Pseudohomophone Effects in Lexical Decision: Still a Challenge for Current Word Recognition Models

Johannes C. Ziegler
Centre National de la Recherche Scientifique and Macquarie University

Dorothee Klippel
Universität Marburg

Arthur M. Jacobs
Katholische Universität Eichstätt

Computational models that implement a serial mechanism of phonological assembly predict interactions between the size of the pseudohomophone (PsH) effect and stimulus length. Models with frequency-sensitive word representations predict baseword frequency effects. These predictions were tested in a lexical-decision task. The results showed constant PsH effects across different word lengths (in favor of parallel phonological activation) and baseword frequency effects (in favor of frequency-sensitive representations). However, the baseword frequency effect was opposite of what the models predicted. This result is most easily accommodated by models that assume an orthographic verification mechanism. The plausibility of such a mechanism was further supported by the results of 2 additional experiments investigating the effects of response speed and spelling probability (feedback consistency) on the size of the PsH effect.

The pseudohomophone effect in the lexical-decision task is probably the oldest and strongest effect in support of the idea that phonological information is automatically activated in visual word recognition (for reviews see Berent & Perfetti, 1995; Frost, 1998; Van Orden, Pennington, & Stone, 1990). This effect reflects the finding that nonwords that sound like real words, so-called pseudohomophones (PsHs), are harder to reject in the lexical-decision task than matched spelling controls (SCs) that do not sound like real words. For example, participants typically take longer and make more errors when rejecting the PsH FEAL than rejecting the matched control FEEP.

This effect was first observed by Rubenstein, Lewis, and Rubenstein (1971) and Coltheart, Davelaar, Jonasson, and Besner (1977). Although the early reports of this effect were criticized for potential orthographic confounds (Martin, 1982; Taft, 1982), later studies clearly established the phonological nature and reliability of the effect (e.g., Besner & Davelaar, 1983; McCann, Besner, & Davelaar, 1988; Stone & Van Orden, 1993; Van Orden, 1991; Van Orden et al., 1992; for a thorough review, see Dennis, Besner, & Davelaar, 1985). In more recent years, the PsH effect has become one of the standard effects for developing and testing computational word recognition models (e.g., Coltheart & Rastle, 1994; Jacobs, Rey, Ziegler, & Grainger, 1998; Seidenberg, Petersen, MacDonald, & Plaut, 1996).

The standard explanation for the PsH effect in lexical decision assumes that a PsH (e.g., FEAL), by virtue of being phonologically identical to a real word (FEEL), contacts a lexical entry in the phonological lexicon. This contact slows down no decisions in the lexical-decision task because the phonological information signals the presence of a real word (i.e., the phonological form of FEEL), whereas the orthographic information signals the absence of a real word. If we assume that resolving this conflict takes time, it follows that PsHs obtain longer decision latencies than matched SCs (for an alternative explanation, see Seidenberg et al., 1996).

Current computational word recognition models differ with respect to how they account for the PsH effect. The dual-route cascaded model (DRC; Coltheart, Curtis, Atkins, & Haller, 1993; Coltheart & Rastle, 1994; Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001) assumes that nonword phonology is assembled via the nonlexical route. This route converts individual letters or letter groups into phonemes according to grapheme–phoneme conversion rules. The nonlexical route is slower than the lexical route. It only starts after a fixed number of cycles. Furthermore, the non-
lexical route operates in a serial manner. That is, each letter becomes available to the nonlexical route from left to right with a fixed number of cycles in between each letter. To simulate lexical-decision performance, the most recent version of the DRC (Coltheart et al., 2001) implements the multiple read-out mechanism proposed by Grainger and Jacobs (1996) in their Multiple Read-Out Model (MROM). If an individual word unit passes a critical activation threshold, the model makes a yes response. If no single unit passes the activation threshold before a response deadline, then the model makes a no response. The response deadline is variable, however. It moves back when the global activation in the lexical system is high.

This response mechanism allows the DRC to capture basic findings about word/nonword decisions, such as longer decision latencies for nonwords with many orthographic neighbors (N-effect; Coltheart et al., 1977). The DRC can simulate this effect because nonwords with many orthographic neighbors produce more global activation in the orthographic lexicon. This increase in global activation delays the response deadline and is responsible for longer no decisions for nonwords with many neighbors (for simulations, see Coltheart et al., 2001). The PsH effect is accounted for in very much the same way as the N-effect. Because PsHs are phonologically identical to real words, they produce via top-down feedback loops a greater amount of global activation in the lexicon than PsHs. Consequently, no decisions will be longer for PsHs than for Scs. Occasionally, it may even happen that a PsH will drive the activation of an individual word unit above the critical activation threshold. If this happens, the model will give an incorrect yes response; an error has occurred.

The multiple read-out model including phonology (MROM-P; Jacobs et al., 1998) is similar to the DRC in many ways (e.g., interactive activation, localist representations, orthographic and phonological lexicons). However, in contrast to DRC, the activation of phonology is neither delayed nor serial. At the same time as letter units activate whole word units in the orthographic lexicon, they will also activate word specific phonological units via a set of phoneme units. Letter units are connected to phoneme units not via rules but via a spreading activation mechanism that operates in parallel. Apart from this, the MROM-P uses the same decision mechanism as the MROM (Grainger & Jacobs, 1996) and the DRC (Coltheart et al., 2001). That is, nonwords are timed out and the deadline is flexible (i.e., it moves back as a function of global orthographic activation). The PsH effect is predicted by the model because PsHs, by virtue of being phonologically identical to real words, will produce greater global activation, which will delay the response threshold (for simulations of the PsH effect, see Jacobs et al., 1998).

Finally, parallel distributed models, like the model by Plaut, McClelland, Seidenberg, and Patterson (1996; PMSP) or its predecessor by Seidenberg and McClelland (1989), consist of a network of distributed representations that take orthographic patterns as input and produce phonological patterns as output. They are typically trained on a large set of words using an error-correcting algorithm. Once the model has learned, it can make lexical decisions by using some sort of error score, a goodness of fit measure between the input and the model's output. If the error score is below a criterion, the model makes a yes response; if the error score is above the criterion, the model makes a no response. With respect to PsH effects, it has been traditionally argued that models without lexical representations corresponding to individual words will have a hard time to simulate PsH effects (e.g., Besner, Twilley, McCann, & Seergobin, 1990, but see Seidenberg et al., 1996).

