total of 43 students studying first-year psychology at Macquarie University participated in return for course credit. Of these 23 participated in the nonword filler condition, and 20 participated in the word filler condition.

Stimuli

Two separate sets of nonwords were used, one for the rime consistency manipulation (high vs. low), and one for the BodyN manipulation (many vs. few). In the rime consistency group, 30 nonword pairs were used. Both members of each pair had the same vowel, but one nonword had a higher rime consistency score than the other nonword. The idea was to see whether people would give the most dominant vowel spelling more often if a vowel was embedded in a highly consistent rime than if it was not. For instance, the two nonwords /fɛːp/ and /nɛːk/ both use the same vowel. However, the vowel spelling -ur occurs in 50% of words that use the /ɛːp/ rime (e.g., *burp*), but only 27% of the time in words that use the /ɛːk/ rime (e.g., *lurk*). Thus if the rime–body context biases people's spellings, it would be expected that people would use the -ur spelling more often for the stimuli /fɛːp/ than /nɛːk/. Across the two groups, the BodyN value of the spelling with the most common vowel was kept stable, as was the feedforward consistency of the spellings with the most common vowels. Since we were interested in the number of times the dominant vowel spelling was given, rime consistency was calculated by using two groups for each item, which were: (1) all bodies that used the dominant vowel ("for" counts); and (2) all bodies that do not use the dominant vowel ("against" counts). For example, the rime /ɛːk/ has six different body spellings: *-urk* (which is in 3 words), *-urke* (1), *-erk* (3), *-irque* (1), *-irk* (6), and *-ork* (1). Thus the rime consistency value for all words that contain the most common grapheme -ur (for counts) is 3 (urk) + 1 (urke)/15 (i.e., the total number of bodies). BodyN was calculated as the sum of all bodies with the most common vowel. Thus the BodyN value of /ɛːk/ would be 4, since 4 bodies use the -ur spelling. All orthographic bodies used in the experiment were used twice, although different onsets were used with each orthographic body.

The same design was used in the BodyN manipulation, except that rime consistency was kept stable while BodyN was manipulated. Thus there were two groups: one that used dominant vowel spellings

[6]Note that the phoneme–grapheme effect may trade off with other effects (such as rime consistency) as found in Perry et al. (2002b). That is, a high-frequency grapheme may not be used as often when embedded in an inconsistent rime–body relationship (e.g., /aɪt/ has many spellings that do not use the most common grapheme, e.g., *might*) as when embedded in a consistent rime–body relationship (e.g., /aɪf/ has only one body spelling, -ife, and that body spelling also uses the most common phoneme–grapheme relationship). Therefore, to examine that effect, phoneme–grapheme contingency effects would need to be examined separately in different groups, as done by Perry et al. (2002b), rather than examined with only a high–low comparison across all possible groups at once, as done here.

that occurred in a high BodyN context, and another that used dominant vowel spellings that occurred in a low BodyN context. The nonwords were chosen pairwise such that a nonword in the high BodyN group would always have a higher BodyN than a vowel matched nonword in the low BodyN group. A deliberate attempt was made in the low BodyN group to keep the values of BodyN as low as possible. Only 18 nonwords were used due to stimulus constraints.

Across the rime consistency and BodyN manipulations, 12 of the same stimuli were used twice. When the same stimuli were used in more than one group, instead of presenting the same stimuli twice, the stimuli were only presented once, and the result was simply used in both groups. Stimuli were repeated instead of creating new stimuli with different onsets but the same rimes to try to reduce the effect of priming between similar-sounding stimuli that occurs even when there are many intervening items (e.g., Campbell, 1983).

The fillers consisted of 76 nonwords in the nonword filler condition and 76 words in the word filler condition. None of the nonword fillers used a vowel that was used in either of the groups designed to examine the manipulation at hand, and nor did any of the word fillers. The items used appear in Appendix B.

Procedure

In the condition where all of the filler items were nonwords, participants were told that the task was nonword spelling, and that all of the stimuli were nonwords. They were asked to write the first spelling that they could think of for each nonword. In the condition where word fillers were used, participants were told that the task was word and nonword spelling, and that they were to write either the spelling of the stimulus if it was a word, or the first spelling that they could think of if the stimulus was a nonword. Participants wrote their answers on a numbered sheet that was provided.

The nonwords and words were played through Harman/Kardon speakers to each participant individually. Participants controlled the pace of presentation by pressing a key on the computer.

The nonwords were pseudorandomized for each participant with the constraint that a filler word always appeared after a test word. This meant that the same vowel never appeared in two consecutive nonwords.

Results

Three nonwords, /glɜːθ/, /miːð/, and /diːð/ and their matched pairs were removed from the analysis due to high error rates, all involving fricative misperceptions. Responses that were either lexicalizations (nonword filler: 3.11%; word filler: 3.02%) or appeared to be sound misperceptions were removed from the analysis (nonword filler: 5.05%; word filler: 0.22%). Results for the rime consistency/BodyN analysis were scored by counting the number of times the most common vowel was given and the number of times another vowel was given. A ratio was then calculated by dividing the most common vowel count by the number of participants minus the number of lexicalizations and sound misperceptions. Thus, for each item, a probability score was derived by dividing the number of times the most common vowel was given and comparing it to the number of times all other vowels were given, taking out erroneous responses.

In terms of larger sized relationships (i.e., rime consistency and BodyN), two separate and independent analyses were done, one for the rime consistency group and one for the BodyN group. These analyses were repeated for both nonword and word filler groups. We used two separate analyses for the two filler groups because independent groups of subjects were used in each of the conditions. Thus, even if the groups were examined together, it would be necessary to use a between- rather than a within-subjects analysis, thus reducing the power of the design. Note also that most studies examining filler manipulations in independent groups have used multiple experiments and thus separate analyses (e.g., Deavers & Brown, 1997; Stone & Van Orden, 1993).

In the nonword filler group, a two-tailed *t* test performed on the rime consistency group showed that participants were more likely to give a dominant vowel spelling when the nonword spelling was of high consistency than when it was not (high: 61.38%; low: 52.00%), although that result was only significant by participants, $t_1(22) = 7.12, p < .001$; $t_2(52) = 1.45$, $p > .1$. The same was not true for the BN manipulation, however (high: 57.03%; low: 52.78%). There was no significant difference between the two groups, both *t*s < 1. A very similar pattern was found with the word filler group. An effect of rime consistency (high: 63.83%; low: 48.93%) was found that was significant by participants, $t_1(19) = 5.65, p < .001$, but not items, $t_2(52) = 1.73, p = .088$. No significant effect was found with the BodyN manipulation (high: 48.33%; low: 49.54%), both *t*'s < 1.

In terms of smaller sized relationships, a different pattern was found when the effect of phoneme–grapheme contingency was examined. This pattern was examined by comparing the number of times the most common vowel spelling was given to the number of times the second most common vowel spelling was given, in each of the four possible groups (i.e., in the two BodyN and the two rime consistency groups). Note that for this analysis we used two analysis of variance (ANOVAs) rather than *t* tests because it may be possible that BodyN or rime consistency affects the number of times people use certain phoneme–grapheme relationships (e.g., Perry et al., 2002b). That is, people might be more prone to give the most common correspondences in groups where the rime consistency value or BodyN also favours those correspondences, thus increasing the difference between how commonly the most common and

TABLE 3A
Percentage of times that participants used the most and second most common vowel spelling as a function of rime consistency and filler type

	Words		*Nonwords*	
Vowel contingency	*High rime consistency*[a]	*Low rime consistency*[b]	*High rime consistency*[a]	*Low rime consistency*[b]
First[c]	64	49	61	52
Second[d]	23	37	26	37

[a]e.g., -arb. [b]e.g., -urk. [c]e.g., -i-e. [d]e.g., -igh.

