Journal of Experimental Psychology:
Learning, Memory, and Cognition
2002, Vol. 28, No. 1, 207–214

0278-7393/02/$5.00 DOI: 10.1037//0278-7393.28.1.207

Theories of Word Naming Interact With Spelling–Sound Consistency

Eamon Strain
Anglia Polytechnic University

Karalyn Patterson
Medical Research Council Cognition and Brain Sciences Unit

Mark S. Seidenberg
University of Wisconsin—Madison

In a previous study (E. Strain, K. Patterson, & M. S. Seidenberg, 1995), the authors concluded that word naming is characterized by an interaction between spelling–sound typicality and word imageability, thus implicating a role for word meaning in the naming process. J. Monaghan and A. W. Ellis (2002) reject E. Strain et al.'s conclusion, arguing that it is age of acquisition (AoA) and not imageability that interacts with spelling–sound typicality. In this article, the authors question their alternative interpretation (a) by raising a number of conceptual and methodological issues germane to this debate and (b) by presenting new data that confirm a significant interaction between spelling–sound typicality and imageability in word-naming latencies, an interaction that is reliable when word AoA is controlled in a regression analysis.

In 1995, we reported three experiments in which normal adult readers were invited to name aloud single words with either typical or atypical spelling–sound correspondences (Strain, Patterson, & Seidenberg, 1995). The novel contribution of this study was a demonstration that the well-known impact of spelling–sound typicality on naming lower frequency words interacted with rated-word imageability, a variable widely accepted to index an aspect of word meaning. Participants were slower and more error prone when naming exception words with abstract meanings (e.g., *scarce*) than when naming either abstract regular words (e.g., *scribe*) or imageable exception words (e.g., *soot*). We interpreted these findings as indicating that semantic representations are automatically activated in the course of translating orthography to phonology (O→P), and that when this translation process is slow or noisy as in the case of low-frequency exceptions, naming of words with richer semantic representations (i.e., high-imageability words) is most likely to benefit from the additional activation of phonology by word meaning.

Monaghan and Ellis (2002) report a somewhat similar study, except that the variable in focus is not imageability but *age of acquisition* (AoA): the age at which a given word is typically learned. The central goal of the Monaghan and Ellis article is to demonstrate that spelling–sound typicality interacts with AoA, such that readers are slower and more error prone to name late-acquired exception words than to name either early acquired irregular words or late-acquired regular items. An additional aim of their article is to establish that spelling–sound consistency does not interact with imageability.

Effects of AoA on skilled reading raise important issues that are the focus of continuing research and debate (e.g., Zevin & Seidenberg, in press). Our goal in writing this article was not to assess Monaghan and Ellis's (2002) evidence for this effect; rather, it was to address their rejection of the central conclusion of the Strain et al. (1995) article, that spelling-sound typicality interacts with imageability. This finding is of some theoretical importance insofar as it is consistent with properties of the triangle model developed by Seidenberg and McClelland (1989) and Plaut, McClelland, Seidenberg, and Patterson (1996).

The structure of this article is as follows. We need to note several respects in which Monaghan and Ellis (2002) have either misinterpreted or miscriticized the original Strain et al. (1995) study and also several potential problems in the Monaghan and Ellis experiments; but, to focus on issues rather than on disagreements, these points are incorporated under three conceptual or methodological questions to which this debate is germane. Following these questions, we present data from a new experiment that confirms an interaction between spelling–sound typicality and word imageability in word-naming latencies—an interaction that is significant even when the AoA of the stimulus materials is controlled in a regression analysis.

The Questions

1. What Is the Interrelationship Between Various Factors That Might Interact With Spelling–Sound Typicality?

The primary factor originally (e.g., Seidenberg, Waters, Barnes, & Tanenhaus, 1984) and repeatedly (e.g., Jared, McRae, & Seidenberg, 1990) shown to interact with spelling–sound consistency

Eamon Strain, Department of Psychology, Anglia Polytechnic University, Cambridge, England; Karalyn Patterson, Cognition and Brain Sciences Unit, Medical Research Council, Cambridge, England; Mark S. Seidenberg, Department of Psychology, University of Wisconsin—Madison.

We thank Jason Zevin for helpful discussions and for providing normative information, and we thank Josie Monaghan for supplying additional data for our analyses.