Because the above described models differ with respect to how phonology is represented and activated, they make different predictions about how the PsH effect interacts with two major variables: stimulus length and baseword frequency. With respect to length (as measured in number of letters), models that implement a serial mechanism of assembly (e.g., DRC) should predict that the size of the PsH-effect interacts with stimulus length because shorter items activate whole word phonology more quickly than longer items. In contrast, models that assume a parallel activation of phonology (e.g., MROM-P or PMSP) are more likely to predict no interactions with stimulus length. With respect to baseword frequency (i.e., the frequency of the real word that the PsH is derived of), models that assume that PsHs will activate frequency-sensitive whole word representations (e.g., DRC or MROM-P) are more likely to predict baseword frequency effects than models that have no frequency-sensitive whole word representations (e.g., PMSP). For the DRC and MROM-P, these predictions were confirmed in simulations using a subset of the items used in Experiment 1 (see Figure 1).

As can be seen in Figure 1, the DRC predicts an interaction between the size of the PsH effect and word length, whereas the MROM-P predicts constant PsH effects. It is worthwhile mentioning that even the MROM-P produces a small overall main effect of length for both PsHs and Scs. Such effects can result from differences in neighborhood structure because representations of shorter words typically have denser neighborhoods than those of longer words. However, for the purpose of the present study, we are not so much concerned with the overall length effect but rather with the issue of whether the size of the PsH effect varies as a function of length. With regard to this issue, the MROM-P predicts PsH effects of similar size across all word lengths, whereas the DRC predicts interactions with word length. As concerns base-word frequency, both models predict a main effect with longer latencies for PsHs based on high-frequency words but a main effect with shorter latencies for PsHs based on low-frequency words. This main effect is somewhat qualified by word length with DRC predicting the biggest base-word frequency effect for five-letter words, whereas MROM-P predicts the opposite.1

---

1 Only a subset of the items could be used because of the word frequency effect for five-letter words, whereas MROM-P predicts constant PsH effects. It is worthwhile mentioning that even the MROM-P produces a small overall main effect of length for both PsHs and Scs. Such effects can result from differences in neighborhood structure because representations of shorter words typically have denser neighborhoods than those of longer words. However, for the purpose of the present study, we are not so much concerned with the overall length effect but rather with the issue of whether the size of the PsH effect varies as a function of length. With regard to this issue, the MROM-P predicts PsH effects of similar size across all word lengths, whereas the DRC predicts interactions with word length. As concerns base-word frequency, both models predict a main effect with longer latencies for PsHs based on high-frequency words but a main effect with shorter latencies for PsHs based on low-frequency words. This main effect is somewhat qualified by word length with DRC predicting the biggest base-word frequency effect for five-letter words, whereas MROM-P predicts the opposite.2

2 A word of caution is in order. These simulations are based on a limited set of items. Although this allows us to get a fairly good idea about the models' general behavior in this domain, this does not mean that the models will always behave in exactly this way whenever confronted with PsHs of different lengths and base-word frequencies. In other words, because these simulation models are extremely sensitive to a variety of item-specific factors (e.g., the presence of higher frequency neighbors), it is not possible to simply generalize these simulation results to any other set of items. Ideally, each new set of items needs to be accompanied by a new set of simulations.
Figure 1. Predictions of German implementations of the dual-route cascaded model (DRC; Ziegler, Perry, & Coltheart, 2000) and the multiple read-out model including phonology (MROM-P; Jacobs et al., 1998). The figures present global activation values in the orthographic lexicon for pseudohomophones (PsHs) and their corresponding spelling controls (SCs) as a function of word length (Panel A) and baseword frequency (Panel B). The higher the global orthographic activation, the longer the no latencies of both models. With respect to word length, the DRC predicts an interaction between the size of the PsH effect and word length, whereas the MROM-P predicts constant PsH effects. With respect to baseword frequency, both models predict a baseword frequency effect with longer latencies for PsHs based on high-frequency words than for PsHs based on low-frequency words. F = frequency.

With regard to the empirical side of these predictions, a number of studies have investigated word length effects in general (e.g., Dorfman & Glanzer, 1988; O’Regan & Jacobs, 1992; Weekes, 1997). However, no study has investigated whether length modulates the size of the PsH effect. As concerns baseword frequency effects, more data are available (for a recent review, see Borowsky & Masson, 1999). Most studies investigated baseword frequency effects in the naming task. Some found that PsHs derived from high-frequency words were indeed pronounced more rapidly than PsHs derived from low-frequency words (e.g., Taft & Russell, 1992). However, when PsHs were mixed with nonwords in the same block, baseword frequency effects in naming seem difficult to obtain (e.g., Grainger, Spinelli, & Ferrand, 2000; Herdman, LeFevre, & Greenham, 1996; McCann & Besner, 1987; McCann et al., 1988; Seidenberg et al., 1996). In lexical decision, some have found robust baseword frequency effects (Van Orden et al., 1992), whereas others have failed to find such effects (McCann et al., 1988; Seidenberg et al., 1996). Notice that the positive findings of baseword frequency effects of Van Orden and colleagues are nevertheless opposite to the predictions of both models (see Figure 1). Although Van Orden and colleagues found faster latencies for PsHs derived from high-frequency words, both models predict more global activity, hence longer decision latencies, for high-frequency PsHs.

In tasks other than naming and visual lexical decision, positive findings of baseword frequency effects have been reported in semantic categorization and proofreading (Van Orden et al., 1992) and in the letter search task (Ziegler, Van Orden, & Jacobs, 1997). Finally, baseword frequency effects were reliably obtained in the phonological lexical decision task, in which participants have to decide whether a visually presented nonword sounds like a real word (e.g., Grainger et al., 2000; Taft & Russell, 1992).
The existence of baseword frequency effects has strong theoretical ramifications. Positive findings of baseword frequency effects were often taken to suggest that PsHs activate frequency-sensitive word representations in the mental lexicon (Taft & Russell, 1992). As argued earlier, such an explanation is problematic for models of the lexicon that use distributed representations (e.g., Plaut et al., 1996; Seidenberg & McClelland, 1989).

Not surprisingly, Seidenberg and colleagues (1996) tried to demonstrate that earlier findings of baseword frequency effects in naming did not replicate. Being successful in this attempt at first, their results have recently been criticized, however. Borowsky and Masson (1999) showed that Seidenberg et al. (1996) were not likely to find a baseword frequency effect with their stimulus material because the real words, from which the PsHs were derived, did not produce a word frequency effect to start with. Taken together, although the empirical situation concerning baseword frequency effects is not completely settled, both DRC and MROM-P predict baseword frequency effects in lexical decision. According to the models, PsHs derived from high-frequency basewords should produce longer no latencies than PsHs derived from low-frequency words. In contrast, the PMSP model should not predict baseword frequency effects.