TABLE 3B
Percentage of times that participants used the most and second most common vowel spelling as a function of Body-N and filler type

	Words		*Nonwords*	
Vowel contingency	*Large BodyN*[a]	*Small BodyN*[b]	*Large BodyN*[a]	*Small BodyN*[b]
First[c]	48	50	57	53
Second[d]	36	31	31	25

[a]e.g., -ice. [b]e.g., -arb. [c]e.g., -i-e. [d]e.g., -igh.

second most common correspondences are given. If this is the case, such an influence should show up as a significant interaction in the ANOVA. Mean results can be found in Table 3A and Table 3B.

In the nonword filler group, a 2 × 2 ANOVA with rime consistency and phoneme–grapheme contingency as factors showed that there was a significant effect of phoneme–grapheme contingency (56.69% vs. 31.34%), $F_1(1, 22) = 20.59$, $p < .001$; $F_2(1, 52) = 17.61$, $p < .001$, but the interaction between the effects of rime consistency and phoneme–grapheme contingency was only significant by participants, $F_1(1, 22) = 31.0$, $p < .001$; $F_2(1, 52) = 2.98$, $p = .09$. A 2 × 2 ANOVA with BodyN and phoneme–grapheme contingency showed a similar pattern, with a significant effect of phoneme–grapheme contingency (54.91% vs. 27.92%), $F_1(1, 22) = 18.73$, $p < .001$; $F_2(1, 30) = 9.16$, $p < .01$, although no significant interaction, $F_1(1, 22) = 1.34$, $p = ns$; $F_2 < 1$.

A similar pattern was found in the word filler group. A 2 × 2 ANOVA with rime consistency and phoneme–grapheme contingency as factors showed that there was a significant effect of phoneme–grapheme contingency (56.38% vs. 29.89%), $F_1(1, 19) = 26.73$, $p < .001$; $F_2(1, 52) = 17.61$, $p < .001$, but the interaction just failed to reach significance by items, $F_1(1, 19) = 42.49$, $p < .001$; $F_2(1, 52) = 3.70$, $p = .06$. Similarly, a 2 × 2 ANOVA with BodyN and phoneme–grapheme contingency showed a significant effect of phoneme–grapheme contingency (48.93% vs. 33.30%), although only by participants, $F_1(1, 19) = 14.41$, $p < .005$; $F_2(1, 30) = 1.85$, $p = ns$, and no significant interaction, $F_1(1, 30) = 2.12$, $p = ns$; $F_2 < 1$.

Discussion

The results of this experiment differ from those found in the previous experiment. In particular, the effects of rime consistency and BodyN were reversed. That is, there was an effect of rime consistency but no significant effect of BodyN. These results support the possibility that there is a weak rime consistency effect that is not modulated by the frequency of orthographic bodies in any meaningful way, at least when those bodies exist and thus contain sequences of letters that occur frequently. In addition, despite only a significant effect of rime consistency being found by participants, the results appear robust in that they are very similar to other findings, as the absolute magnitude of the effect and the significance levels were very similar to those published in Perry et al. (2002b). Perry et al. also found that the effect of the rime–body relationship was systematic enough to be significant by participants, but not by items.

Although it is not obvious why the results continually fail to reach significance by items but reach significance by participants, one potential reason is that the variance in the responses that participants give to nonwords is high (at least in English). This can be seen in the results of the phoneme–grapheme analysis, where the most and second most common phoneme–grapheme patterns were examined. As can be seen, in all of the groups, there were still many answers that did not contain the first or second most common phoneme–grapheme relationship. Those responses tended to be varied and idiosyncratic, with participants giving many such answers on some items and few on others. This suggests that there is a reasonable amount of item variability, and thus it is one reason that finding significant by-items results might be difficult.

In terms of the phoneme–grapheme contingency effect, the results are similar to those reported in the previous experiment, although the phoneme–grapheme effect in this experiment was much stronger than the rime–body effect, reaching greater levels of significance and being of a larger magnitude. This effect was modulated to some extent by larger relationships, however, with an interaction between phoneme–grapheme contingency and rime consistency being significant by participants. This appeared to be caused by a larger difference between the number of times the most and second most common phoneme–grapheme responses were given in the high rime consistency group, compared to the low rime consistency group, with the low rime consistency group failing to reach significance in post hoc tests, even without using corrected *p* values: High rime consistency, word filler: $t_1(20) = 5.86, p < .001$; $t_2(27) = 4.36, p < .001$; High rime consistency, nonword filler: $t_1(22) = 6.57, p < .001$: $t_2(27) = 5.06, p < .001$; Low rime consistency, word filler, $t_1(20) = 1.81, p = ns$; $t_2(27) = 1.07, p = ns$; Low rime consistency, nonword filler: $t_1(22) = 1.99, p = .059$; $t_2(27) = 1.52, p = ns$.

EXPERIMENT 4

The previous experiment examined the effect of phoneme–grapheme contingency and rime consistency by using a tightly controlled small group of items. Such a small-scale design may have underestimated the effect of phoneme–grapheme contingency, because in order to get a small group where the variables of interest (BodyN and rime consistency) could be factorally manipulated, a set of vowels with many possible spellings were used. In particular, it is probably the case that the nonwords used in this study would be less likely to produce a big difference between the most common and second most common vowel spellings than nonwords with fewer alternative vowel spellings (e.g., /ɪ/). This is because it is likely that participants would more commonly use dominant spellings with nonwords that have fewer alternative vowel spellings. This means that the size of the effect for the set of vowels used in Experiment 2 would be potentially reduced compared to a set of vowels chosen from a more typical spelling distribution. In addition, the /iː/ vowel was disproportionately overrepresented compared to normal text, and people typically spell this vowel with an -ee spelling, rather than the more common -ea spelling (e.g., Barry & Seymour, 1988; Perry et al., 2002b). This may have also caused the size of the phoneme–grapheme contingency effect to be underestimated.

To get around the problem of small item sets and to get additional evidence for the effects of phoneme–grapheme contingency and rime consistency, in this experiment, we used a correlational analysis on a much larger number of items (more than 900). In particular, we asked participants to spell a set of nonwords that included every existing monosyllabic rime in English. The effects of rime consistency and phoneme–grapheme contingency were then examined in correlation analyses using this set of items. Testing participants on such a large number of items is extremely time consuming and demanding. Therefore, it was not possible for us to test as many subjects as in the previous experiments. Whilst clearly the optimum situation is to use a large number of participants even when a large number of items are used (e.g., Frost, Ahissar, Gotesman, & Tayeb, 2003), the approach of using a smaller number of participants when large numbers of items is used is not uncommon and is useful as a complement to small-scale factorial designs (e.g., Ziegler et al., 1998).

Method

Participants

A total of 14 psychology students at Macquarie University participated in the experiment.

Stimuli

A computer program extracted every existing monosyllabic rime from the English CELEX database. A total of 938 nonwords were then constructed by semirandomly appending an existing rime onto an existing onset, with the constraint that the two together would not form a real word. Every monosyllabic rime in the English language was used at least once; 15 were used twice.