Correspondence concerning this article should be addressed to Eamon Strain, Department of Psychology, School of Applied Sciences, Anglia Polytechnic University, Cambridge, CB1 1PT, England. E-mail: e.p.strain@apu.ac.uk

is word frequency, usually manipulated with reference to one of several published sets of written word-frequency norms. It is well-known, however, that this variable correlates highly with other measures that are themselves correlated with efficiency of O→P processing, including spoken-word frequency, rated-word familiarity, AoA, imageability, and probably others as well. Even within the more restricted domain of written-word frequency itself, the best measure is not obvious: Monaghan and Ellis (2002) correctly note that the Strain et al. (1995) stimulus words, which were frequency manipulated and matched on the basis of the Kučera & Francis (1967) norms, are not well matched on CELEX (Baayen, Piepenbrock, & Van Rijn, 1993) frequencies, and they imply, in the absence of any direct evidence, that this lack of matching contributed to the apparent imageability effect. Zevin and Seidenberg (in press) make the same point about studies of AoA (including Monaghan and Ellis's Experiment 2) in which stimuli that were matched on Kučera–Francis frequency differ on other frequency measures such as CELEX or the educator's word frequency guide (WFG; Zeno, 1995), such that early AoA words were higher in frequency than late AoA words. They present regression analyses based on three large-scale studies of naming and lexical decision that yield no reliable effects of AoA independent of other correlated stimulus properties such as frequency, length, and imageability.

We do not claim to have a good solution to the thorny problems of intercorrelated variables or discrepancies between different measures of the same variable, but we can provide direct evidence of the presence or absence of a potential confound of word frequency in the studies of concern here. We have redone the by-items analyses for both Strain et al.'s (1995) Experiment 2 and Monaghan and Ellis's (2002) Experiment 2, separately adding both CELEX and WFG frequency values (log transformed) as covariates in the analysis and controlling for AoA in the Strain et al. data. The result is that neither of these additional frequency measures eliminates the significant interaction between regularity and imageability in the Strain et al. data. Interaction F ratios were as follows: $F(1, 58) = 6.81$, $MSE = 1{,}075.3$, $p < .025$, for WFG; $F(1, 58) = 5.81$, $MSE = 1{,}070.9$, $p < .025$, for CELEX. The picture is different when the interaction between regularity and AoA is considered. First, this interaction is not significant in the by-items analysis, even before frequency is controlled, $F(1, 68) = 1.55$, $MSE = 1{,}885.2$, $p = .22$. Furthermore, both frequency measures, when added as covariates, eliminate the (initially borderline) main effect of AoA. AoA F ratios were as follows: $F(1, 75) = 1.75$, $MSE = 1{,}800.2$, $p = .21$, for WFG; $F(1, 75) = 3.5$, $MSE = 1{,}861$, $p = .07$, for CELEX. These analyses support the conclusion that although imageability and frequency account for largely independent portions of the latency variance, AoA is confounded with frequency in its prediction of naming latencies. They also show that the interaction between AoA and regularity, on which Monaghan and Ellis focus so heavily in their discussion, receives only tenuous support from their data.

2. *How Should One Define Spelling–Sound Typicality and Therefore Classify Words as Regular or Irregular?*

There is no consensual answer to this question. Monaghan and Ellis (2002) opted for a definition based on the relative numbers of friends and enemies within a target word's *body* neighborhood (the body being the vowel and terminal consonant or consonants of a monosyllabic word). To use their own examples, *hake* is a regular and consistent word because all —*ake* words rhyme with it; *wand* is an exceptional, inconsistent word because the majority of words ending in —*and* are pronounced like *band*. Monaghan and Ellis criticize Strain et al. (1995) for using a mixed set of criteria for assigning words to regular and exceptional categories and for the fact that the Strain et al. criteria classified a word like *chasm* as irregular on the basis of its having an atypical pronunciation of the segment *ch*. In terms of the Monaghan and Ellis word-body criterion, *chasm* is regular and consistent because its few body neighbors (like *spasm*) rhyme with it.

We have three responses to this point. The first is that the Strain et al. (1995) criterion for atypical spelling–sound words was not a hybrid, as Monaghan and Ellis (2002) suggest, but rather a two-stage classification system. Every irregular word contained an orthographic segment with a phoneme pronunciation that is statistically atypical for its context (e.g., the *a* in *caste*, the *oo* in *soot*, or the *ch* in *chasm*), and every regular word contained segments with only the most typical phoneme pronunciations. What made our classification scheme more complex and apparently led to Monaghan and Ellis's confusion was that irregular words meeting our first criterion were excluded from the list if the atypically pronounced segment happened to be typical for the word's body neighborhood. Thus, although we included *soot* as irregular, we would not include *hook* because even though it contains the atypical pronunciation of the single segment *oo*, when *oo* is followed by *k*, this is the statistically typical pronunciation for the body neighborhood. However, we only considered the consistency of word body in this second stage of our classification system if that was where the atypical segment was located: this is a two-stage classification, not a mixed or hybrid or ambiguous one.

