Running head: FEEDBACK CONSISTENCY

Feedback Consistency Effects in Single-Word Reading

Brett Kessler and Rebecca Treiman

Washington University in St. Louis

John Mullennix

University of Pittsburgh at Johnstown

Feedback Consistency Effects in Single-Word Reading

The goal of this chapter is to examine one of the most intriguing claims that has been made during the past decade of research in single-word reading—the idea that feedback consistency influences fluent adults' performance. It has been reported in several studies (Lacruz & Folk, 2004; Perry, 2003; Stone, Vanhoy, & Van Orden, 1997; Ziegler, Montant, & Jacobs, 1997) that adults are slowed when reading a word like *hurl* because other words that *hurl* rhymes with, such as *girl* and *pearl*, have different spellings of the same rhyme. This effect is surprising, because inconsistency in the sound-to-letter direction, something that might logically make writing difficult, would seem to play no necessary role in reading, which involves mapping letters to sounds. We begin this chapter by reviewing the literature on the feedback consistency effect, pointing out some methodological problems in several of the studies that have argued for its existence. We then report new analyses designed to determine whether feedback consistency has a reliable effect on the speed with which people read individual words.

Prior Research

The speed with which readers can process individual written words has long been a focus of research in literacy and lexical processing (Balota, 1994). Such data can be crucial for the development and verification of models of reading. For example, it is well known that words that a reader has seen with great frequency are processed faster than other words. For this, among other reasons, the DRC (dual-route cascaded) model of Coltheart, Rastle, Perry, Langdon, and Ziegler (2001) includes a lexical route that can retrieve high-frequency words from the mental lexicon quicker than one can access a word via letter-to-sound rules. Likewise, connectionist models such as those of Plaut, McClelland, Seidenberg, and Patterson (1996) are trained most thoroughly on high-frequency words, thereby building up particularly strong connections for them. Such findings support the popular idea that fluent readers process some words as wholes: so-called sight words.

It is also commonly agreed that reading involves phonological processing as well, at least for words of lesser frequency. In phonographic orthographies such as that of English, individual written words are built up from units that represent the sounds of the spoken words. One productive line of research has looked at how various aspects of these phonographic representations affect processing. Several studies have reported that words containing ambiguous letter-to-sound mappings are processed more slowly than other words (e.g., Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Glushko, 1979; Lacruz & Folk, 2004; Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welty, 1995). All other things being equal, a word like *stead* will be processed more slowly than a word like *shell*, because analogous words in English suggest two different pronunciations for the former: the correct /stɛd/[1] (cf. *head*) and the incorrect /stid/ (cf. *bead*). In contrast, all words that end in *ell* are pronounced to rhyme with *shell*. Again, all models of reading need to take such facts into account. In DRC, for example, words with the most typical letter-to-sound correspondences, as in *shell*, can be processed by the moderately efficient phonological assembly route, while words that have letters with unusual sound correspondences can be retrieved only by the lexical route, which can be fairly slow for low-frequency words like *stead*. On a more practical level, this *consistency effect* indicates that phonological decoding is involved even in the skilled reading of college students, the subjects of most adult reading research, and may therefore encourage reading teachers to acknowledge the usefulness of phonics-based reading instruction.

Although the consistency effect has been supported by much research, it is not unchallenged, nor have all the details of its effects been settled, such as whether it applies to high-frequency words (Jared, 1997, 2002), or exactly how consistency is to be measured (Massaro & Jesse, 2005). For example, some researchers treat consistency as a binary measure: If the letters that spell the rhyme of the word have different pronunciations in different words, then the word is inconsistent and will be processed more slowly than other words, ceteris paribus (Glushko, 1979). Other researchers treat consistency as a continuous measure that may be computed over different parts of the word, not just the rhyme: Consistency is the proportion of all words with the same letters that also have the same pronunciation, and different degrees of inconsistency may slow readers by different amounts (Treiman et al., 1995). But however consistency is defined, consistency theory has intuitive appeal. If it is accepted that readers do any phonological decoding at all, then it makes sense that words like *stead*, which could theoretically be read two different ways, would cause them to slow down or stumble.

In 1997, Stone et al. announced a much less intuitive consistency effect: that of feedback consistency. In their main experiment, a lexical decision task, they presented subjects with a series of 100 strings of letters and asked them to quickly judge whether each string was a word or a nonword. All the real words were monosyllabic and of fairly low frequency and had rhyme letters, that is, the letters from the vowel to the end of the word, that were completely consistent in the spelling-to-sound direction. However, half of the words, such as *heap*, were inconsistent in the reverse, or feedback, direction: the sounds /ip/ can also be spelled as in *deep*. The other half of the words, like *probe*, contained rhyme pronunciations that can only be spelled one way; these were called feedback consistent words. Stone et al. (1997) found that it took, on average, 774 ms to correctly identify the feedback consistent stimuli as words, while the feedback inconsistent words took 33 ms longer to identify as words. The error rate was also higher on the feedback inconsistent words compared to the feedback consistent words: 9.8% as opposed to 3.9%.