Procedure

Participants were tested in small groups of either three or four. They were told that the task was nonword spelling and that all of the stimuli were nonwords. They were asked to write their answers on a writing pad that was provided. The stimuli were read in a random order by a native speaker of Australian English.

Results

All nonwords that were spelled as real words (1.79%) and all nonwords that were obvious sound misperceptions (10.61%) were removed from the data. For each nonword, the different vowel responses were grouped and summed. The percentage of times each vowel was given in each nonword was then calculated by dividing the number of times the vowel was given by the number of participants (14) minus the number of lexicalizations and sound misperceptions. For each vowel group, phoneme–grapheme contingency and vowel predictability based on the consistency of the rime were then used as predictor variables for the regression analysis. Vowel predictability based on the consistency of the rime was calculated for each rime by summing the type frequency of all bodies that used a particular vowel spelling (even if they were different bodies) and dividing that sum by the total type frequency of all of the bodies that used that particular rime. We henceforth refer to vowel predictability based on the consistency of the rime simply as rime consistency.

An example of the data points generated by this procedure can be seen from the nonword /dɑːfs/. The procedure for generating points in the correlation for this nonword was as follows:

1. Different vowel spellings given by participants were separated into groups, and the probability of each vowel spelling being given was then calculated. In this data set, the vowel spelling -*ar* (e.g., *darfs*) was given nine times, and the vowel spelling -*a* (e.g., *daffs*) was given once (there were four lexicalizations and errors). Therefore, there were 10 usable answers. These occurred in two groups (and hence generated two points). One of the groups (-ar) contained 90% of the answers (.9 probability), and the other (-a) contained 10% of the answers (.1 probability).

2. The values of the predictor variables were calculated for each of the two points. Thus we wanted to know the rime consistency and phoneme–grapheme contingency values of /ɑː/→ar and /ɑː/→a, as calculated from a database. Based on the assumption that

frequency-based information influences people's nonword spelling, we would expect that the predictor values would correlate with the answers that people give. Thus, if a sound–spelling correspondence occurs frequently, we would also expect that people would use that correspondence frequently. What is interesting is what form of frequency information is the best predictor (rime consistency or phoneme–grapheme contingency).

3. In our example, the sound–spelling contingency value of /ɑː/→ar was .47. This is so because according to the statistics of Perry et al. (2002a), there are 322 words that use the /ɑː/ phoneme, but only 150 are spelled with -ar (150/322). Based on rime consistency, the value for /ɑː/→ar was .14. The /ɑː/→ar value is .14 because the /ɑːfs/ rime occurs 7 times, but it is only spelt with an -ar body 1 time (*scarfs;* 1/7; vs. *halfs*, *laughs*, etc.). For the second spelling (/ɑː/→a), the vowel sound–spelling contingency was .39 and the rime consistency value was .57. The vowel sound–spelling contingency was .39 because of the 322 words that use an /ɑː/ phoneme, 124 are spelt with an –a (e.g., *bath*). The rime consistency was .57 because /ɑːfs/ is spelt with an -a 4 times (e.g., *staffs*; 4/7).

4. We therefore had two triplets of values from the nonword /dɑːfs/. The first value in each triplet was the probability that a certain correspondence was used by participants for a given nonword (two values in our example, .9 for the –ar spelling and .1 for the –a spelling). The second two values in each triplet were the two predictor variables (rime consistency and phoneme–grapheme contingency). These represented the probability, based on database statistics, that the correspondence would be used. Based on these two triplets, and the triplets generated from all of the other nonwords, a correlation between the values obtained from subjects and the values obtained from the predictor variables was possible.

To examine the data, an overall correlation was first performed on the results. The results appear in Table 4.

As can be seen from Table 4, there was a strong correlation with people's spelling for both the phoneme–grapheme and the rime–body measures. The two measures were highly intercorrelated, however ($r = .76$). Since this was the case, we performed two partial correlations, one examining the effect of phoneme–grapheme contingency with rime consistency partialled out, and the other examining rime consistency with phoneme–grapheme contingency partialled out. The results of the correlation showed that with rime–body consistency partialled out, phoneme–grapheme contingency still significantly correlated with the data, $r_{partial}(2165) = .64$, $p < .001$. The same was true when rime consistency was examined with

TABLE 4
Simple and partial correlations between people's spelling responses and values for vowel contingency and rime consistency

	% Responses	
Predictors	*Simple correlation*	*Partial correlation*
Vowel contingency	.84	.65
Rime consistency	.73	.24

Note: $N = 2168$; all ps < .001.

phoneme–grapheme contingency partialled out. The result was much weaker, however, $r_{partial}(2165) = .24$, $p < .001$.

Overall, the results from this experiment support the results from the previous experiment, that whilst there is an effect of rime consistency that can be found, the effect of phoneme–grapheme contingency is much stronger. Since these results come from over 900 nonwords, with a much greater selection of vowels than in the previous experiment, we take this as strong evidence for the importance of the phoneme–grapheme relationship in skilled spelling.

EXPERIMENT 5

The previous experiments provided evidence that people rely to a greater extent on small-size relationships than on large-size relationships during the process of spelling. However, it may be the case that the larger relationships that people use are not those that we examined. Treiman et al. (2002), for instance, showed that vowel spellings were strongly affected by initial consonants in three situations: (1) When the /ɑ/ vowel[7] was preceded by a /w/ phoneme (e.g., *wart*), the –a grapheme was used far more often than in combination with other consonants; (2) when the /ɝ/ vowel was preceded by the /w/ phoneme (e.g., *work*), the -or grapheme was used disproportionately more often than when it was preceded by other phonemes; and (3) when the /u/ vowel followed a noncoronal consonant (e.g., moon), the -oo grapheme was used more often than when it was following other consonants.

One problem with interpreting Treiman et al.'s (2002) results, in terms of whether the onset is important in determining the vowel spelling, is that the results that they found came from a limited number of cases. This meant that to get enough stimuli for an item analysis, the same sound–spelling patterns were repeated many times. To examine the generality of these results—that is, to see whether onset–vowel effects can be found in more than an extremely limited number of circumstances—we constructed a set of nonwords that varied on onset-vowel consistency, but covered a much larger range of onset–vowel combinations. An example of this are the two nonwords /tʃɜːf/ and /dʒɜːf/. In terms of onset consistency, we would expect people to use the -ur spelling more often in /tʃɜːf/ than in /dʒɜːf/, because of the eight words that start with /tʃɜː/, five use a chur-spelling (e.g., *church*, *churn*), and three use an alternative spelling (e.g., *chirp*) whereas no monosyllabic words start with dur-, but five /dʒɜː/ words start with other spellings (e.g., *dirt*, *dearth*). In addition, to make sure that our participant population could be compared to Treiman et al.'s, we got participants to spell the Treiman et al. set of nonwords that examined onset-vowel effects.

[7]Note that these are the phonetic symbols used in Treiman et al. (2002), where participants spoke Hoosier English. In general, the participants in this study spoke Standard Australian English (a small number tended towards Broad Australian). Both forms of Australian English, in terms of the number of phonemes used, are similar to Received Pronunciation.

Method

Participants

A total of 21 first-year psychology students at Macquarie University participated in the experiment in return for course credit.