Second, although we along with many other researchers in the field are impressed by the influence of body consistency on speed and accuracy of O→P translation for lower frequency words, we challenge Monaghan and Ellis's (2002) faith in bodies as the only influential level of spelling–sound typicality. We reassert our claim that the word *chasm* has an atypical spelling–sound relationship, despite its typical body, and wonder what account Monaghan and Ellis would give of the fact that surface alexic patients are so prone to mispronounce the first phoneme of this word in its more typical fashion (as in *charm*).

Finally, we disagree with some of Monaghan and Ellis's (2002) classifications of exception words. They consider *bead* an exception word because there are more words in this neighborhood that rhyme with *head* (10 words) than with *bead* (5 words), but this is not a massive body bias toward the irregular and furthermore we would argue that the correct pronunciation of *bead* is supported by *beat*, *beam*, *beach*, and so on. They also consider *wand* to be exceptional. Here it is true that the body neighborhood (*band*, *hand*, *land*, etc.) is overwhelmingly ganged up against *wand*, but if the consistency of the onset+vowel combination plays a role, then we would expect *wand* to be supported by *want*, *wander*, *wad*, *watch*, and so on. We have in fact published data showing that both surface alexic patients and the Plaut et al. (1996) connectionist network are significantly more successful at pronouncing *wa*— words like *wand* correctly than matched words with atypical bodies but without typical onset+vowel combinations (Patterson et al., 1996).

3. Does Meaning Interact With Regularity, and What Do Interactions Mean?

In their Experiment 4, Monaghan and Ellis (2002) not only collected new data by using the Strain et al. (1995) stimuli but also reanalyzed the original (Experiment 2) Strain et al. data and used the results as a principal basis for rejecting the Strain et al. conclusion. In the original Strain et al. data, with AoA entered as a covariate, the significant main effect of imageability disappears. According to Monaghan and Ellis (2002)

> the form of interaction claimed by Strain et al. [1995] was that imageability affected the naming of low-frequency, irregular but not regular words. Such an interaction would require a main effect of imageability, which seems not to occur once AoA is controlled. (p. 191)

Our response to this statement comes in three parts.

(1) Monaghan and Ellis are simply not correct to assert that a significant interaction (even of the noncrossover type) requires a significant main effect.

(2) Although the original Strain et al. Experiment 2 did produce a significant main effect of imageability, we never predicted such a main effect; we only, and explicitly, predicted the observed interaction: "We therefore predicted that low-frequency exception words, which have had the smallest impact on setting weights for orth-to-phon translation, should be the items most likely to reveal a role for word meaning in word naming" (Strain et al., 1995, p. 1141).

(3) Monaghan and Ellis themselves confirm that in the original Strain et al. data with AoA now controlled, the interaction between spelling–sound typicality and imageability remains significant, although they mistakenly dismiss this result, as described in the first point, above. Lack of regard for this reliable interaction seems puzzling in light of the title of their article.

When Monaghan and Ellis (2002) analyzed the new data that they collected using the Strain et al. (1995) stimuli, once again the predicted interaction was significant in the by-subjects analysis. Neither the main effect for consistency nor its interaction with imageability was significant in the by-items analysis, however. We suggest that this can be explained by a lack of power in this analysis. As Monaghan and Ellis noted early in their article, when close stimulus matching has been performed, as in this case, by-items analyses lack power and are more likely to lead to Type II errors (see also Raaijmakers, Schrijnemakers, & Gremmen, 1999). Furthermore, Monaghan and Ellis's participants made more errors than those in the Strain et al. study, and these errors, which of course were excluded from the latency analysis, were more heavily concentrated in the exception-word condition. This suggests that the consistency effect (and interactions with it) might most clearly be seen in the error data. This interpretation is confirmed when the errors from this experiment are subjected to a by-items analysis of variance, which yields a strong consistency effect, $F(1, 60) = 21.71$, $MSE = 16.85$, $p < .01$, and a very reliable interaction of the predicted form between consistency and imageability, $F(1, 60) = 10.22$, $MSE = 16.85$, $p < .01$. This interaction remains significant when AoA is added as a covariate, $F(1, 59) = 10.3$, $MSE = 17.05$, $p < .01$. AoA is not significant in this analysis. When errors are considered, then, the new Monaghan and Ellis data actually provide further support for the Strain et al. conclusions.