The idea that sound-to-spelling consistency should affect reading is surprising because there does not seem to be any practical reason for people to consult the spelling of rhyming words as part of the task of deciding whether *heap* is a word. However, the results of the experiment made sense in the context of Stone and Van Orden's (1994) theory of *recurrent networks* in word perception. In a recurrent network, the flow of activation is inherently bidirectional. When, in the course of reading, the activation of letter units in one's perceptual input results in the activation of sound units in one's mental lexical networks, those sound units, in turn, automatically activate letter units, just as if those sound units had been activated by sensory input. When reading a word like *probe*, the letters activate the sounds /prob/, which in turn activate the letters *probe*, an internally consistent state of affairs that allows the network to quickly settle on a decision. In contrast, reading *heap* activates the sounds /hip/, which in turn activate, in addition to *heap*, the conflicting spelling *heep*. This conflicting information creates an unstable feedback loop, which requires more time to settle into a state that represents a confident decision.

Thus the possibly counterintuitive results of Stone et al.'s (1997) experiment could be taken as evidence in favor of recurrent networks, a rather exciting theoretical development. In the context of Stone et al.'s recurrent network theory of reading, the traditional consistency of the letter-to-sound mappings was now considered *feedforward* consistency, and the new role played by consistency of the mapping in the reverse direction, from sounds to letters, was called *feedback* consistency. Other researchers have adopted this terminology regardless of their theoretical perspective, and several have joined the search for feedback consistency effects in the lexical decision task, with varying degrees of success. Perry (2003) and Lacruz and Folk (2004) reported a feedback consistency effect in English lexical decision, but Massaro and Jesse (2005) reported there was none. Balota et al. (2004) reported mixed results: Effects were not significant by items but only by subjects, and only for slower participants. In French, a language that shares with English the fact that sounds may be spelled inconsistently across words, Ziegler et al. (1997) reported a feedback consistency effect in lexical decision, but those results were contradicted by Peereman, Content, and Bonin (1998); and see further Ziegler and Van Orden (2000) for a counterrebuttal.

Several factors may account for the conflicting results in lexical decision tasks, some of which will be addressed below. One important issue is that of covariables. So many different factors affect lexical processing that it is very difficult to tease them apart or to design a factorial experiment that perfectly balances lexical stimuli on all factors besides feedback consistency (Cutler, 1981; Massaro & Jesse, 2005). For example, Peereman et al. (1998) noted that Stone et al.'s (1997) attempts to balance words by their frequency appeared to be faulty because the latter used frequency estimates based on a corpus of inadequate size. More subtly, it is possible that only certain types of words may evoke feedback effects. It is often believed, for example, that phonological factors like feedback consistency would not affect the processing of high-frequency words (though see Lacruz and Folk, 2004, for a demurral).

Another important issue is how feedback consistency is defined and measured. As mentioned above, some researchers have treated consistency as a binary property. As applied to feedback consistency, this means that a sound either has a consistent spelling or it does not. One may immediately object that, by that definition, every word of English is feedback inconsistent, because every word has at least one sound that can be spelled more than one way (all vowel sounds, for example, are inconsistent). This objection has generally been rendered irrelevant by the fact that most research has only looked at one-syllable words and has assumed that the rhymes of words are processed as units (e.g., Balota et al., 2004; Lacruz & Folk, 2004; Peereman et al., 1998; Stone et al., 1997; Ziegler et al., 1997); many rhymes have consistent spellings in both English and French. This assumption of a prominent role for the rhyme is venerable in reading research (e.g., Glushko, 1979) and has much empirical and theoretical support (summarized by Treiman & Kessler, 1995), but it is far from obvious that feedback consistency effects should manifest themselves at the level of the rhyme and nowhere else. Perry (2003) looked at the feedback consistency of vowels, which was made possible by the fact that he adopted a graded definition of consistency: Vowels can be more or less consistent. It is conceivable that readers are sensitive to only certain types of feedback consistency and that incompatible findings may be due to the fact that experimenters are measuring somewhat different things.

