JOURNAL OF MEMORY AND LANGUAGE **39,** 151–174 (1998)
ARTICLE NO. ML982573

Is Perception a Two-Way Street? The Case of Feedback Consistency in Visual Word Recognition

Ronald Peereman

Université de Bourgogne and CNRS, Dijon, France

Alain Content

Laboratoire de Psychologie Expérimentale, Université libre de Bruxelles, Brussels, Belgium

and

Patrick Bonin

LAPSCO, Université Blaise-Pascal, Clermont-Ferrand, France

It is generally assumed that during reading, the activation produced over orthographic units feeds forward to phonological units. Supporting interactive models of word recognition, Stone, Vanhoy, and Van Orden (1997) recently claimed that phonological activation reverberates to orthographic processing units and consequently constrains orthographic encoding. They found that the consistency of the relations between phonology and orthography (feedback consistency) influenced lexical decision performance. We explored the effect in five experiments conducted with French words. Although feedback consistency affected writing performance, no significant effect was observed in lexical decision even when inconsistency was defined so as to maximize the effect. We also show that previous reports of consistency effects in French (Ziegler, Montant, & Jacobs, 1997a) may be due to a confound between consistency and word frequency, as assessed by subjective frequency estimates. We conclude that there is at present little evidence that sound-to-print consistency influences orthographic encoding in visual word recognition.

Several findings suggest that language processing is accompanied by the activation of multiple knowledge sources. For example, rhyme judgments on spoken words are influenced by the orthographic similarity of the rhymes (Seidenberg & Tanenhaus, 1979) and phoneme detection is influenced by spelling (Schneider & Healy, 1993). Within the visual modality, a large body of findings demonstrates that word recognition is affected by the phonological characteristics of the letter string. Performance in semantic categorization tasks decreases when the nonmember target words are homophonic with a member of the semantic category (Jared & Seidenberg, 1991; Van Orden, Johnston, & Hale, 1988) and (nonhomophonic) homographs produce longer latencies in lexical decision (Kawamoto & Zemblidge, 1992). Also, recent observations of semantic ambiguity (Borowsky & Masson, 1996) and word imageability effects (Strain, Patterson, & Seidenberg, 1995) on naming performance suggest that semantic knowledge contributes to lexical decision and naming.

Although current models of visual word recognition stipulate parallel activation of multiple

This work was supported by the French CNRS and by a grant from the Direction générale de la Recherche scientifique—Communauté française de Belgique (ARC Grant 96/01-203). We thank Sandy Pollatsek, Guy Van Orden, and an anonymous reviewer for their constructive comments. ASCII files with item data (mean RTs, number of errors, subjective frequency, subjective consistency) can be downloaded by anonymous file transfer from ftp.ulb.ac.be/pub/packages/psyling/FBConsistency.

Address correspondence and reprint requests to Ronald Peereman, Laboratoire d'Etude des Apprentissages et du Développement, CNRS, ESA 5022, Université de Bourgogne, 6, Bd Gabriel, F-21000 Dijon, France. E-mail: peereman@u-bourgogne.fr.

0749-596X/98 $25.00

codes during word processing, they make different assumptions about the flow of information between the coding levels (see Jacobs & Grainger, 1994, for discussion). For instance, whereas the fuzzy logical model propounded by Massaro and Cohen (1991, 1994) and the horserace model described by Paap and Noel (1991) assume multiple but noninteracting sources of information, constraint-satisfaction connectionist models such as the adaptive-resonance model (Grossberg & Stone, 1986; Stone & Van Orden, 1994; Van Orden & Goldinger, 1994) or the interactive activation model and its derivatives (McClelland & Rumelhart, 1981; Grainger & Jacobs, 1996) hold that the different knowledge domains strongly interact during word processing. Since the issue of interactivity encompasses major design features distinguishing among current models, it seems appropriate to adduce more empirical evidence about the information flow between the multiple representational domains involved in word recognition. Furthermore, even if the architecture admits reciprocal connections between all levels of representation, the influence of each set of connections may vary as a function of their strength as well as the detailed time-course of processing. The purpose of the present research is to explore whether word recognition, as indexed by the lexical decision task, entails interactive activation between orthographic and phonological codes.