Stimuli

A total of 21 pairs of nonwords were constructed that used the same rime but a different onset. This means that each pair had identical rime consistency values, but different onset consistency values. Onset consistency was calculated by counting the number of times an onset-plus-vowel spelling-sound relationship occurs (e.g., chur→/tʃɜ/ in *church*) and then dividing that number by the total number of times words sharing the same onset-vowel phonology occur, including those with different spellings (e.g., chur→/tʃɜ/ in *church*, *chir*→/tʃɜ/ in *chirp*). Thus the onset consistency of chur→/tʃɜ/ would be the number of times that chur→/tʃɜ/ occurs divided by the total number of words that start with /tʃɜ/ (e.g., *church*, *churn*, *chirp*). The onsets were chosen such that the onset consistency values of the spelling-sound relationships with the most frequent vowel grapheme-phoneme relationships were higher in one group than the other, as discussed in the example above.

The stimuli appear in the Appendix B. In addition, all of the nonword stimuli used in Treiman et al. (2002) to examine onset–vowel consistency were added. Finally, 18 nonword fillers were used. None of the fillers used the same vowel as any of the nonwords in either the Treiman et al. nonwords or the new stimulus set.

Procedure

The procedure was the same as that in Experiment 3.

Results

Items that used the /ɑʊ/ vowel were often incorrectly recognized as having a /æl/ sound.[8] We therefore excluded the three matched pairs that contained that sound. In terms of the Treiman et al. (2002) nonwords, 1.5% of the responses were removed from the /wɝ/ group (0.25% incorrect sound perceptions and 1.25% lexicalizations); 3.75% of responses were removed from the /wɝ/ group (0.25% incorrect sound perceptions and 3.5% lexicalizations); and 4.0% of responses were removed from the /u/ group (1.5% incorrect sound perceptions and 2.5% lexicalizations). In terms of the new item set, 5.0% and 7.78% of the responses were removed for being incorrect sound perceptions and lexicalizations, respectively. Mean results for the new item set appears in Table 5.

The results obtained using the Treiman et al. (2002) nonwords were essentially the same as those reported in Treiman et al. (2002). The phoneme /wɑ/ was spelled using the -a grapheme far more often than when it occurred after other consonants (69.92% vs. 1.50%), $t_1(19) = 11.43, p < .001$; $t_2(18) = 11.04, p < .001$. The phoneme /wɝ/ was spelled using an -or

[8] Due to allophonic variation of the phoneme /l/ in the syllable coda position (where it is described as a "dark" /l/) the tongue tip gesture for the /l/ is often lost while the secondary tongue back gesture remains. This tongue back gesture closely resembles the gesture for /w/, thus making /ɑʊ/ sound quite similar to /æl/. This may have caused the high incidence of incorrect recognition. We thank Marjia Tabain for this suggestion.

TABLE 5
Percentage of times that participants used the most and second most common vowel spelling as a function of onset consistency

	Onset consistency	
Vowel contingency	*High*[a]	*Low*[b]
First[c]	66	56
Second[d]	19	21

[a]e.g., chur-. [b]e.g., dur-. [c]e.g., -i-e. [d]e.g., -igh.

grapheme far more often than when it occurred after other consonants (28.39% vs. 0%), $t_1(19) = 3.46$, $p < .005$; $t_2(18) = 8.33$, $p < .001$. The phoneme /u/ after a noncoronal consonant was spelled with an -oo grapheme more often than when it occurred after other consonants (82.56% vs. 37.08%), $t_1(19) = 7.66$, p < .001; $t_2(18) = 6.33$, $p < .001$. The results from the less extreme manipulation produced a slightly different pattern of results. There was an effect of onset consistency (65.76% vs. 55.77%) but the effect was only significant by participants, $t_1(19) = 4.96$, $p < .001$; $t_2 < 1$.

In terms of phoneme–grapheme effects, a 2 × 2 ANOVA with group (high vs. low onset-vowel consistency) and phoneme–grapheme contingency (first vs. second) showed a highly significant effect of phoneme–grapheme contingency (60.76% vs. 20.27%), $F_1(1, 19) = 126.52$, $p < .001$; $F_2(1, 34) = 19.44$, $p < .001$. However, the interaction between the two groups and the number of times the most and second most common vowel was given was significant only by participants, $F_1(1, 19) = 11.93$, $p < .005$; $F_2 < 1$.

Overall, the results suggest that an effect of onset consistency on the spelling of the vowel can be found. This was especially so for the small set of cases examined by Treiman et al. (2002). However, when a less extreme manipulation that covered a broad set of onset-vowel combinations was used, the size of the effect was greatly reduced. This suggests that the effect that Treiman et al. found was probably carried by the particular items they used, and that when a broader (less extreme but still large) manipulation is used, the size of the effect is much smaller. In addition, in this new stimulus set, the phoneme–grapheme contingency effect was much stronger than the effect of onset-vowel consistency. However, there was an interaction between the effects of phoneme–grapheme contingency and onset–vowel consistency, which reached significance by participants, suggesting that people's use of smaller spelling units was somewhat affected by the onset–vowel relationship. Interestingly, the absolute magnitude of the effect from the manipulation that examined a broad range of onsets appeared to be very similar to that found in the rime–body manipulation of Experiment 2.

GENERAL DISCUSSION

The dual-route model of spelling assumes that people rely on two procedures when engaging in spelling: a lexical look-up procedure and a nonlexical procedure that uses phoneme–grapheme rules. In the present research, we investigated whether larger grain sizes also play a role in spelling and how they compare to small-size phoneme–grapheme relationships. In addition, we investigated whether purely orthographic units can explain some of the variance typically attributed to the mapping between sound and spelling.

To examine whether frequency-sensitive orthographic processes and sound–spelling relationships of different sizes affect people's spelling, we examined the effects of phoneme–grapheme contingency, rime consistency, and BodyN. The idea was that BodyN reflects an orthographic variable, rime consistency reflects a variable associated with larger subsyllabic sound–spelling relationships, and phoneme–grapheme contingency reflects a variable associated with smaller sound–spelling relationships. Although these variables are highly correlated, it was possible to examine the effect of each variable independently by using regression techniques on large-scale databases and by using small-scale factorial experiments.

In Experiment 1, we examined a large database of homophone spellings that included all monosyllabic homophones that have two potential spellings. The results suggested that BodyN had an effect on people's homophone spellings over and above the effect of phoneme–grapheme contingency. The same was not true for rime consistency, however. There appeared to be little effect of that variable over and above phoneme–grapheme contingency. The same pattern of results was found in Experiment 2, where extreme item subsets were automatically extracted from the database using a computer program. For these subsets, the proportions of responses in high and low BodyN, rime consistency, and phoneme–grapheme contingency groups were examined, rather than a set of correlations. Experiment 3 examined whether this finding was specific to homophone spellings, which are potentially more prone to lexical effects than are nonwords. In that experiment, manipulations of BodyN and rime consistency were again performed, but their effect on nonwords was examined. The results showed a weak effect of rime consistency that was significant by participants. No significant effect of BodyN was found. There was a strong effect of phoneme–grapheme contingency, however. Experiment 4 replicated these results using a large-scale correlational design instead of the small-scale factorial design used in Experiment 3. Experiment 5 examined whether other types of large relationship might play an important role in spelling. The results showed that there was an effect of onset-vowel consistency on vowel spelling. This effect was of a very similar absolute magnitude to that found in the rime–body experiments, at least when a large selection of onsets was used.

Together, the results support the suggestion of both Perry et al. (2002b) and Treiman et al. (2002) that people tend to exploit both orthographic information as well as sound–spelling consistency when spelling words and nonwords. However, in terms of the specific grain-size used, the results support Perry et al.'s and Barry and Seymour's (1988) interpretation that, in general, people tend to use smaller phoneme–grapheme-sized relationships, with the effect of rime–body relationships being weaker. This is because, throughout the study, the effect of rime consistency was smaller and less reliable than the effect of phoneme–grapheme contingency both in small-scale factorial and in large-scale correlational experiments.