A third issue that calls into question the lexical decision results is that of ecological validity. The lexical decision task has a long and respectable history, but it is difficult to know what it reveals. It is far from clear that lexical decision per se ever occurs in the course of natural reading. As Henderson (1989, p. 358) noted, the task “obliged the reader to journey exactly as far as the portals of the lexicon, to ring the bell and, if someone answered, to run home without further ado to report this happy domestic circumstance.” Certainly lexical access occurs, and sometimes it fails. But we do not fully understand what happens when we ask someone to explicitly judge whether lexical access is possible for a given stimulus. Different people could approach this artificial task in different ways, and the approaches could be influenced by subtle differences in participant samples or in experiment administration. It is not inconceivable that subjects who weigh accuracy over speed may choose to assess the lexicality of *heap* by decoding its pronunciation to /hip/, checking whether /hip/ means something, then asking themselves how to spell /hip/, to see if the results match the spelling they were originally presented with. Such a suspicion is supported by Balota et al.’s (2004) finding that feedback consistency effects in the lexical decision task were reliable only for participants whose average word response latencies were slower than the median. If inconsistency in the sound-to-letter direction slows the task of deciding whether /hip/ is spelled *heap*, that is interesting enough in its own right. But if the participant has more or less explicitly generated a spelling subtask, sound-to-letter consistency has become a feedforward factor, and the experiment arguably is not addressing the key theoretical issue of whether feedback is an automatic component of the normal reading process.

Naming studies are much less subject to charges of the lack of ecological validity. In these studies, participants are shown a word and asked to read it aloud as quickly and accurately as possible, and the experimenter records how long it takes the participant to begin saying the word. While the exact conditions of a naming experiment are, as in most experiments, rather artificial, it is also clear that the task much more closely approximates the natural process of reading, which is often done orally, quickly, and with concern for accuracy. In particular, the naming task omits the mysterious and potentially misleading judgment component of the lexical decision task. For this reason, the naming task has been the second major avenue for researching feedback consistency effects.

In their experiment on French word naming, Ziegler et al. (1997) reported a feedback consistency effect. It took participants, on average, 8 additional milliseconds to pronounce feedback inconsistent words as compared to feedback consistent ones, and they made 2.7% more errors. But these numbers were much smaller than those obtained from their own lexical decision experiment, which involved a 33-ms difference and an 8.7% increase in error rate. Most importantly, the significance levels from the naming experiment were not reassuring. The difference in response time was significant when computed by subjects but not by items, and the difference in error rate was not significant either way, using a cutoff of .05. When Massaro and Jesse (2005) obtained similar results in English, they, more conservatively, reported a lack of evidence for feedback consistency. Peereman et al. (1998) also reported no significant feedback consistency effect on either response time or error rates in French. On the other hand, Lacruz and Folk (2004) reported success in English, and Balota et al. (2004) reported that feedback consistency of both the onset and the rhyme affected naming response times in all their analyses, with very strong significance. Intriguingly, their results were more robust for naming than for lexical decision, the opposite of the pattern reported by Ziegler et al.

The lack of agreement in results across the naming studies may be due to many of the same problems that afflict lexical decision studies of feedback consistency. Lexical covariables are just as likely to mask or imitate feedback consistency in the naming task, and there are many different ways of conceptualizing and measuring consistency. In addition, naming studies have a serious handicap that lexical decision studies do not have. The differences in response time between feedback consistent and inconsistent words are small, with values of around 10 ms or less commonly reported. At such small time scales, the physics of speech production becomes a major player. For two words that have the same lexical access time—which is essentially the datum that one wishes to measure—it can take different amounts of time to plan and realize their articulation. Moreover, even after the articulation begins, it may take different amounts of time for the vocal tract to emit sufficient acoustic energy for a measuring device to register sound above background levels. In experiments using standard devices such as voice keys, these differences often are an order of magnitude larger than the purported difference between consistent and inconsistent words (Kessler, Treiman, & Mullennix, 2002).

Such a problem would be considered nothing more than an annoying source of noise, were it not for its quasi-systematicity. On the one hand, the variability is too irregular to control with simple mathematical adjustments. An experiment can get completely different patterns of errors due to differences in the sensitivity of the microphone, the distance between the speaker's mouth and the microphone, the threshold setting on the voice key, and how loud the speaker speaks. On the other hand, there is just enough systematicity to be troubling. On average, a loud response will appear to have been uttered sooner than a soft response, because the word will more quickly reach a threshold of acoustic amplitude. Or a response uttered at a faster tempo will generally appear, falsely, to have been uttered before a response uttered at a slower tempo, even if the response actually began at the same time, because at a faster tempo the speaker will generally get through any subthreshold initial portion of the word more quickly. Naturally, the experimenter does not want to confound soft amplitude or slow tempo with slow response time, since these phenomena may have quite different explanations in terms of the functioning of a cognitive network.