This brings us to Monaghan and Ellis's (2002) Experiment 3, which yielded a significant consistency effect but no interaction with imageability. We suggest that the former is not a genuine effect of consistency but is the result of uncontrolled nuisance variables and that the lack of an interaction with imageability is therefore unsurprising. Earlier, we highlighted a number of items categorized by Monaghan and Ellis as inconsistent by word body, which we suggest are either statistically typical in terms of smaller segments like *ea-* or quite consistent by onset+vowel. Over half of their "inconsistent" items could be so classified, and we refer to these as *ambiguously inconsistent* (see Appendix A). Furthermore, although several studies of word naming with very large stimulus sets (Seidenberg & Waters, 1989; Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welty, 1995) have established that initial phoneme is the largest single source of unique variance in word naming as measured by a voice key, Monaghan and Ellis did not control for this factor. For example, their set of inconsistent words in Experiment 3 contained twice as many items with fricative onsets as did their "matching" set of consistent words. In addition, the inconsistent words were on average longer than the regular words (4.7 vs. 4.2 letters) and had fewer orthographic neighbors in terms of Coltheart's *N* (7.0 vs. 10.3). When these factors have been controlled for in a by-items analysis of covariance (ANCOVA), the initially highly reliable consistency effect, $F(1, 68) = 17.22$, $MSE = 1{,}582.8$, $p < .01$, observed by Monaghan and Ellis is eliminated, $F(1, 56) = 3.53$, $MSE = 935.76$, $p = .07$. Two further by-items analyses confirm that the weakness of the consistency manipulation is due to the subset of items we have identified as ambiguously inconsistent. In the first we included all the consistent items, along with only the clearly inconsistent items. With the ambiguous items removed, this reduced set of items produced a significant consistency effect, $F(1, 48) = 19.23$, $MSE = 1{,}560.3$, $p < .01$, which remained highly reliable when the confounding variables were controlled for in the ANCOVA $F(1, 30) = 7.65$, $MSE = 896.42$, $p < .01$. On the other hand, when the data were reanalyzed including the consistent items along with the ambiguously inconsistent words only, the initially significant consistency effect, $F(1, 52) = 6.24$, $MSE = 1{,}371.76$, $p < .05$, is completely eliminated when the confounding factors are controlled for in the ANCOVA $F(1, 40) = 0.025$, $MSE = 455.8$, $p = .87$.

Of the 18 items Monaghan and Ellis (2002) categorize as low-imageability inconsistent, only 6 appear to have truly atypical O→P translations. This is simply too few items to allow any meaningful or reliable investigation of the interaction between consistency and imageability. In the following experiment, we show that, with a large set of items that produce a clear and unconfounded consistency effect, word naming is characterized by a significant interaction between regularity and imageability, even when AoA has been controlled.

The Experiment

In two key experiments discussed in the previous section (i.e., Experiment 3 in Monaghan and Ellis, 2002, and Experiment 2 in Strain et al., 1995), the approach taken during stimulus generation was to match sets of stimuli on various factors (e.g., initial phoneme, length, frequency), while systematically manipulating two or more additional factors that were the focus of theoretical interest (regularity, imageability, or AoA). It could be argued that the constraints imposed by this approach resulted in small and somewhat atypical stimulus sets. Furthermore, despite the effort expended in attempting to match the stimuli on what were perceived to be critical variables, it has been possible in both cases to identify factors on which the stimulus sets were not well matched (e.g., initial phoneme in Monaghan and Ellis and AoA in Strain et al.). In this experiment, we took a different approach, one that permits the selection of a larger, more typical set of words, while controlling for key factors through a combination of stimulus matching and regression techniques.

Method

Participants

Twenty-four undergraduate students at Anglia Polytechnic University (APU) volunteered to take part in the experiment. The participants were aged between 19–45 years and had normal or corrected-to-normal vision; about two thirds were women.

Materials

The stimuli used in the analysis of this experiment were a subset of those used in Strain and Herdman (1999), selected from a pool of 160 monosyllabic words for which imageability ratings had been collected.