In sum, Experiment 1 investigated whether stimulus length and baseword frequency affect the size of the PsH effect in lexical decision. This manipulation was of particular interest as current models of word recognition make opposite predictions concerning these variables. With respect to word length, the MROM-P clusters together with the PMSP model to predict no interactions with word length, whereas the DRC predicts an interaction. With respect to baseword frequency, the MROM-P clusters together with the DRC to predict baseword frequency effects, whereas the PMSP model predicts no baseword frequency effect. The present experiments were done in German, a language with a phonologically deep writing system but relatively consistent spelling-to-sound mapping (Jacobs, 1999; Scheerer, 1987; Ziegler, Perry, & Coltheart, 2000).

Because similar phonology effects were previously found in a letter search task for both German and English, it seems that these languages can be compared with respect to phonology's role in word recognition (Ziegler & Jacobs, 1995; Ziegler, Van Orden, et al., 1997).

**Experiment 1**

**Method**

Participants. Thirty-two undergraduate psychology students at the Philipps-University of Marburg participated in the experiment. All were native speakers of German. They received course credit for their participation. All reported normal or corrected-to-normal vision. Participants were tested individually.

Stimuli. The critical stimulus set contained 192 stimuli (96 German words and 96 nonwords). An additional 48 stimuli (24 words and 24 nonwords) served as fillers; they were excluded from all analyses. Of the 96 nonword trials, half were PsHs and half were SCs. PsHs were derived from basewords by changing (i.e., replacing, adding, or removing) one letter so that stimulus phonology, but not spelling, was identical to that of the baseword. SCs were directly matched to their corresponding PsHs and basewords. That is, they were derived from the same basewords as the PsHs and matched in terms of number of letters and syllables. They were constructed by changing the same letter (at the same position) as it was done in the case of the PsH (e.g., the SC SARL and the PsH SAHL were derived from the same baseword SAAL). Where possible, a vowel was replaced by another vowel and a consonant was replaced by another consonant. Because of these tight construction principles, PsHs and SCs have identical scores on standard orthographic similarity indexes (e.g., Kwantze & Marmurek, 1994; Weber, 1970).

Of the 48 PsHs, an equal number (16) contained three, four, and five letters. Half of the PsHs of each length were derived from high-frequency words, that is, they had a high baseword frequency (365, 388, and 309 occurrences per million, for 3-, 4-, and 5-letter PsHs, respectively). The other half of the PsHs had a low baseword frequency (10, 13, and 8 occurrences per million, for 3-, 4-, and 5-letter PsHs, respectively). Frequency estimates were taken from the CELEX database (Baayen, Piepenbrock, & van Rijn, 1993). Frequencies greater than 1,000 occurrences per million were truncated to 1,000 per million. PsHs and SCs were matched on the number of orthographic neighbors (3.1 vs. 3.4 neighbors for PsHs and SCs, respectively).

Of the 96 word stimuli, one third were three letters long, one third four letters long, and one third five letters long. Half of the word stimuli of each word length were of high word frequency (337, 310, and 316 occurrences per million for 3-, 4-, and 5-letter words respectively); the other half was of low word frequency (9.3, 9.4, and 9.1 occurrences per million for 3-, 4-, and 5-letter words, respectively).

Procedure. Participants were seated in front of a computer screen at a distance of approximately 50 cm and were given written instructions. They were told that they were going to see letter strings, some of which were German words and some were nonwords. Their task was to decide as accurately and as quickly as possible whether the letter string formed a word of their language or not. Participants gave their response by pressing one of two buttons on the computer keyboard. No feedback was given. Participants were presented with 20 practice trials consisting of 10 words and 10 nonwords to familiarize them with the task. Any questions concerning the task were answered. The experimental trials were presented in a randomized order for each participant.

Each trial began with a 500-ms presentation of a fixation mark (·) in the center of the screen. The fixation mark was replaced by the stimulus, which remained on the screen until the participants gave their response. If participants did not give a response within 2,000 ms, the next trial was initiated. The stimuli were displayed in black on a white background. They were typed in uppercase letters using a standard Macintosh font (Courier, size 24). The experiment was controlled by an Apple Macintosh 7200/90. Latency was measured from the onset of the stimulus until the participant's response. The experimental session lasted about 20 min.

**Results**

Three participants were excluded for high error rates (above 35%). Two PsHs (LEHR and LAKEI) were mistaken for their basewords by over 75% of the participants. They were excluded along with their corresponding SCs. Latencies that were three standard deviations (SDs) beyond the individual grand mean were identified as outliers (2.32% of the PsH latencies, 0.59% of the SC latencies). Because the percentage of outliers was not evenly distributed across experimental conditions, there was reason to believe that a portion of the PsH outliers were not true outliers in the sense of being unrelated to the experimental manipulation but reflect some genuine cognitive processes related to the rejection of PsHs. Therefore, outlier reaction times (RTs) were not excluded but truncated by the 3SD cut-off value. The data were submitted to

Incidentally, recent simulations showed that models with frequency-sensitive representations were indeed able to predict Taft and Russell's (1992) baseword frequency effects in naming (Coltheart et al., 2001).
an analysis of variance (ANOVA) with participants ($F_1$) and items ($F_2$) as random variables. In the participant analysis, homophony (PsHs vs. SCs), word length (three, four, five), and baseword frequency were treated as within-subject factors. In the item analysis, homophony was treated as within-item factor, and length and baseword frequency were between-item factors. The theoretically relevant RT effects are illustrated in Figure 2.

**Latency data.** The ANOVA on RTs exhibited significant main effects of homophony, $F_1(1, 28) = 27.4, p < .0001$, and $F_2(1, 40) = 39.5, p < .0001$; length, $F_1(2, 56) = 58.7, p < .0001$, and $F_2(2, 40) = 11.5, p < .0001$; and baseword frequency, $F_1(1, 28) = 31.5, p < .0001$, and $F_2(1, 40) = 9.8, p < .01$. None of the interactions approached significance, all $F$s < 1.

The homophony main effect reflects the finding that PsHs took longer to reject than their matched controls (735 vs. 688 ms, respectively). Importantly with regard to our predictions, this main effect was not qualified by an interaction with length; that is, the PsH effect was of similar size (about 50 ms) across all word lengths. This result is presented in Figure 2A. As can also be seen in this figure, the data exhibited systematic length effects with shorter items being quicker to reject than longer items. Finally, the main effect of baseword frequency reflects the finding that items derived from high-frequency basewords were faster to reject than items derived from low-frequency basewords (695 vs. 727 ms). Again, there were no interactions with length; that is, the size of the baseword frequency effect was approximately identical (about 50 ms) across all word lengths. This result is illustrated in Figure 2B.

Although SCs were derived from the same baseword as their yoked PsH, one could argue that only PsHs, but not SCs, have a unique baseword, namely, their sound-alike word mate. We therefore carried out an additional analysis that looked at the baseword frequency effect on PsH trials only. Again, we found a significant baseword frequency effect with high-frequency PsHs being rejected more quickly than low-frequency PsHs, $F_1(1, 28) = 15.2, p < .001$, and $F_2(1, 40) = 6.4, p < .05$.