Interactive processes in visual word recognition have mainly been discussed in the context of the finding that letters are better identified within words than within nonwords (Reicher, 1969; Wheeler, 1970). The interactive activation (IA) model of McClelland and Rumelhart (1981; Rumelhart & McClelland, 1982) holds that activation of letter units is controlled both by perceptual events and top-down influences from word nodes. The key mechanism explaining the word advantage is that letter strings that form real words cause activation of their corresponding lexical nodes which, in turn, reverberate activation to their constituent letters. Unless the nonwords resemble words, no such feedback activation to lower processing levels is expected. Following the original IA framework, more recent models of word recognition based on cascaded processing have incorporated the word–letter interactivity assumption (e.g., Coltheart, Curtis, Atkins, & Haller, 1993; Grainger & Jacobs, 1996).

However, the above account of the word superiority effect and other related findings has been questioned on both empirical and simulation grounds (Allen, Wallace, & Weber, 1995; Jacobs & Grainger, 1992; Mewhort & Johns, 1988; Paap, Newsome, McDonald, & Schvaneveldt, 1982). Hence, it remains unclear whether reading really involves interactive processes allowing activation at a higher processing level to reverberate and influence activation at a lower processing level. Furthermore, the demonstration of interactive processing between one pair of representational layers does not imply that similar interactions exist between other pairs of layers. Even within the resonance/interactive activation framework, the extent and strength of connectivity between levels should presumably vary with their functional relevance. Hence, the significance of interactive links needs to be empirically grounded for each pair of levels separately.

Recently, Stone, Vanhoy, and Van Orden (1997) reported empirical evidence suggesting bidirectional influences between orthographic and phonological codes during word recognition. To our knowledge, Stone et al.'s study represents the first direct investigation of interactive processes between orthography and phonology in word reading. Because the present study was initiated to extend Stone et al.'s observations to the reading of French words, we will consider in detail the critical structural variables that were manipulated.

Stone et al.'s (1997) demonstration was based on a manipulation of the consistency of the mappings between phonology and orthography. They took advantage of the multiplicity and variability of the orthographic transcodings of phonological rimes in English and contrasted words varying along this dimension, which they termed *feedback consistency*. Words were categorized as feedback (FB) inconsistent when several alternative spellings existed for their rime (the phonological string comprising all phonemes except the initial consonants, e.g., HEAP, in which the rime /iːp/ can also be spelled EEP). FB consistent words were those for which a unique transcription of the rime exists in English. Stone et al. observed longer response times and more errors for FB inconsistent than for FB consistent words.

Stone et al. interpreted their findings in the framework of interactive activation/resonance models (Stone & Van Orden, 1994; Van Orden & Goldinger, 1994), which assumes reciprocal connections between orthographic and phonological units. In this framework, word identification is essentially a function of the activation dynamics of the whole set of orthographic and phonological units. The gradual evolution toward a stable pattern of activation over orthographic units is determined by the connections between orthographic and phonological units as well as between phonological and orthographic units. Thus, the interactivity assumption intrinsic to the resonance framework leads to the prediction that word recognition should not only depend on the characteristics of the mapping from orthography to phonology (which corresponds to the traditional notion of print-to-sound consistency and is hereafter designated as *feedforward* (*FF*) consistency) but also on the correspondences from phonology to orthography (feedback consistency).

The present research was undertaken to investigate whether sound-to-print consistency influences lexical decision in French. There were several reasons to launch such a project. One reason was that the stimuli of Stone et al. were not matched on several variables that are known to affect performance in lexical decision. In particular, feedback consistency was confounded with neighborhood size, a variable that influences lexical decision performance (Andrews, 1989, 1992). The problem was acknowledged by the authors and addressed through post hoc analyses, in which feedback consistency still influenced performance after the contribution of neighborhood size was partialled out. We felt it appropriate to try to replicate the FB consistency manipulation with a different stimulus set.

A second reason that motivated our interest is that Stone et al. only considered FF and FB consistency for the body/rime unit. Although this choice is easy to understand given the characteristics of the English orthography, the consistency of other units is known to be relevant, at least in the FF direction. Thus, Jared (1997) reported that grapheme-to-phoneme consistency influences phonological conversion; Treiman, Mullennix, Bijeljac-Babic, and Richmond-Welty (1995) observed a contribution of the onset consonant cluster consistency in addition to the effect of body/rime consistency; and Kay (1987) and Taraban and McClelland (1987) found effects determined by the initial consonants plus vowel group.