These results differ from those reported by Treiman et al. (2002), who found a somewhat larger effect of rime consistency. However, the stimuli we used in Experiment 2 differ from those used by Treiman et al. in a number of ways. First, the orthographic bodies were not repeated many times. Second, the control nonwords always formed potentially commonly occurring orthographic patterns—Treiman et al.'s control nonwords often would have to be spelled using nonexisting orthographic patterns (e.g., *-ighf*) to score in the same way as the target words. Third, a control group of nonwords was spelled by participants to see whether purely orthographic frequency differences might also have affected the results (a variable highly correlated with rime consistency). Given that our results showed no significant effects of BodyN in nonword spelling, one can be assured that the rime consistency effects that we found were not simply due to variations in BodyN.

The results are most easily interpreted in terms of the dual-route framework (e.g., Barry & Seymour, 1988; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Ellis, 1984). In particular, the BodyN effect found in the analysis of the homophone data provides evidence for some form of frequency-sensitive lexical orthographic body representation. The idea here is that words with common orthographic body representations are easier to spell than those with uncommon orthographic body representations. Apart from this study, such a level of representation is supported by a number of studies that reported evidence for orthographic body units in reading (Forster & Taft, 1994; Montant & Ziegler, 2001; Ziegler & Perry, 1998; Ziegler et al., 2001). In terms of the rime consistency effect, it suggests that the nonlexical phonology–orthography rules would need to include context sensitivity between phoneme units such that more common rime–body relationships would occasionally be used to the detriment of the most common phoneme–grapheme relationships.

In terms of current connectionist models of spelling (e.g., Bullinaria, 1997; Houghton & Zorzi, 1998; Shallice, Glasspool, & Houghton, 1995), the results constrain the type of representation that is used and the way that the mapping from phonology and orthography needs to be learned. This is because, with words, an effect of BodyN was found, whereas no effect of rime consistency was found. The reverse pattern was found with nonwords, however, where an effect of rime consistency was found but no effect of BodyN. This means that, depending on the lexicality of the stimuli, the models would need to produce a double dissociation between these effects. This is unlike the reading model of Zorzi, Houghton, and Butterworth (1998), where facilatory effects of both BodyN and spelling–sound consistency are found with both words and nonwords (which parallels the pattern of data found in English). Only further simulations with these models would be able to determine to what extent these patterns can be simulated. Demonstrating such a dissociation between words and nonwords with connectionist learning models would certainly be impressive.

Another potential difficulty that these models might encounter is related to the statistical learning procedure. In the experiments reported here, both rime and onset-vowel consistency produced effects that were almost identical, both quantitatively and qualitatively. This is different to the pattern of statistical relationships that is obtained when the entire lexical database is analysed. In particular, in a large-scale statistical database analysis, Kessler and Treiman (2001) noted that, in terms of how much the onset of a word can help predict its vowel spelling, "improvements between the vowel and the coda are about seven times the size of corresponding improvements between the vowel and the onset" (p. 610). This has an obvious consequence for models that learn sound–spelling correspondences based on statistical regularities within such a training corpus. These models will need to exhibit similar effects for onset–vowel and vowel–coda relationships despite the fact that the statistical pressures on learning are different for these two types of relationship. However, given that our stimuli were not perfectly matched across experiments, this conclusion should be taken as tentative. Note, however, that in terms of consistency scores, our rime consistency manipulation was of very similar strength to our onset consistency manipulation (rime consistency: .70 vs. .33, difference = .37; onset consistency: .57 vs. .23, difference = .34).

In conclusion, the results of this study show that orthographic frequency effects can be differentiated from sound–spelling effects. In addition, they show that the effects of small-sized sound–spelling relationships are much stronger than large-sized sound–spelling relationships. These results are important for three reasons. First, they suggest that people's spelling needs to be understood in terms of both the relationships between sound and spelling at various grain sizes and purely orthographic frequency at and below the whole word level. Second, they add support to the possibility that some form of orthographic body representation is used when spelling real words, a result concordant with recent findings in reading. Third, they support the claim by Perry et al. (2002b) that people dominantly use phoneme–grapheme units when spelling, and that the use of larger relationships, whilst still occurring, is more limited.

REFERENCES

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.

Bonin, P., Peeremen, R., & Fayol, M. (2001). Do phonological codes constrain the selection of orthographic codes in written picture naming? *Journal of Memory and Language*, *45*, 688–720.

Brown, G. D. A., & Deavers, R. P. (1999). Units of analysis in nonword reading: Evidence from children and adults. *Journal of Experimental Child Psychology*, *72*, 208–242.

Bruck, M., & Treiman, R. (1990). Phonological awareness and spelling in normal children and dyslexics: The case of initial consonant clusters. *Journal of Experimental Child Psychology*, *50*, 156–178.

Bullinaria, J. A. (1997). Modeling reading, spelling, and past tense learning with artificial neural networks. *Brain and Language*, *59*, 236–266.

Campbell, R. (1983). Writing nonwords to dictation. *Brain and Language*, *19*, 153–178.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. C. (2001). DRC: A Dual Route Cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*, 204–256.

Deavers, R., & Brown, G. D. A. (1997). Rules versus analogies in children's spelling. Evidence for task dependence. *Journal of Experimental Child Psychology*, *73*, 208–242.

Ellis, A .W. (1984). *Reading, writing and dyslexia: A cognitive analysis*. Hove, UK: Lawrence Erlbaum Associates Ltd.

Fischer, F. W., Shankweiler, D., & Liberman, I. Y. (1985). Spelling proficiency and sensitivity to word structure. *Journal of Memory and Language*, *24*, 423–441.

Forster, K. I. (2000) The potential for experimenter bias effects in word recognition experiments. *Memory & Cognition*, *28*, 1109–1115

Forster, K. I., & Taft, M. (1994). Bodies, antibodies, and neighborhood-density effects in masked form priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 844–863.

Frost, R., Ahissar, M., Gotesman, R., & Tayeb, S. (in press). Are phonological effects fragile? The effect of luminance and exposure duration on form priming and phonological priming. *Journal of Memory and Language*.

Goodman-Schulman, R., & Caramazza, A. (1987). Patterns of dysgraphia and the nonlexical spelling process. *Cortex*, *23*, 143–148.

Holmes, V. M., & Caruthers, J. (1998). The relation between reading and spelling in skilled adult readers. *Journal of Memory and Language*, *39*, 264–228.

Houghton, G., & Zorzi, M. (1998). A model of the sound–spelling mapping in English and its role in word and nonword spelling. In M. A. Gernsbacher (Ed.), *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society* (pp. 490–501). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

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

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*, *18*, 765–776.

Kreiner, D. S., & Gough, P. B. (1990). Two ideas about spelling: Rules and word-specific memory. *Journal of Memory and Language*, *29*, 103–118.

Montant, M., & Ziegler, J. C. (2001). Can orthographic rimes facilitate naming? *Psychonomic Bulletin & Review*, *8,* 351–356.

Perry, C., Ziegler, J. C., & Coltheart, M. (2002a). A dissociation between orthographic awareness and spelling production. *Applied Psycholinguistics*, *23,* 43–73.

Perry, C., Ziegler, J. C., & Coltheart, M. (2002b). How predictable is spelling? An analysis of sound–spelling contingency in English. *Quarterly Journal of Experimental Psychology*, *55A*, 897–915.