**Accuracy.** The error data exhibited significant main effects of homophony, $F_1(1, 28) = 34.1, p < .0001$, and $F_2(1, 40) = 10.8, p < .01$, and baseword frequency, $F_1(1, 28) = 14.5, p < .001$, and $F_2(1, 40) = 7.7, p < .01$. The main effect of length was significant by participants but not by items. $F_1(2, 56) = 5.5, p < .05$, and $F_2(2, 40) = 1.6, p > .19$. The effect of homophony was qualified by an interaction with baseword frequency, $F_1(1, 28) = 40.8, p < .0001$, and $F_2(1, 40) = 6.9, p < .05$. No other interaction was significant by participants and items.

The main effect of homophony reflects the finding that participants committed more errors when rejecting PsHs than when rejecting their corresponding SCs (8.5% vs. 3.1%, respectively). The absence of an interaction with word length suggested that the size of the PsH effect was constant across word lengths (6.8% vs. 2.8%, 7.3% vs. 2.8%, and 11.3% vs. 3.9%, for 3-, 4-, and 5-letter items, respectively). Thus, this part of the error data mirrored the RT data.

The main effect of baseword frequency reflects the finding that low-frequency PsHs produced more errors than high-frequency PsHs (7.8% vs. 3.8%, respectively). This main effect was qualified by a significant interaction with homophony. This interaction resulted from the fact that the baseword frequency effect on errors was entirely produced by the PsHs (12.7% vs. 4.3% for high- vs. low-frequency PsHs, respectively). Post hoc analyses showed that the baseword frequency effect was significant on PsH trials only, $F_1(1, 28) = 27.6, p < .0001$, and $F_2(1, 40) = 8.5, p < .01$. No baseword frequency effect was obtained for SCs (3.3% vs. 3.0% for high- vs. low-frequency SCs, respectively). Finally, the absence of an interaction between baseword frequency and length suggests that the direction and size of the baseword frequency was not affected by word length. The size of the baseword frequency effect was similar across all word lengths (13.2% vs. 0.4%, 10.2% vs. 4.0%, 14.7% vs. 8.5%, for 3-, 4-, and 5-letter items, respectively).

**Word trials (yes latencies).** It is worthwhile pointing out that the data on word trials (yes responses) exhibited a significant Length $\times$ Frequency interaction with reliable and systematic length effects for low- but not for high-frequency words (for similar results in naming, see Weekes, 1997). Because the focus of the present article is on PsH effects, however, the length effects on...
word trials are not discussed any further in this article. Nevertheless, the item means for the word trials can be downloaded.4

Discussion

Experiment 1 replicated the PsH effect in a lexical-decision task in German: PsHs produced longer no latencies and more errors than matched SCs. This effect occurred despite the fact that items were constructed according to strict principles advocated by Martin (1982) and Taft (1982). That is, PsHs and SCs were identical in terms of orthographic similarity and neighborhood size; both were constructed with respect to the same baseword by changing a single letter. The results showed an overall length effect, no interaction between the size of the PsH effect and word length, and a baseword frequency effect. These effects are discussed below.

Length effects. The most important result is that the size of the PsH effect was not affected by stimulus length. That is, 3-, 4-, and 5-letter PsHs produced almost identical PsH effects (about 50 ms) in each of the length conditions. As seen in Figure 1B, the DRC predicted interactions with word length by virtue of the serial assembly of phonology, whereas the parallel MROM-P predicted constant effects across all word lengths. Thus, the results are more compatible with the predictions of the MROM-P. To the extent that the PMSP model would predict a PsH effect in lexical decision to begin with, the PMSP is more likely to join the predictions of the MROM-P. Therefore, it seems that models in which phonology is activated in parallel are more appropriate to capture lexical decision performance in German than models that activate phonology in a serial manner. One caveat, however, is that the present versions of both the German DRC and MROM-P operate with the original parameter set developed for English (see Jacobs et al., 1998; Ziegler et al., 2000). It is possible that some of the linguistic properties of German (e.g., the more regular orthography—phonology mapping) may require the development of a German-specific parameter set. Whether a parameter set could be found that allows the models to predict the present effects is not trivial and requires additional simulation work and parameter fitting that go beyond the scope of the present study.

Baseword frequency effects: Activation versus verification. The results exhibited reliable baseword frequency effects similar to those obtained by Van Orden and colleagues in the lexical-decision task (Van Orden, 1991; Van Orden et al., 1992). PsHs derived from low-frequency basewords took longer to reject than PsHs derived from high-frequency basewords. There was no interaction with word length, which indicated that the size of the baseword frequency effect was constant across word lengths. Both the DRC and the MROM-P predicted baseword frequency effects. However, a closer inspection of Figure 1B shows that both models predict more global activation resulting in longer no decisions for high-frequency than for low-frequency PsHs. Thus, the human data are opposite to what the models predicted. It is unlikely that the reversed baseword frequency effect was specific to German or our particular item set because Van Orden obtained similar results in English (Van Orden, 1991; Van Orden et al., 1992). Thus, to the extent that baseword frequency effects can be detected in lexical decision, they are consistently going in the same direction.5

So what is wrong with the models? In both the DRC and the MROM-P, nonwords are timed out when no single word unit reaches a critical activation level before a response deadline has elapsed. The response deadline is variable and depends on how much activity there is in the lexical system (global activation). The greater the global activation, the longer the response deadline. In both models, word-specific representations are frequency sensitive. High-frequency words have higher resting levels than low-frequency words. Accordingly, when PsHs activate word units in the lexicon, word units coding high-frequency words have a head start, and thus activation rises more quickly than for units coding low-frequency units. Consequently, PsHs based on high-frequency words will always activate their basewords more strongly than PsHs based on low-frequency words. Therefore, any activation model that implements such a decision mechanism will necessarily predict stronger PsH effects for high-frequency PsHs than for low-frequency PsHs, just the opposite of what was obtained.

If an activation account will always predict the wrong direction of the baseword frequency effect in lexical decision, what other mechanism could be responsible for the effect? It has been suggested that verification mechanisms could account for some aspects of the frequency effect (e.g., Becker, 1976; Paap, McDonald, Schvaneveldt, & Noel, 1987; Paap, Newsome, McDonald, & Schvaneveldt, 1982). But if orthographic representations of low-frequency words are less accurate than those of high-frequency words, then it should be harder to verify PsH spellings derived from low-frequency words than those derived from high-frequency words. For example, it may be the case that the orthographic representation of the low-frequency word LEER is less accurate than the representation of the high-frequency word FEEL. If so, it may be harder to verify the spelling of the low-frequency PsH LEER than that of the high-frequency PsH FEAL. Therefore, a verification account will predict the correct direction of the baseword frequency effect, with PsHs based on low-frequency words being harder to reject than PsHs based on high-frequency words. Of course, these are qualitative predictions that will have to be verified by future quantitative studies. We are not aware of any implemented simulation model that includes phonological processes and spelling verification.