Most perniciously, response time measurements are correlated with the phonetic segments that occur at or near the beginning of a word, for both articulatory and acoustic reasons. For example, words beginning /tɑ/, /sɑ/, /si/, and /stɑ/ may all have systematically different response times associated with them. Unfortunately, the details are difficult to predict in a natural experimental setting, even to the extent of their relative order: /stɑ/ may be associated with unusually fast response times or unusually slow response times, depending on such factors as the measurement technology (Sakuma, Fushimi, & Tatsumi, 1997) and how loud the participant is speaking. The particular danger lies in the fact that the identity of the same phonemes that contribute to systematic measurement bias is an essential part of the computation of the feedback consistency measurement. That is, there is a quasi-systematic relationship between feedback consistency and response time measurements that has nothing to do with the cognitive process under investigation. This bias, naturally, is most troublesome in a study that investigates the feedback consistency of the onset (prevocalic) consonants or of the word’s head (onset plus vowel), because phonemes at the beginning of the word have the strongest effect. But problems can arise even if the study looks only at the consistency of the rhyme of one-syllable words (the vowel plus coda, i.e., the following consonants), because there are many words that begin with a vowel, and vowels can influence response time measurement bias even when preceded by onset consonants (Kessler et al., 2002).

Most studies of feedback consistency that measure naming response time have taken some notice of these measurement biases and have attempted to work around them in some way. Ziegler et al. (1997) added a delayed-naming task. After the participants had gone through the normal trials, pronouncing the words as quickly as possible, they went through a second round where they were given time to mentally prepare before being asked to pronounce each word. Because feedback consistency did not significantly affect response time measurements in the delayed condition, the authors concluded that the feedback effects measured in the normal trials must reflect the time taken for mental preparation, that is, settling of the mental lexical networks. But Massaro and Jesse (2005) got the opposite results from using the delayed-naming task. They found that feedback consistency affected response time just as strongly in delayed trials as in normal, immediate trials, and concluded that measured feedback effects must therefore be due at least in part to processes that follow the act of reading. Of course, while such demonstrations are suggestive, they have some shortcomings. For example, in repeated measures designs, participants may behave systematically differently on delayed trials, whether because of fatigue or familiarity with the materials. Also, delayed trials may subtract out some articulatory contributions to response time in addition to the cognitive processing time they are intended to target, because speakers may, for example, position their articulators to pronounce the first phoneme while anticipating the cue to speak.

Massaro and Jesse (2005) used several other controls in addition to the delayed-naming task. Instead of relying on voice keys, they studied digitally produced waveforms to determine the onset of speech. They also adopted the technique of Kawamoto, Kello, Jones, and Bame (1998), whereby the participant is asked to drone a neutral vowel, /ə/, up to the point where the response is uttered. When the word in question begins with an obstruent, measuring the end of the drone vowel can be a more accurate estimate of the beginning of the word than an attempt to detect the sound of the consonant itself. These methods greatly reduce bias due to acoustic factors. However, bias due to articulatory factors necessarily remains, even under perfect measurement conditions. Some initial sounds will measure systematically differently because they take longer to articulate. These differences are real but regrettable, because it can be difficult to tell whether a slow response is due to articulatory factors or to feedback inconsistency. Massaro and Jesse attempted to compensate for remaining phonetic biases by balancing their experimental stimuli by manner of articulation of the initial consonant and by using words that do not vary by number of initial consonants.

Massaro and Jesse (2005) addressed phonetic bias problems with unusual and exemplary thoroughness. However, not all of their techniques can easily be adopted in all experiments. In particular, megastudies, which have been gaining in popularity, are not easily analyzed in these ways. Lexical megastudies are experiments that use thousands of words as stimuli, often thoroughly covering appreciable subsets of a language's vocabulary. Seidenberg and Waters (1989), Kessler et al. (2002), and Balota et al. (2004) describe experiments where participants read virtually all common, simple, one-syllable words of English and their response times for each word were recorded. Such studies are typically analyzed using regression analyses rather than factorial ANOVAs, freeing the experimenter from having to equate lexical stimuli across conditions on the many covariables that can affect performance. Megastudies have several other advantages as well (see Balota et al., 2004, for a thorough discussion), not least of which is a great potential for reusing data for other purposes than they were originally collected for. But they require the researcher to revisit the question of how to address phonetic bias, because the manual analysis of waveforms may be prohibitively expensive or may not be available at all when retrospectively analyzing megastudy datasets for which only voice key measurements are available.