Definition of Consistency (Regularity)

A word was classified as an exception if its pronunciation was inconsistent with the most typical pronunciation of an orthographic segment corresponding to a single phoneme (Venezky, 1970). As in Strain et al. (1995), however, and as we described earlier, we excluded from the exception set any word belonging to an orthographic body neighborhood in which the pronunciation of a large majority of the members conflicts with typicality as defined by the first criterion. For example, this excluded items with word bodies such as —*ead*, —*old*, and —*ind*. Exception words with no body neighbors (i.e., unique words) were included only if they had a Coltheart's *N* of at least 2, to ensure that they were not very orthographically strange. Included among the original Strain and Herdman (1999) exceptional stimuli were a number of words with a *wa* onset, which, as noted in the previous section, are quite consistent by word-onset neighborhood. In line with our earlier argument that such items may not have particularly atypical O→P mappings, these words (and their matched items) were not included in the analysis of the data, although our participants named them during the experiment.

A word was classified as regular if both (a) its pronunciation corresponded to the most typical pronunciation of all of its orthographic segments and (b) it belonged to a largely consistent orthographic body neighborhood. Each regular word was matched with an exception word in terms of initial phoneme, which was coded for the regression analysis in terms of 10 binary variables (following Treiman et al., 1995). Regular and exception words were also matched as groups on length (number of letters), imageability (ratings from Strain & Herdman, 1999), CELEX written word frequency, Coltheart's *N*, and positional bigram frequency (both of the latter two measures calculated from the Medical Research Council Psycholinguistics Database; Coltheart, 1981). We also ensured that imageability was evenly distributed across the regular and exception word sets, so that they could be divided into high- and low-imageability sets with equal numbers of items and with matched mean imageability ratings. In all, there were 75 pairs of regular and exception items in the original stimulus set, of which 60 pairs were used in the final analysis stage of this experiment. The characteristics of these 120 items are shown in Table 1 (actual items are in Appendix B).

Table 1
Characteristics of Items Used in the Experiment (With Standard Deviations)

Factor	Regular	Exception
Imageability	463 (151)	468 (154)
High imageability	593 (57)	599 (48)
Low imageability	332.3 (90)	336.8 (103)
Frequency (CELEX)	149 (125)	149 (127)
Number of letters	4.6 (0.7)	4.7 (0.9)
Coltheart's *N*	7.1 (5.5)	6.7 (5.9)
Positional bigram frequency	3,795 (2,926)	3,780 (2,452)
Age of acquisition	376 (103)	395 (112)

AoA

To control for any confound due to AoA at the data-analysis stage, we collected ratings on this variable for the complete set of 150 items from 60 undergraduate APU students. Each participant rated 75 words, presented on sheets in four different randomized orders, so that each word was rated by 30 participants. The mean ratings, which were quite similar for regular and exception words, are presented in Table 1.

Apparatus

For the word-naming experiment, the stimuli were presented in the center of a 14-in. (35.56-cm) color screen, using PsyScope (Cohen, MacWhinney, Flatt, & Provost, 1993) running on an Apple Macintosh Quadra 650 computer. The words were presented in black, lowercase font, New York 24 point on a white screen placed approximately 60 cm from the participant. Naming responses were recorded using a headset microphone connected to the voice-key port of a Carnegie Mellon University button box (see Cohen et al., 1993, for details), which timed response latencies in milliseconds.

Design and Procedure

Participants in the experiment named the complete set of 150 words in a series of four blocks (separated by brief rest periods). Regular and exception words were spread evenly through the blocks, and each block was matched for mean CELEX frequency, positional bigram frequency, Coltheart's *N*, and imageability. Each of the experimental blocks began with either 2 or 3 starter items (low-frequency regular and exception items), so that there were 40 items to be named per block.

Participants, tested 1 at a time in a quiet room, were instructed to name the words aloud as quickly and accurately as they could. The instructions were followed by a block of 15 practice trials and then by the four experimental blocks. The intertrial interval in both practice and experimental blocks was 500 ms. Within each trial, participants first saw a fixation point, which remained on the screen for 750 ms. Immediately at the offset of the fixation point, the word to be named appeared. As soon as the participant named the word, it disappeared and the cycle was repeated after the intertrial interval. The words within each block were presented in a different random order for each participant, and the order of presentation of the four experimental blocks was randomly counterbalanced between participants. The experimenter recorded mispronunciations and voice-key errors by hand during the experiment, and these were checked with tape recordings of the sessions afterward. At the end of each session, participants were shown the complete set of items and asked to circle any items with which they were not familiar. They were also re-presented with any words that they had mispronounced during the experiment and asked to read these aloud again, taking their time. This was to check whether the participants knew the correct pronunciations of all the items in the experiment.