In summary, the present results suggest that when it comes to lexical-decision performance in German, parallel models like MROM-P seem to be more appropriate than serial models like the DRC because the former predicts no interactions with length, whereas the latter does. However, both models predict the wrong direction of the baseword frequency effect, which suggests that some sort of verification may be at work. This possibility is further investigated in Experiment 2.

Experiment 2

The reversed baseword frequency effect suggests that spelling verification may play an important role in lexical decision. If we assume that people's response speed in lexical decision was determined by the efficiency of spelling verification, then fast par-

---

4 Items and item means can be downloaded at ftp://www.maccs.mq.edu.au/pub/ziegler/JEPHPP.

5 McCann et al. (1988) reported null effects of baseword frequency in lexical decision. Given that later studies found reliable effects, it could be argued that their "failure to find evidence of a baseword frequency effect in lexical decision is possibly a Type 2 error" (D. Besner, personal communication, November 1998).
participants should show smaller PsH effects than slow participants. Such a pattern was observed by Seidenberg and colleagues (1996). In their study, PsH effects in lexical decision were found for slow readers but not for fast readers. Although such a pattern is compatible with the idea that faster readers dispose of more efficient verification procedures and more accurate orthographic representations, the same pattern could also be accounted for by the bypass hypothesis. This hypothesis assumes that skilled readers can bypass phonological information through the use of direct orthographic access, whereas beginning readers rely exclusively on phonological processes (Baron, 1973; Doctor & Coltheart, 1980).

If this was the case, then fast participants should also show smaller overall length effects, given the standard assumption that direct orthographic access occurs in parallel (McClelland & Rumelhart, 1981).

In sum, both the verification and the bypass hypothesis predict that faster participants should show smaller PsH effects than slower participants. However, the bypass hypothesis also predicts that faster participants should show smaller length effects. In contrast, if response speed was determined by the quality of orthographic representations and the efficiency of verification, then speed should affect the size of the PsH effect but not the size of the overall length effect. These predictions were initially tested in a reanalysis of the data of Experiment 1. However, the number of participants in each response speed group was too small to obtain stable differences. Therefore, in Experiment 2, the number of participants was doubled. In addition, to exclude the possibility that participants simply did not know the correct spelling of low-frequency basewords, we used only PsHs that were derived from fairly familiar basewords.

Method

Participants. Sixty students (24 men, 36 women) at the Philipps-University of Marburg participated in the experiment. None of them had participated in the previous experiment. All were native speakers of German and reported normal or corrected-to-normal vision. Participants were tested individually. They received course credit for their participation.

Stimuli. The experimental stimulus set consisted of 162 stimuli. Half of them were words (66 trials), and half were nonwords (no trials). Of the nonwords, 27 were PsHs, 27 SCs, and 27 nonword fillers. One third of each stimulus type was either three, four, or five letters long. PsHs and SCs were constructed as in the previous experiment. They were matched for orthographic similarity and neighborhood size. To exclude the possibility that baseword spellings were simply unknown to participants, we used PsHs based on fairly frequent words. Mean baseword frequency was 448 occurrences per million. Baseword frequency was matched across stimulus length, $F(2, 24) = 0.4$, ns.

Procedure. The procedure was identical to that of Experiment 1. The experimental trials were organized into three blocks of 54 trials each, with 27 words and 27 nonwords. Both the order of the three blocks and the order of the items within the blocks were randomized over participants.

Results

Three participants were excluded because of a high error rate (above 35%). Unlike in Experiment 1, outliers that were 3 SDs beyond the individual grand mean were evenly distributed across conditions. Therefore, they were simply eliminated from the analysis (less than 1.1% of the data). Participants were divided into fast, medium, and slow groups according to their overall speed in rejecting nonwords. Each group was composed of 19 participants. The data were analyzed in an ANOVA resulting from the factorial combination of homophony (PsH vs. SCs), length (3 vs. 4 vs. 5), and response speed (slow vs. medium vs. fast). In the participant analysis ($F_1$), homophony and length were treated as within-subject variables, and response speed was treated as a between-subject variable. In the item analysis ($F_2$), homophony and response speed were within-item variables, and length was treated as a between-item variable.

As seen in Figure 3A, the data showed an interaction between the participants' response speed and the size of the PsH effect. Fast participants exhibited smaller PsH effects than slow participants did. In contrast, as illustrated in Figure 3B, the overall length effect was not affected by the participants' response speed. This pattern was confirmed in the statistical analysis. The RT analysis showed a significant overall PsH effect, $F_1(1, 162) = 47.3, p < .001$, and $F_2(1, 48) = 10.7, p < .005$. More important, there was a significant interaction between the size of the PsH effect and participants' response speed, $F_1(2, 162) = 4.3, p < .05$, and $F_2(2, 48) = 3.1, p < .05$. This interaction resulted from the size of the PsH effect being smaller for fast participants than for slow participants. With respect to length, there was a significant main effect of length, $F_1(2, 162) = 18.9, p < .0001$, and $F_2(2, 24) = 7.6, p < .005$, but no significant interaction between participants' response speed and length, $F_1(4, 162) = 0.4, p > .70$, and $F_2(4, 48) = 1.1, p > .30$. Apart from a strong response speed effect, $F_1(2, 162) = 215.2, p < .0001$, and $F_2(2, 48) = 550.1, p < .0001$, no other effects approached significance.

With regard to error data, there was a significant PsH effect, $F_1(1, 162) = 96.0, p < .0001$, and $F_2(1, 48) = 29.8, p < .005$, reflecting the finding that PsH rejections were associated with more errors than SC rejections (13.8% vs. 4.2% errors). The interaction between the size of the PsH effect and speed was not significant, all $Fs < 1$, ps > .70, suggesting that errors to PsH were not significantly affected by response speed (12.1%, 16.0%, 13.2% for fast, medium, and slow participants, respectively). Although there was a tendency for fast participants to commit fewer errors, overall error rate was not significantly different across different speeds, $F_1(2, 162) = 1.8, p > .15$, and $F_2(2, 48) = 2.4, p > .10$. 