Several researchers (Balota et al., 2004; Spieler & Balota, 1997; Treiman et al., 1995) addressed the problem of phonetic bias in megastudy regression analyses by introducing several variables to stand for the articulatory features of the first phoneme of the word, indicating, for example, whether it is a bilabial sound, or a fricative, or voiced. When this technique was introduced in 1995, 10 binary variables probably seemed more than enough to account for known voice key biases. Nowadays, however, it is better understood that the biases extend beyond the first phoneme of the word. More importantly, the contributions of different features to voice key bias are simply not additive (Kessler et al., 2002). This point is crucial because of the logic of investigating feedback consistency via linear regression. If one claims that an analysis shows that feedback consistency makes a significant contribution to predicting response time that cannot be attributed to other variables, that is tantamount to saying that all effects caused by the covariables, including the phonetic features, have been adequately accounted for by the additive, linear, model of the regression. If that claim is not credible, then the claim of feedback consistency effects is not credible.

Because there is no well-established way to decompose into linear components the phonetic bias contributions of all the different heads that can begin simple monosyllabic words in English, one might think instead to treat the different heads as levels of a category variable. Within the context of an ordinary linear regression, the categorical variable could be represented by dummy variables. Unfortunately, there are over 600 different heads, and therefore over 600 dummy variables. Such a huge number of variables can prove intractable, quite apart from whether one believes one could trust the results generated with such a huge model.

A Retrospective Analysis of Feedback Consistency Effects in Naming

In this section we introduce a new way of addressing whether there are true feedback consistency effects in word naming. In order to deal with the issues identified in the previous section, we have experimented with a hybrid solution to the problem of running a regression analysis while controlling for phonetic heads. In the first step, an ordinary regression is run, using response time in a reading task as the dependent variable and, as independent variables, a variety of quantitative measures that may have an effect on response time. This step omits the feedback consistency measures that are the real target of the research, as well as the categorical variables, the heads of the words. The goal of this first step is to fit a model that accounts for as much of the variation in response time as we possibly can, before we get to the target variable and the less tractable heads.

As a second step, we apply the fitted linear model to all the words in the same study, to find the difference between the observed response time and the time predicted by the model. If the first step was reasonably accurate, these residuals will comprise a mixture of random variation and the effects of variables we have not yet considered.

In the last step, we see whether these residuals can be accounted for by the feedback consistency measures. We see whether the two numbers correlate in a monotonically ascending or descending order, a more liberal test than seeing whether they stand in a straight-line relation, as a standard regression would do. At the same time, we test for significance using permutation tests that only rearrange data between words that have the same phonetic heads. That way, the *p* value will disregard any variability that is due to the phonetic heads. This procedure therefore completely eliminates all known sources of phonetic bias, whether articulatory or acoustic in origin.

Materials

Data were analyzed from four previous megastudies that used voice keys to measure response time of U. S. and Canadian college students in naming tasks: Kessler et al. (2002) , henceforth KTM; Seidenberg and Waters (1989), SW; Spieler and Balota (1997; Balota et al., 2004), SB; and the English Lexicon Project (Balota et al., 2005), ELP. For the purpose of comparability, analyses here are limited to data available for all four studies: response times averaged across all participants for each of 2,326 simple, one-syllable words. Error rates were also analyzed, although the data are not completely commensurable across studies. In KTM and SW, errors were coded by the experimenter, whereas in SB and ELP, errors were noted only if reported by the participants themselves immediately after naming the word in question.

Table 1 shows the dependent variables that were collected for each word for use as covariables in the regression step. Most of these variables were chosen because previous studies had indicated that they may have an effect on response time, and most of them have been taken into account in careful studies of feedback consistency. A few others were included because we suspected that they may affect response time and correlate with feedback consistency, so failing to include them could lead to spurious effects. For example, spelling–sound correspondences that have low feedback consistency, such as the *gn* /n/ of *gnat*, will usually have spellings that are less frequent in text: More words have onset *n* than *gn*. It is reasonable to surmise that readers may hesitate when they encounter less common letter groups like *gn*, but such behavior would not constitute evidence for recurrent networks. In the absence of the written frequency variables, some of the effect due to spelling frequency could be spuriously attributed to feedback consistency.

For the last step, the correlation, we collected also feedback consistency measures using measures analogous to the feedforward measures. The phonetic head was also determined for each word.

Procedure

For each of the four studies, ordinary least squares linear regressions were run using response time as the dependent variable and, as independent variables, the measures listed in Table 1. Residuals were extracted for each word, giving one residual per study per word. Correlations were run between each of the four sets of residual measurements and each of the five ways of computing feedback consistency (onset, head, vowel, rhyme, and coda). The correlation measure consisted of multiplying the residual and the consistency measure for each word, then summing those products across all words. This measure was chosen because the sum of products is minimized as the pairs of numbers approach inverse rank order. That is, if the words were to be ordered by increasing residual value, such a measure would be at a minimum if the corresponding consistency measures were in decreasing order. Therefore, significance of the correlation measure was determined by randomly rearranging the associations between residual and consistency 10,000 times and counting how many times the rearranged measure was less than or equal to the observed measure. This tested the hypothesis that increased consistency is associated with decreased (faster) response time. To factor out any effect of phonetic heads from the significance measure, the random rearrangements only took place between words that have the same phonetic head.