Given the high level of print-to-sound consistency in French, it is much easier to control FF consistency dimensions in the selection of experimental materials. This follows from the fact that, whereas French and English orthographies are equally inconsistent when phonology-to-orthography mappings are considered, the French orthography is much more consistent than English in print-to-sound correspondences (Content & Peereman, in preparation; Peereman & Content, 1997; Ziegler, Jacobs & Stone, 1996; Ziegler, Stone, & Jacobs, 1997b).

Additionally, it may be particularly appropriate to investigate the FB consistency effect on FF consistent words. Indeed, both empirical and theoretical reasons led us to expect that the FB consistency effect should be more salient on FF consistent words. In Stone et al.'s second experiment, FB consistency and FF consistency were orthogonally manipulated, and a significant effect of FB consistency was obtained on reaction times only for FF consistent words. Moreover, FF consistent words may provide a cleaner test condition because a unique phonological code corresponding to the rime is evoked, so that its sound-to-print consistency constitutes the only relevant dimension to be considered. Conversely, with FF inconsistent words (e.g., SHEATH, see Fig. 3b in Stone et al., 1997), the orthographic body activates several phonological rimes (−/iːθ/ and −/ɛːθ/), each of which is FB consistent or inconsistent. Though Stone et al. (1997) classified words according to the FB consistency of the *correct* phonological rime only, the effect might be diluted by the FB contribution of the (incorrect but dominant) inconsistent phonological rime. Thus the FB activation of -EETH caused by −/iːθ/ could possibly be counterbalanced by the FB activation of -EATH due to the −/εːθ/ rime.

EXPERIMENT 1: FEEDBACK CONSISTENCY IN LEXICAL DECISION, NAMING, AND WRITING

The aim of Experiment 1 was to examine whether feedback consistency affects reading performance. We started our investigation by manipulating consistency without regard to a particular segmentation unit, to avoid any *a priori* assumption about the nature of processing units. Whereas print-to-sound consistency for English words is higher for body/rime correspondences than for other subsyllabic units (Treiman et al., 1995), such is not the case for French (Peereman & Content, 1997). Moreover, both in English and French, sound-to-print consistency is low and analyses as a function of the subsyllabic units reveal no FB consistency advantage for rime/body correspondences over other units (Content & Peereman, in preparation). It might, therefore, be premature to assume that bodies and rimes are important units in French.

Three tasks were used with the same sets of consistent and inconsistent words matched on word frequency. Experiment 1a used a lexical decision task as in the study by Stone et al. (1997) and Experiment 1b used the naming task. To validate the selection of the words, the stimuli were presented auditorily for writing in Experiment 1c, with the expectation that a consistency effect should occur. Several observations indicate that words including phonemes that can be represented in different ways impair spelling performance both in children and adults (Alegria & Mousty, 1994; Kreiner, 1992, 1996). Participants in the naming task also performed a delayed naming task to ensure that inconsistent and consistent items did not differ in terms of articulatory ease.

Method

Participants. Sixty psychology students at the University of Bourgogne participated in Experiment 1 for course credit. All of the participants in this and subsequent experiments were native speakers of French. Twenty students took part in Experiment 1a (lexical decision), 20 in Experiment 1b (naming), and 20 in Experiment 1c (writing).

Stimuli. The target words consisted of 28 feedback inconsistent words and 28 feedback consistent words selected from the LEXOP database (Peereman & Content, 1998) in which the consistency of the C_1 (onset), Vowel, C_2 (Coda), C_1V (Lead), and VC_2 (Rime) units was computed for a set of 2449 monosyllabic French words. Most of the inconsistencies were carried by the vowel or the rime. Mean consistency on the onset, vowel, and rime appears in Table 1 together with mean word frequency, mean bigram frequency, neighborhood size, and print-to-sound consistency. Nineteen of the 28 stimulus pairs were matched for the initial phoneme and the initial letter.

As shown in Table 1, consistent and inconsistent words were also matched for subjective frequency in print, as estimated by an independent group of 22 students. Booklets including all stimulus words were prepared. Six squares were printed in front of each word. The first square was labeled ‘‘unknown’’ and the last ‘‘very frequent.’’ Students were asked to rate each word of the list for its frequency by putting a cross in the square corresponding to their choice. Instructions emphasized the requirement to estimate the frequency in printed materials specifically. Participants' judgments were converted to numerical values ranging from 1 (unknown) to 6 (very frequent). The same procedure was used to collect subjective frequency data in the subsequent experiments.