6 Because there was some interest in the question whether the presence of homophones affects the size of the PsH effect (Dennis, Besner, & Davelaar, 1985; Taft, 1982; Underwood, Roberts, & Thomson, 1988), participants in this experiment were initially divided into two experimental groups. In one group, 27 homophones were included among the 81 word (ver) trials, whereas in the other group the homophones were replaced by nonhomophonic words. The results showed that both groups produced a significant and comparable PsH effect on both RTs and errors (RTs: 34 ms vs. 29 ms; errors: 8.5% vs. 9.0% for the homophone and the nonhomophone group, respectively). The statistical analysis confirmed that there were no differences between the groups (all $Fs < 1.1$; all ps > .29). More importantly, the size of the PsH effect did not interact with group (all $Fs < .5$; all ps > .49). Thus, this finding replicates previous reports that PsH effects can be obtained even in the absence of homophones (McCann et al., 1988, footnote 3, p. 696). The two groups were collapsed for all further analyses.
Discussion

The present results clearly demonstrated that fast readers exhibited smaller PsH effects than slow readers did, a pattern previously reported by Seidenberg et al. (1996) for lexical decision in English. In addition, the results showed that the overall length effect was not modulated by participants’ response speed. This pattern is consistent with the verification hypothesis, which assumes that faster readers dispose of better spelling verification than slower readers. Although smaller PsH effects for faster participants are also compatible with the bypass hypothesis, this hypothesis would have also predicted smaller length effects for fast than for slow participants. This effect however, was not found. The size of the length effect was almost identical for slow, medium, and fast participants.

In previous research, the bypass hypothesis has been rejected on other grounds as well. Van Orden (1987) conducted a semantic categorization experiment, in which he looked at false-positive errors to homophonic foils (is BEATS a VEGETABLE?). The bypass hypothesis assumes that phonology can be bypassed for high-frequency words. Therefore, it predicted that high-frequency homophones like BEATS should be less likely to produce a false-positive error than low-frequency homophones. In contrast, the verification hypothesis predicted that the frequency of the real category exemplar BEETS rather than the frequency of the homophone foil BEATS should determine the size of the homophone effect. If the real exemplar BEETS is of high-frequency, then readers should have better knowledge of its spelling. Accordingly, they should be better at rejecting the incorrectly spelled homophone BEATS. These two alternative hypotheses were tested by manipulating the frequency of the homophone foil independent of the frequency of the real category member. Contrary to the predictions of the bypass hypothesis, high-frequency homophone foils did not produce fewer errors than low-frequency foils. In contrast, the verification hypothesis was supported: Errors to homophone foils increased as the frequency of the real category exemplar decreased.

Notice that the pattern in Van Orden’s homophone experiments mirrors the pattern obtained for the baseword frequency in Experiment 1. In both Van Orden and our experiments, the size of the homophone effect increased as the frequency of the baseword decreased. Keep in mind that a pure activation account would have predicted the opposite, namely, that the size of the PsH effect increases with increasing baseword frequency. The next experiment provides a further test of the verification hypothesis.

Experiment 3

As argued above, PsH effects have often been accounted for in terms of how strongly PsHs activate phonological forms of their corresponding basewords: The stronger a PsH activates its baseword, the stronger the effect (i.e., the activation account). However, our previous experiments together with Van Orden’s (1987) homophone experiments suggested that spelling verification plays an important role in lexical decision or semantic categorization. The present experiment attempts to further pit the activation account against the verification account. If spelling verification plays a role in lexical decision, then we would predict that spelling probability should affect lexical-decision latencies over and above activation variables, such as orthographic similarity or regularity/consistency. Spelling probability can be defined as the probability with which a phonological pattern maps into a particular spelling pattern; this is also known as feedback consistency (e.g., Stone, Vanhoy, & Van Orden, 1997; Ziegler, Montant, & Jacobs, 1997).

The logic of the experiments is as follows. Words whose phonology is predominantly spelled the same way are thought to acquire more stable spellings, thus better orthographic representations (much like high-frequency words in Van Orden’s experiments). Hence, if a PsH activates a baseword with a dominant spelling pattern, verification should be easy. In contrast, if a PsH activates a baseword with a subordinate (infrequent) spelling pattern, then verification should be harder because these words should have less stable spellings and may suffer from competition coming from the dominant spelling pattern. Therefore, if verification plays a role in lexical decision, then PsHs derived from basewords with dominant spellings should be easier to reject than PsHs derived from basewords with subordinate spellings.
German is ideal to test this prediction because there is considerable uncertainty in spelling (a phonological pattern can often be spelled in multiple ways) but very little uncertainty in reading (a given spelling pattern is almost always pronounced the same way, see Jacobs, 1999). This asymmetry makes it easy to manipulate spelling-to-sound consistency while controlling for activation variables like spelling-to-sound consistency. Thus, we used three groups of items: (a) PsHs derived from basewords with frequent spelling patterns (dominant group), (b) PsHs derived from basewords with no dominant spelling pattern (balanced group), and (c) PsHs derived from basewords with uncommon spelling patterns (subordinate group). The spelling probability manipulation may superficially resemble the orthographic similarity manipulation (e.g., Coltheart, Patterson, & Leahy, 1994; Taft, 1982). In their experiments, it was found that PsHs that were orthographically more similar to the baseword (e.g., FEAL-FEEL) produced stronger PsH effect than PsHs that were orthographically less similar (e.g., PHOCKS-FOX). However, PsHs in our experiment did not differ in terms of orthographic similarity; we derived them from basewords by changing only one letter at a given position. Thus, all PsHs should be equally effective in activating their baseword phonology. Thus, any difference between these three groups of PsHs should be due to the efficiency of verification rather than the efficiency of baseword activation.

**Method**

Participants. Forty students from the Philipps-University of Marburg (23 women, 17 men) participated in the experiment. None of them had participated in the other experiments. All were native German speakers and reported normal or corrected-to-normal vision. Participants were tested individually.

Stimuli. The stimulus set consisted of 316 items (three to five letters long). Half of the items were words and half were nonwords. Of the nonwords, 52 were PsHs, 52 SCs, and 52 nonword fillers. Half of the PsHs were derived from basewords with balanced spellings, the other half were derived from basewords with either dominant or subordinate spellings (see below). PsHs and SCs were matched in terms of stimulus length, baseword frequency, neighborhood size, and orthographic similarity to the basewords. All PsHs and SCs were consistent in terms of orthography-phonology mapping.

Spelling probability (dominant, balanced, subordinate) was estimated through statistical analyses of all German monosyllabic words in the CELEX database. In these analyses, the frequency with which a given phonology mapped onto one or more than one spelling pattern was calculated. For example, in German, the phoneme /a/ in the nucleus position can be spelled either EE or EH, with both possibilities occurring equally often in German (12 words are spelled with EE and 12 are spelled with EH). These analyses were identical to the feedback consistency analyses of Ziegler and colleagues (Ziegler, Jacobs, & Stone, 1996; Ziegler, Stone, & Jacobs, 1997) except that they were done on smaller subsyllabic units, namely, onsets (i.e., the consonant cluster before the vowel), nuclei (i.e., the vowel units), and codas (i.e., the consonant cluster after the vowel) instead of the body/rime unit.