To test the effect of feedback consistency on error rates in each of the four studies, all predictor variables, including the feedback consistency measures, were entered into a simultaneous regression. Because there is no reason to expect that phonetic factors will correlate with error rates, no steps were taken to block by phonetic head.

Results and Discussion

The columns labeled *Response time* in Table 1 show which variables accounted for significant variance in response time in the regression analyses of Step 1. Significance levels vary among the four different studies, but studies generally agree in the direction of any effect. Note that a plus sign in the table indicates greater response time, that is, slower reading. Some of the results are unsurprising. For example, the more frequent and familiar a word is, or the more common the spelling of its rhyme, the faster one can read the word. Other results are perhaps less expected. For example, the more common vowel phonemes are associated with slower response times. Of course, this part of the analysis ignores feedback consistency and the phonetic heads of the words, and so these specific results should not be taken as definitive. More important is the result of the correlation, which happens after the regression step and so is not shown in Table 1. Of the 20 analyses of response time (4 studies with 5 ways of computing feedback consistency), only one showed an effect in the predicted direction: The SB data showed increased feedback consistency in the onset as being associated with faster response times ($p = .04$). The finding of one significant test out of 20 is what one would expect by chance at a significance threshold of .05.

Our regression analyses of error rates are presented in the right half of Table 1. Recall that for error rates, we considered it safe to treat feedback consistency directly in the simultaneous regressions, because there was no reason to expect phonetic voice key biases to interfere with the measurements. The results for error rates do not tell any unified story—sporadic effects in both directions can be seen in the table—but it is not completely clear what effects would be predicted for feedback consistency in the first place. On the one hand, anything that perturbs the equilibrium of networks enough to increase response time might be expected to also lead to increased errors. By this reasoning, we would expect feedback consistency to lead to reduced error rates (minus signs in the table), but the only reliably replicated effect appears to be that vowel feedback consistency is associated with increased error rates. On the other hand, the argumentation behind feedback consistency experiments depends crucially on the assumption that the reader has decoded the correct pronunciation in the initial, feedforward, steps. Why, then, should an error emerge? The only type of error that would specifically be predicted on the basis of recurrent networks would be the result of feedback from the feedback. Upon seeing the word *rose* and decoding it to the correct /roz/, the network might generate, among other respellings, *rows*. In turn, feedback from the feedback spelling *rows* could generate the error pronunciation /rɑʊz/, rhyming with *cows*. It would be intriguing if people do indeed read *rose* as /rɑʊz/. Unfortunately, none of the actual mispronunciations generated in the megastudies have been made available, so it is impossible to evaluate this implication of recurrence theory.

Conclusion

The results of our experiment in factoring out phonetic bias were not encouraging for the hypothesis that feedback inconsistency slows readers' performance in word-naming tasks. This conclusion does not appear out of step with past research. Some of the strongest results supporting a feedback effect on naming have either not controlled for phonetic biases at all (e.g., Lacruz & Folk, 2004) or have done so in a way that is arguably inadequate (e.g., the binary feature variables in Balota et al., 2004). Other results are more tepid (e.g., Ziegler et al., 1997) or agree with us in being negative (e.g., Massaro & Jesse, 2005; Peereman et al., 1998). Further, our study took into account additional predictor variables that may have been confounded with feedback properties in several prior studies, such as the frequency of the letter group that spells the unit whose feedback consistency is being measured. All in all, there are strong grounds for being skeptical of the idea that there is a proven feedback consistency effect in naming.

Does that mean that no effect will ever be demonstrated? Current research protocols make such an undertaking almost hopelessly difficult. Any effect of feedback on the speech stream is surely so small that it is difficult to pick out amidst all the other complicated and noisy variables in the system. Possibly progress on this research front will come from the adoption of methodologies that do not attempt to measure speech onset. Another tack could be to see whether feedback effects are more characteristic of less experienced readers. One could study children, or teach adult subjects an artificial symbol system with the crucial properties of English- or French-like orthographies. This latter approach would have the additional advantage of letting the experimenter disentangle feedback consistency measures from the many other factors that facilitate or inhibit the fluent reading of single words.