For the purpose of the lexical decision task, 56 pronounceable pseudowords were added to the list. The pseudowords matched the target words on number of letters. Half of them were feedback inconsistent whereas the others were feedback consistent.

For the writing task, the target words were

TABLE 1

Characteristics of the Words Used in Experiment 1 (mean values)

Variables	Inconsistent words	Consistent words	*p* values
Length (number of letters)	4.82	4.86	ns
Log frequency	0.79	0.78	ns
Subjective frequency	3.96	4.08	ns
Log bigram frequency[a]	2.86	2.85	ns
Number of orthographic neighbors	2.11	3.96	<.01
Number of higher frequency neighbors	1.2	2.0	.04
FF consistency of C_1	.95 (.95)[b]	.99 (.98)[b]	ns (ns)
FF consistency of V	.95 (.89)[b]	.92 (.93)[b]	ns (ns)
FF consistency of VC	.95 (.96)[b]	.99 (.100)[b]	.03 (ns)
FB consistency of C_1	.90 (.88)[b]	.97 (.98)[b]	.19 (.11)
FB consistency of V	.43 (.43)[b]	.88 (.96)[b]	.001 (.001)
FB consistency of VC	.28 (.17)[b]	.92 (.97)[b]	.001 (.001)
Position of uniqueness point (in number of phonemes)	4.18	4.68	.01
Auditory length duration (in ms)	478	494	ns

Note. Duration of the auditory stimuli used in Experiment 1c (writing); C_1, initial consonant or consonant cluster (onset); V, vowel; VC, vowel + final consonant(s) (rime). *p* values of the *t*-tests reported only when lower than .20.

[a] From Content and Radeau (1988).

[b] Consistency estimates are based on type counts (token counts estimates in parentheses).

recorded by a female talker and digitized using 16 bits analog to digital conversion at a 44.1 kHz sampling rate with the SoundEdit software on a Macintosh computer. Auditory length durations and position of the Uniqueness point (Marslen-Wilson & Welsh, 1978) of consistent and inconsistent words appear in Table 1. Despite our best efforts, there were slight differences in neighborhood size and number of higher frequency neighbors. The stimuli are provided in Appendix 1.

Procedure. The experimental items were divided into two blocks of identical length. The order of blocks was counterbalanced across participants. For all tasks, the experimental session was preceded by 20 practice trials. In lexical decision and naming, the stimuli were displayed in lower case characters on the computer screen, and presentation and timing were controlled by a PC286. Each trial began with a warning signal (a ''*'' sign) for 500 ms, followed by a blank screen for an additional 200 ms. The stimulus was then displayed at the center of the computer screen until the participant's response or for a maximum of 2 s. The intertrial interval was 2 s. Participants in the lexical decision task were asked to decide as quickly as possible whether the letter string was a word or a nonword. Responses were given via the computer keyboard in the lexical decision and via the voice key in the naming task. Participants responded to words with the preferred hand and to pseudowords with the other hand. Incorrect lexical decisions were followed by an auditory signal. Participants in the naming task were asked to pronounce each target word as fast as possible. After completing the immediate naming task, participants performed a delayed naming task with the same stimuli. The main change in procedure was that participants were asked to read the stimulus silently and to wait for a response cue before pronouncing it. The target word was displayed for 1500 ms and was followed by an empty screen during a random delay interval of either 1300, 1400, or 1500 ms. The response cue (a ''???'' sign) was then presented and the time measured until the onset of the participant's response. An auditory warning signal was presented to increase attention to the

TABLE 2

Mean Latencies (in ms) and Percentages of Errors in Experiments 1 and 2

Task		Inconsistent words	Consistent words
Lexical decision	Latencies	690	685
	Errors	11.8	10.9
Immediate naming	Latencies	533	527
	Errors	4.46	2.14
Delayed naming	Latencies	332	335
	Errors	1.1	0.9
Writing	Latencies	974	912
	Errors	21.2	9.8
Masked lexical decision	Latencies	677	677
	Errors	20.4	20.0

response cue 1 s after the disappearance of the target word.

In the writing task, each trial began with an auditory warning signal followed 500 ms later by the auditory stimulus word presented through headphones. The intertrial interval was 5 s. Presentation and timing were controlled by Psyscope 1.0.1 (Cohen, MacWhinney, Flatt, & Provost, 1993) running on a Macintosh LCIII connected with a graphic tablet (WACOM). The participants were instructed to write the stimulus word as fast as possible using a SP-210 contact pen. The time elapsing between the onset of the auditory word and the contact of the pen with the graphic tablet was recorded by the computer. After completion of the experimental session, participants involved in the naming and writing tasks were shown the 56 target words printed on a sheet of paper, and they were asked to circle unknown items.