On the basis of these analyses, a baseword was classified as balanced if its phonology mapped equally often into both spelling patterns. For example, a word with the EE or EH spelling is balanced because 12 words contain the EE spelling and 12 contain the EH spelling. A baseword was subordinate if its spelling occurred in less than one third of the words. For example, the phoneme /a/ is spelled AI in only nine words (e.g., SAITE), whereas it is spelled EI in 100 words (e.g., WEITTE). Thus, a baseword that contained the AI spelling was classified subordinate. In turn, words with phonological units that were spelled the same way more than two thirds of the time were classified dominant. For example, a word that contained the more frequent EI spelling was dominant.

Procedure. The procedure was identical to the one used in the previous experiments.

**Results**

Mean latencies and error rates for the three groups of PsHs and their corresponding SCs are given in Table 1. Because the 3-SD trimming procedure affected more PsHs (45) than SCs (20), outlier RTs were replaced by the 3-SD cut-off value (less than 0.5% of the data). The data were submitted to an ANOVA with participants (F₁) and items (F₂) as random variables. In the participant analysis, homophony (PsHs vs. SCs) and spelling probability (dominant, balanced, subordinate) were treated as within-subject factors. In the item analysis, homophony was treated as a within-item factor, whereas spelling probability was a between-item factor.

The RT data exhibited a significant main effect of homophony, F₁(1, 39) = 14.2, p < .001, and F₂(1, 47) = 4.1, p < .05. The main effect of spelling probability was significant by participants and marginally significant by items, F₁(2, 78) = 9.7, p < .001, and F₂(2, 47) = 2.3, p < .10. The interaction between these two effects was not significant. As in all previous experiments, the main effect of homophony indicates that PsHs took longer to reject than their respective controls (683 vs. 663 ms). The main effect of spelling dominance suggests that nonwords derived from basewords with dominant spelling patterns were easier to reject than nonwords derived from basewords with balanced spelling patterns, which, in turn, were easier to reject than nonwords derived from basewords with subordinate spelling patterns. The accuracy data essentially mirrored the RT data with significant main effects of both homophony, F₁(1, 39) = 26.9, p < .0001, and F₂(1, 47) = 12.3, p < .001, and spelling probability, F₁(2, 78) = 18.7, p < .0001, and F₂(2, 47) = 4.9, p < .01. No interaction was obtained between these two effects, both Fs < 1.

**Discussion**

The present results support the verification hypothesis. Nonwords whose basewords had dominant spellings were easier to reject than nonwords whose basewords had subordinate spellings. The balanced group was in between the other two groups. The activation account predicted no differences between the three groups because all PsHs, being matched on critical activation
variables, should have activated their basewords to the same extent. Thus, the present results join the finding of the reversed baseword frequency effect to suggest that efficiency of verification rather than strength of activation can account for modulations in the size of the PsH effect.

The present spelling probability effect closely resembles the feedback consistency effects previously obtained for words (Stone et al., 1997; Ziegler, Montant, et al., 1997) and homophones (Pexman & Lupker, 1999). In these experiments, it was shown that feedback inconsistent words (i.e., words with rimes that could be spelled in more than one way) obtained longer lexical-decision latencies and more errors than feedback consistent words (their rimes could be spelled only one way). In fact, in Ziegler et al.'s experiments, all feedback inconsistent words had subordinate spelling patterns; that is, they had strong spelling competitors (as indicated by a low consistency ratio, see Ziegler, Montant, et al., 1997). Together then, the results point to a common mechanism of spelling verification that plays a role in both word acceptances and nonword rejections.

General Discussion

The PsH effect in lexical decision is one of the cornerstones for the hypothesis that phonological information plays an important role in silent reading. Current simulation models of word recognition differ with respect to how they account for the effect. Serial models like the DRC predicted interactions between the size of the PsH effect and word length, whereas parallel models (e.g., MROM-P and PMSP) did not. In contrast, models with word-specific frequency sensitive representations (DRC and MROM-P) predicted baseword frequency effects, whereas models with distributed word representations were less likely to predict such effects.

The data of Experiment 1 showed no interactions between the size of the PsH effect and word length, suggesting that parallel models may be more adequate than serial models for simulating lexical-decision performance in German. Experiment 1 also exhibited strong baseword frequency effects, which seem to be a problem for distributed models like PMSP. However, it turned out that the present pattern of the baseword frequency effect also challenges the DRC and the MROM-P because both models had predicted the effect in the opposite direction. The models predicted stronger effects for high-frequency than for low-frequency PsHs, whereas the data showed stronger effects for low-frequency than for high-frequency PsHs. Similar results were previously obtained for English (Van Orden, 1991; Van Orden et al., 1992).

It was argued that the models' failure to predict the correct pattern could have been expected on logical grounds because, in activation-type models like MROM-P and DRC, high-frequency PsHs should always activate their basewords more strongly than low-frequency PsHs (see also Taft & Russell, 1992). The decision mechanism of these models should therefore always lead to longer lexical-decision latencies to high-frequency than to low-frequency PsHs. In contrast, it was argued that a verification mechanism qualitatively accounts for the reversed baseword frequency effect. The verification hypothesis assumes that high-frequency basewords have more stable orthographic representations than low-frequency basewords. Therefore, it should be easier to verify the spelling of high-frequency PsHs than that of low-frequency PsHs.

This hypothesis was corroborated in experiments by Van Orden (1987, 1991). The verification hypothesis made two additional predictions that were tested in Experiments 2 and 3. First, if fast participants have more accurate orthographic representations and better verification procedures than slow participants, fast participants should show smaller PsH effects than slow participants. For this hypothesis to be tested, a potential alternative explanation had to be ruled out, namely, that fast participants could show smaller PsH effects because they may bypass phonology and access lexical information on a purely orthographic basis. This alternative hypothesis predicted, however, that fast readers should also show smaller length effects. This prediction was not the case. Fast readers showed reduced PsH effects, but they did not show reduced length effects. Second, the verification hypothesis predicted that PsHs derived from words with dominant spelling patterns should be easier to verify than PsHs derived from words with subordinate spelling patterns. This pattern was obtained in Experiment 3. Together, then, the overall pattern seems to suggest that the PsH effect reflects not only the quality of the phonological baseword activation but also the efficiency of the verification mechanism. This pattern is more compatible with models that include both processes, activation and verification (Paap & Johansen, 1994; Paap et al., 1982, 1987).