Rapp, B., Epstein, C., & Tainturier, M. J. (2002). The integration of information across lexical and sublexical processes in spelling. *Cognitive Neuropsychology*, *19*, 1–29.

Shallice, T., Glasspool, D. W., & Houghton, G. (1995). Can neuropsychological evidence inform connectionist modelling? Analyses of spelling. *Language and Cognitive Processes*, *10*, 195–225.

Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition*, *55*, 151–218.

Spieler, D. H., & Balota, D. A. (1997). Bringing computational models of word naming down to the item level. *Psychological Science*, *8*, 411–416.

Stone, G. O., & Van Orden, G. (1993). Strategic control of processing in word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *19*, 744–774.

Treiman, R. (1991). Children's spelling errors on syllable-initial consonant clusters. *Journal of Educational Psychology*, *83*, 346–360.

Treiman, R., Kessler, B., & Bick, S. (2002). Context sensitivity in the spelling of English vowels. *Journal of Memory and Language*, *47*, 448–468.

Treiman, R., Mullennix, J., Bijeljac-Babic, R., & Richmond-Welty, E. D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology: General*, *124*, 107–136.

Treiman, R., & Zukowski, A. (1988). Units in reading and spelling. *Journal of Memory and Language*, *27*, 466–477.

Ziegler, J. C., & Perry, C. (1998). No more problems in Coltheart's neighborhood: Resolving neighborhood conflicts in the lexical decision task. *Cognition*, *68*, 53–62.

Ziegler, J. C., Perry, C., Jacobs, A., & Braun, M. (2001). Identical words are read differently in different languages. *Psychological Science*, *12,* 379–384.

Ziegler, J. C., Rey, A., & Jacobs, A. M. (1998). Simulating individual word identification thresholds and errors in the fragmentation task. *Memory & Cognition*, *26*, 490–501.

Zorzi, M., Houghton, G., & Butterworth, B. (1998). Two routes or one in reading aloud? A connectionist dual-process model. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 1131–1161.

Original manuscript received 6 August 2002
Accepted revision received 13 January 2003
PrEview proof published online 25 July 2003

APPENDIX A

Items used and item statistics in the rime consistency manipulation of Experiment 2

Low difference in rime consistency group							*High difference in rime consistency group*						
dVC	*dRC*	*dBN*	*Prop*	*dlogF*	*hom1*	*hom2*	*dVC*	*dRC*	*dBN*	*Prop*	*dlogF*	*hom1*	*hom2*
–.05	–.03	1	.47	–0.72	guise	guys	.09	.33	1	.95	–0.93	surfs	serfs
.08	–.05	0	.72	–2.83	bean	been	–.4	.2	0	.14	–2.31	alms	arms
.08	–.05	0	0	–2.92	quean	queen	–.21	.06	0	0	–1.83	calk	cork
–.18	–.05	0	.53	0.77	heard	herd	–.21	.06	0	.85	0.78	stalk	stork
–.21	–.07	5	.95	1.63	lose	loos	–.27	.29	5	1	1.7	coats	cotes
–.36	–.07	1	.05	–0.97	pi	pie	.01	.06	1	.6	–0.93	hoards	hordes
–.23	–.09	5	1	3.06	shoes	shoos	–.21	.36	5	1	2.71	talks	torques
–.21	–.1	3	.95	3.25	word	whirred	–.08	.39	4	.81	3.44	needs	kneads
–.63	–.12	5	.05	–1	bass	base	–.27	.29	5	.86	–0.7	moats	motes
–.23	–.12	0	1	1.35	shoe	shoo	.35	.11	0	.22	0.87	role	roll
.33	–.14	13	1	2.54	seeds	cedes	0	.62	13	1	2.38	set	sett
–.11	–.16	3	.94	3.16	were	whir	0	.6	3	1	2.25	coins	coignes
–.4	–.17	0	.33	1.26	calves	carves	.01	.07	1	.86	1.04	noughts	naughts
–.08	–.18	13	.95	1.58	are	ah	–.21	.17	13	.14	1.39	grown	groan
0	–.18	1	.35	–3.08	leased	least	.12	.08	1	.05	–2.29	grayed	grade
0	–.3	20	1	1.45	sick	sic	0	.58	19	1	1.1	sack	sac
–.89	–.5	7	.52	1.35	none	nun	–.35	.8	7	.43	0.67	fowls	fouls
–.25	–.6	1	.29	0.92	brooch	broach	–.38	.08	1	.14	0.1	piece	peace
–.36	–.67	0	.1	–0.68	castes	casts	–.31	.05	0	.05	–0.62	hoes	hose
.36	–.84	9	1	0.9	seek	sikh	–.38	.47	9	.7	1.04	wailed	whaled
Average													
–.17	–.22	4.35	.61	.55			–.14	.28	4.4	.59	.49		

Note: dVC = difference in vowel contingency; dRC = difference in rime contingency; dBN = difference in body neighbourhood; Prop = proportion of first homophone responses; dlogF = difference in log frequency; hom1 = homophone 1; hom 2 = homophone 2.

Items removed: the–thee, swats–swots, swat–swot, quad–quod, done–dun, shew–show, kris–crease, lied–lead, wort–wert.

Items used and item statistics in the body neighbourhood manipulation of Experiment 2

Low difference in body neighbourhood group							*High difference in body neighbourhood group*						
dVC	*dRC*	*dBN*	*Prop*	*dlogF*	*hom1*	*hom2*	*dVC*	*dRC*	*dBN*	*Prop*	*dlogF*	*hom1*	*hom2*
.25	.6	–1	.71	–0.92	broach	brooch	0	.67	14	.52	–0.77	tod	todd
0	.18	–1	.65	3.08	least	leased	0	.1	29	.95	2.97	what	watt
.36	.07	–1	.95	0.97	pie	pi	.38	.09	11	1	0.87	low	lo
.05	.03	–1	.53	0.72	guys	guise	.08	.15	7	.85	0.76	slows	sloes
–.38	0	–2	1	2.19	raised	rased	.01	0	12	1	2.37	one	won
.11	.16	–3	.06	–3.16	whir	were	.21	.19	14	.64	–2.88	gnaw	nor
.21	.1	–3	.05	–3.25	whirred	word	.19	.02	12	.11	–2.88	yaws	yours
.63	.12	–5	.95	1	base	bass	.24	.16	8	.9	0.96	aid	aide
.23	.09	–5	0	–3.06	shoos	shoes	.33	.16	13	.1	–2.55	tows	toes
.21	.07	–5	.05	–1.63	loos	lose	.08	.15	7	.1	–1.62	flaw	floor
.89	.5	–7	.48	–1.35	nun	none	.08	.5	8	.38	–1.39	seam	seem
–.36	.84	–9	0	–0.9	sikh	seek	0	.84	16	.76	–0.75	franks	francs
.08	.18	–13	.05	–1.58	ah	are	.17	.19	14	.14	–1.65	hews	hues
–.33	.14	–13	0	–2.54	cedes	seeds	.29	.17	8	.16	–2.54	crewed	crude
0	.3	–20	0	–1.45	sic	sick	0	.34	11	.24	–1.43	gays	gaze
Average													
.13	.23	–5.93	.37	–.79			.14	.25	12.26	.52	–.70		

Note: dVC = difference in vowel contingency; dRC = difference in rime contingency; dBN = difference in body neighbourhood; Prop = proportion of first homophone responses; dlogF = difference in log frequency; hom1 = homophone 1; hom 2 = homophone 2.