Results

Experimental 1a (lexical decision). No decision latency exceeded 1800 ms. Mean latencies and percentages of errors are reported in Table 2. Analyses of latencies showed no significant effect of feedback consistency ($p = .55$ by subjects; $p = .83$ by items). The number of errors did not vary significantly between consistent and inconsistent words ($p = .56$ by subjects; $p = .85$ by items).

Experiment 1b (naming). Naming latencies smaller than 200 ms or longer than 900 ms in the immediate naming task, and smaller than 150 ms or longer than 900 ms in the delayed naming task, were excluded from analyses. In immediate naming, there were seven latencies smaller than 200 ms, and five latencies longer than 900 ms (two inconsistent words, three consistent words). There were 50 anticipatory responses in delayed naming with latencies shorter than 150 ms (4.5% of the observations). Only one response exceeded 900 ms in delayed naming. Words unknown to the participants were also excluded from the analyses both in immediate and delayed naming (3.6% of the observations). Mean naming latencies and error rates appear in Table 2.

Analyses of naming latencies showed that consistent words were not pronounced significantly faster than inconsistent words, either in immediate naming ($p = .11$ by subjects; $p = .80$ by items) or in delayed naming ($p = .68$ by subjects; $p = .50$ by items). In immediate naming, there was a significant effect of consistency on errors, but in the by-subject analysis only ($F_1(1,19) = 4.39$, $MS_e = .96$, $p = .05$; $F_2(1,54) = 2.22$, $MS_e = 1.36$; $p = .14$). There was no significant difference in error rates between consistent and inconsistent words in the delayed naming task.

Experiment 1c (writing). Writing latencies longer than 2000 ms were excluded from the analyses (2.0% of the data) as well as latencies for words unknown to the participants (3.6%). Omissions were counted as errors. As shown in Table 2, writing latencies were shorter for consistent than for inconsistent words ($F_1(1,19) = 22.16$, $MS_e = 1779.6$, $p < .001$; $F_2(1,54) = 4.23$, $MS_e = 11{,}968.6$, $p < .05$). Inconsistent words were also more prone to errors than consistent words ($F_1(1,19) = 45.33$, $MS_e = 1.92$, $p < .001$; $F_2(1,54) = 5.21$, $MS_e = 11.93$, $p < .05$).

Discussion

The results of Experiment 1 are straightforward. Sound-to-print consistency effects were observed in the writing task but not in lexical decision or in naming, except for a small effect on errors, which was significant only by subject. The finding that FB consistency affects writing performance is consistent with previous reports of impaired spelling performance for words including polygraphic phonemes (Alegria & Mousty, 1994; Kreiner & Gough, 1990; Kreiner, 1992, 1996).

The absence of a significant consistency effect in lexical decision does not support the hypothesis that orthographic codes are reactivated through feedback connections between phonological and orthographic units, as claimed by Stone et al. (1997). Because any effect of FB consistency depends on the prior activation of phonological units, one might assume that phonological activation had not enough time to develop to cause activation to flow back to the orthographic units. To increase the reliance on phonological information, and as a consequence, the potential influence of FB connections, Experiment 2 used a masked lexical decision task.

EXPERIMENT 2: MASKED LEXICAL DECISION

Experiment 2 was an exact replication of Experiment 1a except that words and pseudowords were presented briefly and followed by a mask. The motivation was that the degradation of the visual information might enhance the contribution of phonological coding. Several results suggest that visual degradation differentially affects orthographic and phonological codes. In a recent study, Hino and Lupker (1996) demonstrated that phonological information played a more important role in lexical decision when orthographic information was degraded by a luminance reduction. Similarly, it has been suggested that, whereas orthographic processing is highly disrupted by masking, phonological activation is relatively unaffected (Hawkins, Reicher, Rogers, & Peterson, 1976; Spoehr, 1978; Van Orden, 1987). When the target is presented, activation of orthographic processing units should spread to associated phonological units. Since the mask disrupts orthographic activation, any feedback from phonology to orthography should have a larger role than in normal visual conditions. For inconsistent words, the orthographic evidence activated through feedback should strongly compete with the target orthographic codes already weakened by the mask, so that the feedback contribution should be largely detrimental. Hence, we expected that a feedback consistency effect in lexical decision should be more easily observable when the target is masked.