Results

In the following analyses, latencies and errors were pooled over participants for each word and a p value of less than .05 was accepted as significant. A total of 13.4% of the data (386 responses) were excluded from the latency analysis. Of these, 8.9% (256 responses) were removed because of mistakes in naming the target word. Another 1.5% (44 responses) were removed because participants made an articulation error on the target word (e.g., stuttered). A further 1.4% (41 values) were removed because of voice-key errors, including any responses slower than 1,500 ms or faster than 300 ms. Finally, 1.6% (45 responses) of the data were removed because participants, when probed at the end of the experiment, reported that they were unfamiliar with the target word.

The regression model included 17 main variables as described in the *Method* section (i.e., 10 for initial phoneme plus CELEX frequency,[1] length, imageability, AoA, positional bigram frequency, Coltheart's N, and regularity) plus a term for the interaction between regularity and imageability. The distributions of each of the continuous independent variables were plotted, and transformations were applied as required to reduce skew and kurtosis and to improve normality. CELEX written frequency was log transformed, and imageability was square root transformed. All continuous variables were centered (Aitken & West, 1991) before being entered into the regression model. Centering of variables aids in the interpretation of interactions by reducing collinearity effects.

Regression Analyses

Naming latency. The following regression analyses used unique sums of squares. All of the variables entered the regression at once; each independent variable is evaluated in terms of what it adds to the prediction that is not also predicted by the other variables. The outcome of the regression analysis carried out on naming latencies is shown in Table 2. The regression model accounted for 76% of the latency variance; 72% adjusted, $F(18, 101) = 17.88$, $MSE = 698.48$. The largest proportion of unique variance was attributable to initial phoneme: voiced initial phonemes triggered the voice key faster than unvoiced; fricatives activated the voice key later than other phonemes; and bilabials, labiodentals, and velars gave rise to faster voice-key responses than did phonemes with alternative places of articulation. Word frequency also predicted naming latency, with participants responding more quickly to higher frequency words than to lower frequency words. AoA did not account for a significant portion of unique variance in the latency data.

Regularity and imageability each accounted for a significant portion of unique variance, with regular words producing shorter latencies than exception words and with high-imageability words yielding shorter latencies than low-imageability words. Most important, imageability and regularity interacted. As recommended by Aitken and West (1991), the interaction was plotted using the regression equation, and it is represented in Figure 1 as the mean latency of the regular and exception words at high, medium, and

Table 2
Outcome of Regression Analysis Carried Out on the Latency Data

Variable	β	t	p	sr^2
Initial phoneme				16[a]
Voice	−40.81	−5.86	<.01	
Nasal	2.828	0.168	.87	
Fricative	29.22	3.1	<.01	
Liquid	−5.5	−0.357	.722	
Bilabial	−36	−2.38	<.02	
Labiodental	−51.7	−2.81	<.01	
Palatal	−26.33	−1.42	.157	
Alveolar	−21.51	−1.49	.139	
Velar	−30.68	−2.294	<.05	
Glottal	−19.62	−0.94	.35	
Age of acquisition	0.06	1.68	.096	
Word frequency	−27.88	−4.48	<.01	4.80
Word length	7.057	1.39	.167	
Positional bigram frequency	0.002	1.6	.112	
Coltheart's N	−.22	−.33	.74	
Imageability	−2.927	−2.615	<.02	1.62
Regularity	−26.07	−5.338	<.01	6.74
Imageability × Regularity	3.21	2.38	<.02	1.60

Note. Total percentage of variance accounted for = 76% (unique variability = 31%; shared variability = 45%). sr^2 = % of unique variance.
[a] Total for initial phoneme.

low imageability when the other predictor variables are at their mean values. Analysis of the simple slopes reveals that imageability significantly predicts latency for exception words, $t(131) = -2.85$, but not for regular words, $t(133) = -0.5$, as we hypothesized.

Errors. The same independent variables used as predictors in the latency regression analysis were used with the error data. Errors were pooled over participants for each word. The regression model accounted for 39% of the error variance; 30% adjusted, $F(18, 131) = 4.64$, with $MSE = 5.59$, for all the F ratios in this section. Of the initial phoneme variables, only velars accounted for a significant portion of the error variance, $F(1, 131) = 5.8$. Three of the other factors accounted for significant portions of the error variance, namely word length (greater accuracy for shorter words), $F(1, 131) = 8.586$, regularity (an accuracy advantage for regular words), $F(1, 131) = 39.122$, and AoA (higher accuracy for early acquired words), $F(1, 131) = 7.39$. Neither imageability on its own nor its interaction with regularity significantly predicted error rate.