Allowing both activation and verification to play a role can account for changes in the direction of baseword frequency effects across tasks. In naming and letter search, high-frequency PsHs produce bigger effects than low-frequency PsHs (Taft & Russell, 1992; Ziegler, Van Orden, et al., 1997). However, in lexical decision, semantic categorization, and proofreading, high-frequency PsHs produce smaller effects than low-frequency PsHs (e.g., Van Orden, 1991). Because naming is an activation-driven task, spelling verification should have little influence. Therefore, the stronger the PsH activates its baseword, the bigger the PsH effects. Similarly, letter search presumably does not require much verification. However, the stronger a PsH activates its baseword, the stronger the baseword's capacity to overrule the incoming bottom-up information. In both cases, activation alone can predict the right direction of the baseword frequency effect. In contrast, we assume that a verification component is present in lexical decision, semantic categorization, and proofreading because verification can predict the reversed baseword frequency effect in these tasks. Therefore, it seems that the crucial factor underlying the baseword frequency effect in letter search and naming is how strongly and/or quickly corresponding baseword representations are activated. In lexical decision, in contrast, the baseword frequency effect reflects how quickly PsH spellings can be verified.

7 In these experiments, participants were asked to detect a target letter (e.g., a) in a briefly presented and masked PsH (e.g., FEAL). Under the assumption that phonology plays a role during early phases of word recognition, detection errors were expected because the letter a was present in the spelling of FEAL but was absent from the spelling of the phonologically activated baseword FEEL. Such a phonological interference effect was obtained (Ziegler, Van Orden, et al., 1997), and it was stronger for PsHs derived from high-frequency basewords (e.g., FEAL, baseword FEEL) than for PsHs derived from low-frequency basewords (e.g., LEAR, baseword LEER).
Thus, both factors seem important but they may be involved to a different degree in different tasks. Activation and verification do not have to be two distinct mechanisms but can be part of a single mechanism. This is the case in the resonance account of word perception (e.g., Van Orden & Goldinger, 1994; Van Orden, Jansen op de Haar, & Bosman, 1997). According to this account, a visually presented stimulus, word or nonword, activates visual units, phonological units, and semantic units. This initial activation, however, is not sufficient to distinguish words from nonwords. Lexical decisions are based on the global coherence of activation. Coherence implies self-consistent feedback that matches bottom-up activation (i.e., a bidirectional flow of activation, a coupling). Thus, a word will be identified when the patterns of activation across different families of visual, phonological, and semantic units cohere. Known words achieve pattern match among all families of units. Nonwords also activate phonological and semantic units. However, they always entail some mismatch that inhibits global coherence and leads to a relatively disordered pattern of activation. Lexical decisions to nonwords may be based on this mismatch. Bigger mismatch leads to faster nonword rejections. This mismatch idea can explain why SCs can be rejected faster than PsHs. That is, because SCs generate less coherent phonological or semantic activity, they entail more mismatch than PsHs, and therefore they can be rejected faster than PsHs.

The resonance account can also explain the reversed baseword frequency effect and the spelling probability effect. As top-down expectations are matched (verified) against incoming bottom-up activation, PsHs will create a mismatch because of differences in spelling. Because frequent basewords (or dominant spelling patterns) provide stronger top-down expectations than infrequent basewords (or subordinate spelling patterns), this mismatch is amplified for PsHs derived from high-frequency or dominant basewords. Because stronger mismatch allows the system to make faster no responses, both effects are expected.

In a similar line of argument, Taft (1991, see also Taft & van Graan, 1998) has offered a qualitative account of phonology effects in word recognition that involves some of the mechanisms of resonance theory within an interactive-activation framework. In this framework, the orthography and phonology components are composed of a hierarchy of units ranging from graphemes (EA) and phonemes (/i:/) to whole words. Activation passes up the hierarchy as well as between orthographic and phonological units at the same level. The interface with semantics is only at the top of the hierarchy, that is, at the whole-word level. The main idea borrowed from the resonance account is that activation from phonological units can rebound back to orthographic units (i.e., the OPO rebound hypothesis). For example, while the EE unit in English will send activation to the phonological /i:/ unit, the latter will send activation back to the EA and the EE unit. OPO rebound allows the model to account for PsH or homophone effects while maintaining the possibility that access to meaning is not mediated by phonology.

The OPO rebound hypothesis correctly predicts the spelling probability and baseword frequency effect. For example, when the PsH YEER is presented, the orthographic body unit -EER will activate the phonological unit /i:/, which, in turn, will activate the orthographic unit -EAR. Because -EAR is the more dominant (common) pattern, and hence a stronger unit than -EER, it will produce greater competition than if the pattern were the other way around. If the system uses an index of competition (i.e., mismatch) as the basis for nonword rejections, then it would indeed predict that PsHs with dominant basewords were rejected faster. Such a mechanism would also predict the right direction of the baseword frequency effect as high-frequency basewords rebound more strongly into the “correct” orthographic pattern than low-frequency basewords, thus producing greater competition.

Similarly, if the current decision mechanism of DRC or MROM-P was replaced by a mechanism based on a mismatch criterion, then both models would predict the right direction of the baseword frequency effect (i.e., smaller interference effects for PsHs based on high-frequency words than for those based on low-frequency words). Such a mechanism could use mismatch between the orthographic and phonological lexicon as an index for rejecting nonwords, rather than the amount of global activation. This mechanism could work in the following way. A PsH would activate a word unit in the phonological lexicon. Because no corresponding unit is activated in the orthographic lexicon, this mismatch, provided it is big enough, will lead the system to reject the nonword. As high-frequency PsHs activate units in the phonological lexicon more strongly than low-frequency PsHs, the mismatch between the orthographic and phonological activation is greater for high- than for low-frequency PsHs. As a consequence, high-frequency PsHs can be rejected more quickly, which is the pattern that corresponds to the empirical data. In sum, although the present versions of DRC and MROM-P are not able to simulate the baseword frequency effect found in lexical decision in German and English, the implementation of a mismatch criterion may offer ways out of this dilemma. This mechanism however, requires major changes of how the models make lexical decisions.

In conclusion, the results of the present study suggest the operation of two basic mechanisms underlying word recognition performance: automatic phonological activation (as indicated by the existence of reliable PsH effects across all experiments) and spelling verification (as indicated by the reverse baseword frequency effect and the spelling probability effect). It seems that phonological activation and spelling verification are two general mechanisms underlying reading because they are found across a variety of tasks. However, it may be misleading to think of them as two separate, stage-like, strategic processes. Instead, they could be seen as feedforward and feedback aspects of the selfsame process, a coupling between orthography and phonology underlying normal skilled reading.

References


PSEUDOHOMOPHONE EFFECT


Received September 18, 1998
Revision received July 3, 2000
Accepted August 22, 2000