Method

Participants. Eighteen students at the University of Bourgogne took part in the experiment for course credit. None of them participated in the previous experiments.

Stimuli and procedure. The words and pseudowords were identical to those used in Experiment 1. Compared to the lexical decision of Experiment 1, the main change in the procedure was that the stimuli were presented briefly before the appearance of a mask which remained on the computer screen for 300 ms.

The exposure duration of the stimuli was determined individually for each participant during a preliminary experimental phase. Participants started the experiment with a list of 66 words four to six letters long. Exposure duration decreased after each 11 trials by one screen refresh cycle (14 ms approximately). The presentation of the stimulus was immediately followed by a string of hash marks (#) matched in length to the stimulus. Duration was varied from 7 to 2 cycles. In this preliminary phase, the participants were asked to write down the letter string they saw. The exposure duration used in the lexical decision task corresponded to the smallest number of cycles that yielded more than 50% of correct identifications.

The average exposure duration used in lexical decision was 3.3 cycles (approximately 47 ms) and varied between 2 and 4 cycles. Response latencies were recorded from the onset of the mask. The lexical decision began with 28 practice trials. The other aspects of the procedure were identical to Experiment 1a.

Results and Discussion

Two latencies corresponding to consistent words were excluded from the analyses because they were longer than the 1800 ms deadline criterion. The mean lexical decision latencies and the percentages of errors appear in Table 2.

Mean latencies were identical for consistent and inconsistent words and the small difference in error rates was not significant (all *p*s = .80). Hence, although the masking procedure of Experiment 2 was thought to disrupt orthographic unit activation and increase the role of phonology-to-orthography feedback, orthographically inconsistent words were responded to as fast as consistent words.

EXPERIMENT 3: RIME/BODY CONSISTENCY IN LEXICAL DECISION

Experiment 3 was planned to examine further the role of feedback consistency in lexical decision with a new set of stimuli. First, to select word targets more similar to those used by Stone et al. (1997), feedback consistency was defined exclusively with regard to rime/body correspondences. Second, since word selection in Experiment 1 was performed using the Lexop database (Peereman & Content, 1998), the inconsistency of some rime/body correspondences might depend on the particular algorithm employed to parse orthographic strings into onsets and bodies. There were indeed several small differences in the way syllables were segmented in Lexop and in the consistency analysis reported by Ziegler et al. (1996). The stimuli used in Experiment 3 were therefore selected according to both the Lexop estimates and Ziegler et al.'s (1996) tables. Third, the use of the naming and the writing tasks in Experiment 1 imposed several additional constraints on stimulus selection and decreased the number of potential stimuli. Larger sets of words were used in Experiment 3.

Relative to Experiment 1a, Experiment 3 also introduced one minor methodological improvement in evaluating lexical decision performance. When low-frequency words are used, a negative response in lexical decision can result from the fact that the word is simply unknown to the participant. Worse, a ''yes'' response can be attributed to words which are unknown because of the emphasis on response speed. From the participant's point of view, such responses are errors even if they are correct from the experimenter's standpoint. Each participant was asked to circle unknown items at the end of the session, and responses corresponding to unknown words were discarded from the analyses.

Method

Participants. Twenty students at the University of Bourgogne took part in the experiment for course credit. None of them had participated in the previous experiments.

Stimuli and procedure. The stimuli included 45 FB consistent and 45 FB inconsistent low-frequency words selected from the Lexop database. Only 12 of the inconsistent items were used in the previous experiments. Words were categorized as FB consistent or inconsistent exclusively as a function of rime/body consistency using both the Lexop database (Peereman & Content, 1998), and Ziegler et al.'s (1996) consistency tables. All words were FF consistent and were four to six letters long. Each of the FB inconsistent words was individually matched to a consistent word as similar as possible for length, frequency, bigram frequency, and neighborhood size. Descriptive statistics about stimulus sets are shown in Table 3. The stimuli appear in Appendix 2.

As in Experiment 1, stimuli were assessed for subjective frequency by an independent group of 28 students from the same population. They were required to estimate printed word frequency using the same six-point scale as previously.