Discussion

The latency analysis of this large set of items supports our claim that imageability has a greater impact on the naming of low-frequency exception words than low-frequency regular words,

[1] We chose to include CELEX frequency in the regression analysis instead of WFG because preliminary analyses indicated that it predicted significantly more of the latency data, and because we felt it better not to include two separate frequency measures. It should be noted, however, that when entered in the regression model, either in conjunction with CELEX or instead of it, WFG frequency did not reduce the Regularity × Imageability interaction.

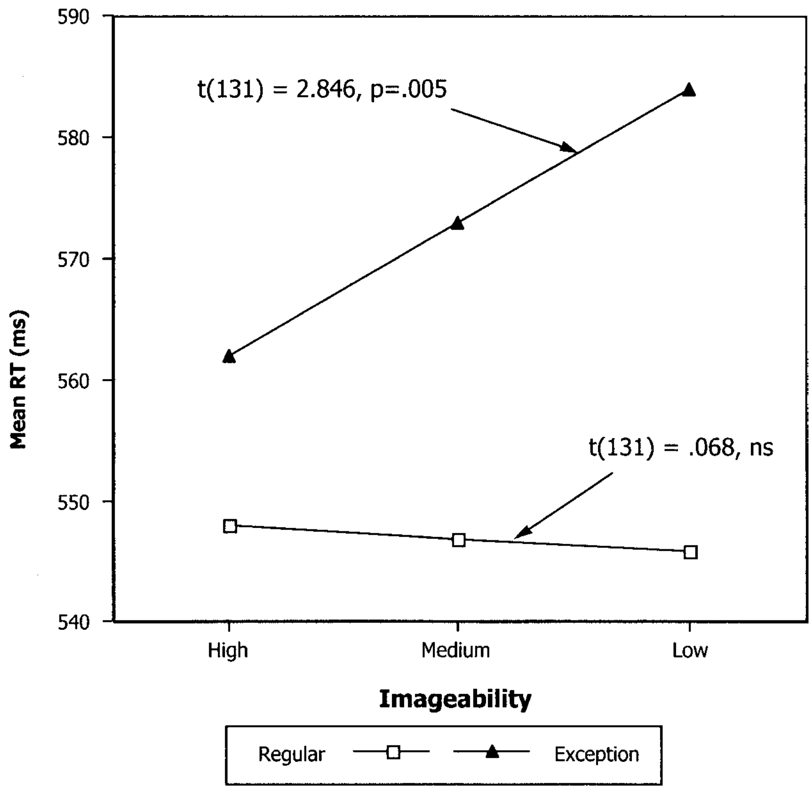

Figure 1. Mean latencies of the regular and exception words at high-, medium-, and low-imageability values. RT = response time.

replicating the findings of Strain et al. (1995) and Strain and Herdman (1999). In this experiment, the Regularity × Imageability interaction remained significant in a regression analysis controlling for AoA.

Concluding Comment

What have we learned about O→P processes from the past nearly quarter of a century of research on word naming? It seems clear that the efficiency (speed and accuracy) of translating a written word to its pronunciation is modulated by the extent to which the word's pronunciation is typical of (i.e., can be predicted from) its spelling pattern. Detailed criteria for how to characterize typicality are still debated, but a genuine consistency or regularity effect is indisputable. It is also clear that spelling–sound typicality interacts with some measure of familiarity; words that we encounter very frequently, even if they have atypical spelling–sound correspondences, are processed efficiently. Written-word frequency seems almost guaranteed to be the source of a genuine interaction with typicality, and current models of word reading attempt to address this effect in different ways.

The question addressed here is whether we need one or more factors in addition to word frequency in our quest to provide a full account of the impact of spelling–sound typicality. Here we mean central lexical factors, not peripheral ones like initial phoneme, though we would like to again emphasize the importance of controlling such powerful peripheral factors. Both Strain et al. (1995) and Monaghan and Ellis (2002) argued for an additional factor—but different ones. Strain et al. set out to provide evidence for an impact of word imageability; Monaghan and Ellis not only did the same for another variable, AoA, but also attempted to deny any genuine role for imageability. On the basis of our arguments in the first section, concerning both Monaghan and Ellis's mischaracterizations of the Strain et al. study and also some problems in Monaghan and Ellis's study, and especially on the basis of the evidence from our new experiment with a larger stimulus set, we argue that there is more compelling evidence for an interaction between typicality and imageability than between typicality and AoA. The message with which we conclude, however, is really the point that we have attempted to make with the title of this article. Almost all of the factors in the domain of word familiarity and meaningfulness are correlated with one another, and it is a brave experimenter who attempts to establish the prominence of one while denying any impact for the others. Given that the field hosts a variety of theoretical positions about what should interact with spelling–sound typicality, the most informative answer to the question of what currently produces this interaction is the following: theories of word naming.