We would be as intrigued as anyone to eventually see convincing proof that reading activates the same sound-encoding processes that were developed for use in writing. However, our appraisal of the current state of knowledge, along with our reanalyses of several megastudies using our new, stronger, controls on phonetic voice key biases, leads us to conclude that there is no reason to require models of oral reading to predict feedback consistency effects, whether by making networks recurrent or by other means.

References

Balota, D. A. (1994). Visual word recognition: The journey from features to meaning. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 303–348). San Diego, CA: Academic Press.

Balota, D. A., et al. (2005). *English Lexicon Project web site*. Retrieved April 1, 2005, from the Washington University in St. Louis web site, http://elexicon.wustl.edu

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General, 133,* 283–316.

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornič (Ed.), *Attention and performance VI* (pp. 535–555). Hillsdale, NJ: Erlbaum.

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

Cortese, M. J., & Fugett, A. (2004). Imageability ratings for 3,000 monosyllabic words. *Behavior Methods and Research, Instrumentation, & Computers, 36,* 384–387.

Cutler, A. (1981). Making up materials is a confounded nuisance: or, Will we be able to run any psycholinguistic experiments at all in 1990? *Cognition, 10,* 65–70.

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 674–691.

Henderson, L. (1989). On mental representation of morphology and its diagnosis by measures of visual access speed. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 357–391). Cambridge, MA: MIT Press.

International Phonetic Association (2005). *The International Phonetic Association.* Retrieved April 26, 2005, from http://www.arts.gla.ac.uk/ipa

Jared, D. (1997). Spelling–sound consistency affects the naming of high-frequency words. *Journal of Memory and Language, 36,* 505–529.

Jared, D. (2002). Spelling–sound consistency and regularity effects in word naming. *Journal of Memory and Language, 46,* 723–750.

Kawamoto, A. H., Kello, C. T., Jones, R., & Bame, K. (1998). Initial phoneme versus whole-word criterion to initiate pronunciation: Evidence based on response latency and initial phoneme duration. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 862–885.

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

Kessler, B., Treiman, R., & Mullennix, J. (2002). Phonetic biases in voice key response time measurements. *Journal of Memory and Language, 44,* 592–617.

Lacruz, I., & Folk, J. R. (2004). Feedforward and feedback consistency effects for high- and low-frequency words in lexical decision and naming. *The Quarterly Journal of Experimental Psychology, 57A,* 1261–1284.

Massaro, D. W., & Jesse, A. (2005). The magic of reading: Too many influences for quick and easy explanations. In T. Trabasso, J. Sabatini, D. W. Massaro, & R. C. Calfee (Eds.), *From orthography to pedagogy: Essays in honor of Richard L. Venezky* (pp. 37–61). Mahwah, NJ: Erlbaum.

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

Peereman, R., Content, A., & Bonin, P. (1998). Is perception a two-way street? The case of feedback consistency in visual word recognition. *Journal of Memory and Language, 39,* 151–174.

Perry, C. (2003). A phoneme–grapheme feedback consistency effect. *Psychonomic Bulletin & Review, 10,* 392–397.

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

Sakuma, N., Fushimi, T., & Tatsumi, I. (1997). Onseiha no shisatsu ni yoru kana no ondokusenji no sokutei: Ondokusenji wa gotōon no chōonhō ni yori ōkiku kotonaru. [Naming latency measurements of kana based on inspection of voice waveforms: Naming latency varies greatly depending on the manner of articulation of the initial phoneme.] *Shinkeishinrigaku, 13,* 126–136.

Seidenberg, M. S., & Waters, G. S. (1989, November). *Reading words aloud: A mega study* [Abstract]. *Bulletin of the Psychonomic Society, 27,* 489.

Solso, R. L., & Juel, C. L. (1980). Positional frequency and versatility of bigrams for two- through nine-letter English words. *Behavior Research Methods & Instrumentation, 12,* 297–343.

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. C. (1994). Building a resonance framework for word recognition using design and system principles. *Journal of Experimental Psychology: Human Perception & Performance, 20,* 1248–1268.

Stone, G. O., Vanhoy, M., & Van Orden, G. C. (1997). Perception is a two-way street: Feedforward and feedback phonology in visual word recognition. *Journal of Memory and Language, 36,* 337–359.

Treiman, R., & Kessler, B. (1995). In defense of an onset–rhyme syllable structure for English. *Language and Speech, 38,* 127–142.

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.

Zeno, S. M., Ivenz, S. H., Millard, R. T., & Duvvuri, R. (1995). *The educator's word frequency guide* [Electronic data file]. Brewster, NY: Touchstone Applied Science Associates.

Ziegler, J. C., Montant, M., & Jacobs, A. M. (1997). The feedback consistency effect in lexical decision and naming. *Journal of Memory and Language, 37,* 533–554.