Items removed: fie–phi, balm–barm, swots–swats, swot–swat.

Items used and item statistics in the vowel contingency manipulation of Experiment 2

Low difference in vowel contingency group							*High difference in vowel contingency group*						
dVC	*dRC*	*dBN*	*Prop*	*dlogF*	*hom1*	*hom2*	*dVC*	*dRC*	*dBN*	*Prop*	*dlogF*	*hom1*	*hom2*
–.31	.06	2	0	–2.35	pried	pride	.38	.11	4	0	–2.36	lade	laid
–.31	.06	2	.5	–0.53	tied	tide	.83	.19	1	.1	–0.4	led	lead
–.31	.05	0	.05	–0.62	hoes	hose	.25	.09	2	.62	–0.68	coos	coups
–.35	.35	8	.86	2.31	hold	holed	.38	.35	8	1	2.04	mean	mien
–.35	.35	8	1	2.29	sold	soled	.63	.29	12	1	2.13	sane	seine
–.35	.8	7	.52	–0.97	fowl	foul	.62	.75	4	.05	–1.07	brake	break
–.35	.8	7	.43	0.67	fowls	fouls	.57	.78	7	1	0.4	poof	pouf
–.35	.75	7	.42	–0.7	fowled	fouled	.95	.71	10	.05	–0.73	gilt	guilt
–.35	.09	3	.47	0.48	polled	poled	.83	.19	1	.59	0.45	red	read
–.36	–.07	1	.05	–0.97	pi	pie	.01	.06	1	.6	–0.93	hoards	hordes
–.36	–.67	0	.1	–.68	castes	casts	.1	0	0	.7	–0.68	sauce	source
–.38	.47	9	.7	1.04	wailed	whaled	.53	.45	10	1	1.11	droop	drupe
–.38	.19	8	.19	–0.75	lain	lane	.35	.21	4	.22	–0.7	bur	burr
–.38	.19	8	.71	1.89	main	mane	.59	.21	7	.65	1.76	danes	deigns
–.38	.19	8	.8	1.62	pain	pane	.44	.14	7	.95	1.56	tie	thai
–.38	.19	8	.29	0.07	plain	plane	.38	.24	9	1	0.18	lees	leas
–.38	.19	8	.13	–0.33	wain	wane	.53	.25	7	.85	–0.3	loots	lutes
–.38	.33	7	.24	0.18	plained	planed	.86	.4	4	.19	0.09	blond	blonde
–.38	.18	6	.76	0.55	mains	manes	.53	.14	5	.95	0.57	loot	lute
–.38	.18	6	.8	0.75	pains	panes	.5	.18	5	.47	0.82	stayed	staid
–.38	.18	6	.24	–0.33	plains	planes	.27	.17	2	.43	–0.37	urns	earns
–.38	.18	6	.29	–0.78	wains	wanes	.48	.17	3	.29	–0.62	throne	thrown
–.38	.08	4	.35	–0.21	ail	ale	.24	.04	2	.57	–0.18	waits	weights
–.38	.08	4	.63	0.79	bail	bale	.53	.14	5	1	0.83	shoot	chute
–.38	.08	4	.95	0.69	hail	hale	.36	.07	4	.95	0.68	die	dye
–.38	.08	4	.29	–0.54	mail	male	.08	.08	4	.48	–0.48	creaks	creeks
–.38	.08	4	.11	–1.33	pail	pale	.25	.12	3	.53	–1.3	coo	coup
–.38	.08	4	.76	–1.09	sail	sale	.08	.09	2	.67	–1.09	steal	steel
–.38	.08	4	.81	0.28	tail	tale	.81	0	1	.71	0.34	use	used
–.38	.09	3	.57	0	ails	ales	.01	.06	1	.14	–0.03	isle	aisle
–.38	.09	3	.65	–1.19	bails	bales	.1	.13	2	.23	–1.23	gauds	gourds
–.38	.09	3	.15	–1.22	mails	males	.05	.06	3	.27	–1.26	rues	ruse
–.38	.09	3	.25	0.57	pails	pales	.09	.11	1	.95	0.47	surge	serge
–.38	.09	3	.6	–1.02	sails	sales	.08	.09	2	.6	–1	heal	heel
–.38	.09	3	.86	–0.31	tails	tales	.37	.15	3	.4	–0.44	spar	spa
–.38	.08	1	.14	0.1	piece	peace	.35	.11	0	.86	0.23	pole	poll
–.4	0	2	.2	–1	calve	carve	.4	0	2	.1	–1.38	barm	balm
–.4	0	1	.1	–1.74	calved	carved	.83	.19	1	0	–1.65	bred	bread
–.4	–.17	0	.33	1.26	calves	carves	.07	.04	1	.75	1.23	shorn	shawn
–.4	.2	0	.14	–2.31	alms	arms	.5	.18	5	.05	–2.28	rayed	raid
–.41	.06	3	.15	–0.88	boos	booze	.08	.08	4	0	–0.78	wreaks	reeks
–.51	.02	1	.42	1.98	oh	owe	.54	.05	2	1	1.76	wise	whys
–.63	–.12	5	.05	–1	bass	base	.47	.13	8	.9	–1.04	wood	would
–.66	.03	1	.19	–1.61	sighs	size	.12	.08	1	.1	–1.41	spayed	spade
–.67	.15	4	.1	–2.58	wights	whites	.29	.17	8	.16	–2.54	crewed	crude
–.67	.13	4	.1	–2.2	wight	white	.11	.12	8	.62	–2.22	shear	sheer
–.67	.13	4	0	–1.85	bight	bite	.25	.11	5	.1	–1.84	strait	straight
–.67	.13	4	.76	2.02	might	mite	.37	.15	3	1	2.2	far	fa
.67	.15	4	.11	–1.99	bights	bites	.27	.23	3	0	–1.95	purl	pearl
–.67	0	0	.86	2.13	find	fined	.01	.01	1	1	2.27	door	dour
–.67	0	0	.95	2.6	mind	mined	.06	.02	1	.94	2.63	ward	warred
–.89	–.5	7	.52	1.35	none	nun	.36	–.84	9	1	0.9	seek	sikh
Average													
–.44	.12	4.08	.42	0.14			.37	.15	4	.55	–0.16		

Note: dVC = difference in vowel contingency; dRC = difference in rime contingency; dBN = difference in body neighbourhood; Prop = proportion of first homophone responses; dlogF = difference in log frequency; hom1 = homophone 1; hom 2 = homophone 2.