References

Aitken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions.* Newbury Park, CA: Sage.

Baayen, R. H., Piepenbrock, R., & Van Rijn, H. (1993). *The CELEX lexical* *database* [CD-ROM]. Philadelphia: University of Pennsylvania, Linguistic Data Consortium.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments and Computers, 25,* 257–271.

Coltheart, M. (1981). The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology, 33A,* 497–505.

Jared, D., McRae, K., & Seidenberg, M. S. (1990). The basis of consistency effects in word naming. *Journal of Memory and Language, 29,* 687–715.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

Monaghan, J., & Ellis, A. W. (2002). What exactly interacts with spelling–sound consistency in word naming? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 183–206.

Patterson, K., Plaut, D. C., McClelland, J. L., Seidenberg, M. S., Behrmann, M., & Hodges, J. R. (1996). Connections and disconnections: A connectionist account of surface dyslexia. In J. A. Reggia, E. Ruppin, & R. S. Berndt (Eds.), *Neural modeling of brain and cognitive disorders* (pp. 177–199). Singapore: World Scientific.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired reading: Computational principles in quasi-regular domains. *Psychological Review, 103,* 56–115.

Raaijmakers, J. G., Schrijnemakers, J. M., & Gremmen, F. (1999). How to deal with "The language-as-fixed-effect fallacy": Common misconceptions and alternative solutions. *Journal of Memory and Language, 41,* 416–426.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96,* 523–568.

Seidenberg, M. S., & Waters, G. S. (1989). Word recognition and naming: A mega study. *Bulletin of the Psychonomic Society, 27,* 489.

Seidenberg, M. S., Waters, G. S., Barnes, M. A., & Tanenhaus, M. K. (1984). When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning and Verbal Behavior, 23,* 383–404.

Strain, E., & Herdman, C. M. (1999). Imageability effects in word naming: An individual differences analysis. *Canadian Journal of Experimental Psychology, 53,* 347–359.

Strain, E., Patterson, K., & Seidenberg, M. S. (1995). Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 1140–1154.

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.

Venezky, R. L. (1970). *The structure of English orthography.* The Hague, the Netherlands: Mouton.

Zeno, S. (1995). *The educator's word frequency guide.* Brewster, NJ: Touchstone Applied Science Associates.

Zevin, J. D., & Seidenberg, M. S. (in press). Age of acquisition effects in reading and other tasks. *Journal of Memory and Language.*

Appendix A

Reclassification of Monaghan and Ellis's (2002) Experiment 3 "Inconsistent" Stimuli

Clearly inconsistent	Ambiguously inconsistent
brooch	bead
caste	font
deaf	foul
draught	grieve
drought	knead
ghoul	plead
hearth	sheath
mould	spook
quay	squat
sew	swamp
sewn	swap
shove	swarm
soot	swat
steak	swatch
swear	wad
vase	wand
	warn
	warp
	wart
	wreath

(*Appendixes continue*)

Appendix B

Word List for the Experiment

Exception	Regular	Exception	Regular
ache	arch	pear	peck
aunt	arc	pint	plug
axe	ail	pour	plum
blown	blotch	quart	quest
bowl	blunt	realm	rude
brooch	broom	scarce	snail
cache	carve	seize	saint
caste	crumb	sew	sauce
choir	clink	sleight	scribe
chord	cloak	shoe	shed
comb	clung	shone	shawl
cough	croft	chute	shrewd
couth	crude	sieve	skate
crow	cage	sloth	slang
deaf	duck	sown	sage
debt	dumb	sponge	starch
draught	dense	steak	spire
dreamt	drench	suave	spout
fete	filth	suede	sock
flood	flame	suite	steam
flown	fraud	swear	stub
gauge	grace	sword	stump
ghoul	gait	tomb	tack
glow	gift	ton	tale
gross	graft	tow	tune
hearth	heave	trough	trend
hearse	hoarse	wolf	waist
leapt	lamb	womb	web
mauve	merge	wool	weed
mould	moan	worm	weep

Received May 21, 2001
Revision received May 29, 2001
Accepted June 25, 2001 ■