Ziegler, J. C., & Van Orden, G. C. (2000). Feedback consistency effects. *Behavioral & Brain Sciences, 23,* 351–352.

Author Note

Brett Kessler and Rebecca Treiman, Department of Psychology, Washington University in St. Louis. John Mullennix, Department of Psychology, University of Pittsburgh at Johnstown.

This work was supported in part by grant BCS-0130763 from the National Science Foundation. We thank David Balota, Howard Nusbaum, Mark Seidenberg, and Melvin Yap for making their data files available to us.

Correspondence concerning this article should be addressed to Brett Kessler, Department of Psychology, Washington University in St. Louis, Campus Box 1125, One Brookings Dr., St. Louis MO 63130-4899. E-mail: bkessler@wustl.edu

Footnote

[1] Phonetic transcriptions follow the conventions of the International Phonetic Association (2005).

Table 1

Significant Contributions to Response Time and Error Rates in Step 1 Regressions

	Response time[a]				Error rate[b]			
Variable	KTM	SW	SB	ELP	KTM	SW	SB	ELP
Whole word[c]								
Frequency[d]	-***	-**	-***	-***	-***	-**	-*	-***
Familiarity[e]	-***	-*	-**	-***	-***		-***	-***
Imageability[f]				-***	-**			-**
Bigrams[g]	-***	-*						
Neighborhood[h]			-***					
Syllable parts[i]								
Letters[j]								
Onset	+*		-**					
Head								
Vowel								
Rhyme	+**	+*		+**	-***			
Coda					+***			
Phonemes[k]								
Onset	+*	+*	+***	+*			+***	
Coda		-**	-*	-*				
Written frequency[l]								

Onset	-*		-**					
Head								+*
Vowel					-***	-***	-***	-***
Rhyme	-**	-*	-*	-*		-**		-**
Coda	+**					-*		-*
Spoken frequency[m]								
Onset	+*							
Head	-**						-***	
Vowel	+***	+**	+**	+***	+**	+***	+***	+***
Rhyme			+*	+*				
Coda						+***		+**
Feedforward consistency[n]								
Onset	-***	-***	-***	-***	-*		-***	
Head		-***		-*				
Vowel					-***	-*	-***	-***
Rhyme	-*		-***	-**	-***	-**	-***	-***
Coda						-***		
Feedback consistency[o]								
Onset	—	—	—	—			-*	
Head	—	—	—	—				-**

Vowel	—	—	—	—	+***	+*	+***	+***
Rhyme	—	—	—	—				
Coda	—	—	—	—		+**		+*
R^2	.32	.26	.39	.34	.16	.05	.14	.17

Note. Each column gives the result of a separate simultaneous linear multiple regression; total variance accounted for is in last row. Signs tell effect of increased level of variable on response time and error rate, and are shown only when two-tailed significance is less than .05.

[a]Time elapsed between presentation of word and tripping of voice key after response was initiated. [b]Percentage of mispronounced responses, excluding those with outlying response times or failures to respond. [c]Measures computed over the entire word. [d]Corpus frequency in Zeno, Ivenz, Millard, and Duvvuri (1995), log transformed. [e]Scaled 1 (least familiar) to 7; most ratings are from Nusbaum, Pisoni, & Davis, 1984. Values for some words not covered by that study were supplied from a small experiment at Wayne State University. [f]Cortese and Fugett (2004). [g]Average text frequency of the two-letter sequences in the spelling (Solso & Juel, 1980), square-root transformed. [h]Coltheart's *N*: Number of words in full collegiate lexicon that differ by substituting one letter (Coltheart, Davelaar, Jonasson, & Besner, 1977); square-root transformed. [i]Measures computed for each phonological parse of the syllable: onset is all consonants before the vowel, coda is all consonants after the vowel, head is onset plus vowel, rhyme is vowel plus coda. [j]Number of letters in syllable part. [k]Number of phonemes in syllable part (constant 1 for vowel). [l]Sum of the natural log of the frequencies (per Zeno et al.) of all words in the Kessler and Treiman (2001) list that have the same spelling of this syllable part, regardless of pronunciation; square-root transformed. [m]Counting words as in footnote (m), the sum of all words in the list that have the same pronunciation of this syllable part, regardless of

spelling. [n]Counting words as in footnote (m), the count of words whose syllable part is spelled the same and has the same pronunciation, divided by the count of all words whose syllable part is spelled the same. [o]Counting words as in footnote (m), the count of words whose onset, etc., is pronounced the same and has the same spelling, divided by the count of all words whose onset, etc., is pronounced the same. Not used in Step 1 regressions on response times.