APPENDIX B

Items used in the rime consistency manipulation of Experiment 3

				Nonword filler condition		Word filler condition	
Dominant spelling	*BN+*	*BN–*	*Rime consistency*	*Proportion 1st vowel spelling*	*Proportion 2nd vowel spelling*	*Proportion 1st vowel spelling*	*Proportion 2nd vowel spelling*
sporce	6	1	.5	.86	.09	.80	.10
dorce	6	1	.5	.87	.04	.95	0
gorch	3	1	.75	.52	.22	.90	.10
norch	3	1	.75	.81	.05	.85	.10
nurge	4	3	.44	.30	.52	.26	.53
gurge	4	3	.44	.52	.26	.50	.20
glabe	1	0	1	.83	.17	.90	.05
nabe	1	0	1	.77	.23	.80	.10
blape	11	0	.92	.78	.17	.85	.10
fape	11	0	.92	.83	.17	.90	.05
glame	12	3	.8	.40	.50	.63	.32
zame	12	3	.8	.61	.35	.89	.11
drave	15	1	.94	.67	.29	.95	.05
blave	15	1	.94	.82	.14	.83	.11
churf	3	1	.75	.57	.26	.72	.17
kurf	3	1	.75	.61	.26	.45	.45
furnt	1	0	.5	.26	.70	.20	.50
slurnt	1	0	.5	.22	.57	.42	.42
furp	2	1	.5	.41	.36	.30	.40
gurp	2	1	.5	.52	.26	.15	.40
burb	2	2	.5	.83	.13	.55	.15
glope	12	1	.92	.61	.17	.60	.25
fope	12	1	.92	.68	.14	.85	.15
doke	12	4	.63	.77	.05	.95	.05
snoke	12	4	.63	.85	.10	.82	.09
bleace	5	3	.5	.41	.32	0	.8
feace	5	3	.5	.25	.40	.2	.5
neethe							
deethe							
gurp							
Average	6.52	1.48	.70	.61	.26	.64	.23
chort	7	1	.28	.78	.06	.93	0
dort	7	1	.28	.76	.10	.70	0
blorth	2	0	.5	.75	.00	.85	0
gorth	2	0	.5	.74	.05	.90	.05
nurt	3	3	.21	.22	.74	.16	.63
plurt	3	3	.21	.59	.32	.55	.35
gafe	2	1	.5	.60	.35	.90	.10
spafe	2	1	.5	.63	.32	.80	.20
skane	13	21	.31	.77	.18	.40	.50
glane	13	21	.31	.26	.65	.80	.20
drale	14	18	.4	.09	.83	.00	.90

(continued overleaf)

APPENDIX B (continued)

Dominant spelling	*BN+*	*BN–*	*Rime consistency*	*Nonword filler condition*		*Word filler condition*	
				Proportion 1st vowel spelling	*Proportion 2nd vowel spelling*	*Proportion 1st vowel spelling*	*Proportion 2nd vowel spelling*
zale	14	18	.4	.39	.57	.70	.30
jate	16	4	.62	.88	.13	.56	.39
drate	16	4	.62	.62	.33	.55	.45
nurk	4	3	.27	.35	.55	.21	.68
hurk	4	3	.27	.55	.41	.30	.45
churve	1	5	.17	.70	.30	.60	.10
snurve	1	5	.17	.10	.76	.10	.60
nurm	0	5	0	.22	.67	.15	.55
plurm	0	5	0	.19	.76	.05	.74
sturth	0	2	0	.24	.24	.11	.47
glone	13	4	.48	.53	.26	.35	.25
vone	13	4	.48	.91	.05	.00	.00
spote	9	11	.45	.91	.09	.90	.10
chote	9	11	.45	.71	.19	.60	.40
dreep	5	13	.28	.21	.58	.05	.75
feep	5	13	.28	.36	.55	0	.74
dreef							
geef							
glurth							
Average	6.59	6.67	.33	.52	.37	.45	.37

Note: Nonwords *italicized* were removed from the analysis; BN+ = body neighbourhood counts for all bodies that share the same vowel; BN– = body neighbourhood counts for all bodies that used a different vowel.

Items used in the body neighbourhood manipulation of Experiment 3

				Nonword filler condition		*Word filler condition*	
Dominant spelling	*BN+*	*BN–*	*Rime consistency*	*Proportion 1st vowel spelling*	*Proportion 2nd vowel spelling*	*Proportion 1st vowel spelling*	*Proportion 2nd vowel spelling*
sturl	4	0	.33	.14	.57	.25	.25
murl	4	0	.33	.14	.59	.11	.50
glope	12	1	.92	.64	.18	.65	.25
fope	12	1	.92	.75	.10	.85	.15
dorce	6	1	.5	.86	.09	.90	.10
sporce	6	1	.5	.87	.04	.95	0
chorn	13	1	.46	.90	.05	1	0
zorn	13	1	.46	1.00	.00	.21	0
slake	20	0	.83	.91	.00	1	0
nake	20	0	.83	.76	.19	.8	.13
zeet	15	11	.56	.14	.68	0	.85
leet	15	11	.56	.48	.43	.2	.7
steech	10	5	.67	.32	.55	.16	.74
theech	10	5	.67	.48	.48	.15	.75
pleek	14	10	.52	.52	.35	.45	.55
feek	14	10	.52	.23	.59	.05	.74
Average	11.75	3.63	.60	.57	.31	.48	.36
furnt	1	0	.5	.26	.70	.20	.50
slurnt	1	0	.5	.22	.57	.42	.42
plobe	4	0	.8	.76	.05	.90	.05
sobe	4	0	.8	.76	.05	.70	.15
horb	1	1	.5	.79	.05	.79	0
thorb	1	1	.5	.95	.00	.94	0
gorch	3	1	.75	.52	.17	.90	.10
norch	3	1	.75	.77	.05	.85	.05
glabe	1	0	1	.83	.17	.90	.05
nabe	1	0	1	.77	.23	.80	.10
bleece	5	3	.5	.41	.32	0	.80
feece	5	3	.5	.25	.40	.20	.50
meesh	1	0	.5	.29	.50	.05	.75
dreesh	1	0	.5	.22	.33	.15	.54
neeve	4	3	.4	.30	.13	0	.35
dreeve	4	3	.4	.33	.33	.12	.59
Average	2.5	1	.62	.53	.25	.50	.31

Note: BN+ = body neighbourhood counts for all bodies that share the same vowel; BN– = body neighbourhood counts for all bodies that used a different vowel.

Items used in Experiment 4

Dominant spelling	*Onset+*	*Onset–*	*Onset consistency*	*Proportion 1st vowel spelling*	*Proportion 2nd vowel spelling*
durf	0	5	0	.37	.21
fobe	3	10	.23	.05	.85
fape	15	16	.48	.05	.85
gurch	0	7	0	.4	.25
gobe	0	12	0	0	1
harb	11	12	.48	0	1
feask	4	10	.29	.17	.5
furb	6	10	.38	.35	.25
horch	6	22	.21	0	.88
boolm	3	6	.33	.13	.27
dorch	2	10	.17	0	1
gorp	6	13	.32	.06	.72
jeab	1	5	.17	.85	0
carth	12	13	.48	.11	.67
loke	7	18	.28	.17	.78
neast	4	9	.31	.38	.19
sove	0	15	0	.33	.56
durl	0	5	0	.38	.08
douk					
goub					
toup					
Average	4.44	11	0.23	.21	.56
churf	5	3	.63	.47	.41
nobe	9	7	.56	0	1
dape	11	5	.69	0	1
hurch	5	8	.38	.35	.35
dobe	13	3	.81	0	.89
marb	17	8	.68	.05	.95
beask	15	10	.6	.22	.61
burb	14	12	.54	.2	.65
foolm	2	2	.5	0	.85
corch	16	16	.5	.38	.31
forch	16	17	.48	0	.9
norp	5	7	.42	0	.94
cheab	3	6	.33	.74	0
marth	17	8	.68	.06	.94
doke	11	3	.79	.06	.76
meast	8	3	.73	.27	.27
tove	7	8	.47	.1	.9
nurl	2	3	.4	.6	.1
bouk					
foub					
foup					
Average	9.78	7.17	.57	.19	.66

Note: Nonwords *italicized* were removed from the analysis; Onset+ = onset–vowel counts for all onset–vowels that use the most common vowel; Onset– = onset–vowel counts for all onset–vowels that do not use the most common vowel.