Ninety pronounceable pseudowords were added to the stimulus list. The experiment

TABLE 3

Characteristics of the Words Used in Experiment 3 (mean values), Mean Lexical Decision Latencies (in ms), and Percentage of Errors

Variables	Inconsistent words	Consistent words	*p* values
Length (number of letters)	4.76	4.76	ns
Log frequency[a]	0.59	0.53	ns
Subjective frequency	3.74	3.98	.17
Log bigram frequency[b]	2.89	2.83	ns
Number of orthographic neighbors	2.26	2.62	ns
Number of higher frequency neighbors	1.30	1.15	ns
Body/rime consistency[c]	.99 (1)	1 (1)	.19 (ns)
Rime/body consistency[c]	.16 (.05)	1 (1)	.001 (.001)
Mean latency	687	680	
Error rate	7.7	5.2	

Note. *p* values of the *t*-tests reported only when lower than .20.
[a] From Imbs (1970).
[b] From Content and Radeau (1988).
[c] By type (by token in parentheses).

started with an additional set of 20 practice trials. In all other respects, the procedure paralleled that of Experiment 1a.

Results and Discussion

Because of an error in stimulus encoding, one inconsistent word was misspelled (TORS) and this item was thus removed from the analysis together with its matched consistent word (VOLT). Overall, 4.8% of the data for the inconsistent words and 4.1% for the consistent words corresponded to words unknown to the participants. Responses to the words unknown by the participants were discarded from the analyses. Three words were excluded from the analyses because they were often declared as unknown (eight participants for the words LUTH and STELE and 16 participants for the word BONZE), and too few correct RTs remained. Finally, RTs smaller than 200 ms or larger than 1800 ms were also excluded from the analyses. The deadline criteria led to the discarding of only one long reaction time for an inconsistent word. Mean lexical decision latencies and percentages of errors as a function of word consistency are reported in Table 3.

The analyses on latencies revealed no significant effect of FB consistency ($p = .36$ by subjects; $p = .73$ by items). In the analyses on errors, FB consistency was nearly significant by subjects ($F(1,19) = 3.95$, $MS_e = 2.52$, $p = .06$), but not by items ($p = .13$). Thus, as in the previous experiments, Experiment 3 suggests that feedback consistency does not affect lexical decision latencies. Unlike Experiments 1a and 2, there was a small, nonsignificant trend in favor of a consistency effect on error rates. However it is unclear how to explain that finding given that analyses of latencies showed that the 5-ms advantage of consistent over inconsistent words was far from significant. Within the framework of the resonance model (Stone & Van Orden, 1994; Van Orden & Goldinger, 1994), feedback activation of orthographic codes from phonology should influence the buildup of orthographic activation and consequently slow word recognition. Inconsistent feedback might occasionally give rise to strong activation of an incorrect orthographic pattern and thus give rise to errors. There seems to be no room, however, for the observation that correct responses (representing more than 90% of the trials) were not affected by consistency.

EXPERIMENT 4: INCREASING ORTHOGRAPHIC MISMATCH BETWEEN THE BODIES

In the previous experiments, feedback consistency was estimated as the proportion of words including a particular phonological unit with the same orthographic counterpart among all words containing the phonological unit. For instance, type-consistency of the rime/body correspondence /ur/ -OURS was equal to .08 since only one word (COURS) contains that correspondence among the 14 words sharing the rime /ur/ (POUR, SOURD, BOURG, . . .). However, such estimates do not reflect the number of letters in common between the different orthographic renderings for a particular phonological unit. For example, the monophonemic rime /ɑ̃/ occurs in 45 words with 13 different orthographic transcriptions (CLAN, BLANC, CAMP, VENT, DANS, FAON, SANG, GENS, GRAND, TEMPS, . . .). Among them, the orthographic bodies -AMP and -ANG each occur in two words. Hence both rime/body correspondences are equally inconsistent (.04). Nevertheless, the letter N occurs much more frequently than the letter M within the different orthographic counterparts of the rime /ɑ̃/. In second serial position, M occurs in three words while N occurs in 39 words. Therefore, although rime/body consistency values are identical, the letter N is much more likely to occur than the letter M. Hence, over and above consistency values for the different units considered, orthographic encoding might be facilitated through the phonology-to-orthography loop for letters that are more likely to occur (such as N for /ɑ̃/). This raises the possibility that we failed to observe feedback consistency effects in the previous experiments because most letters of the inconsistent words occur frequently in the alternative spellings. The aim of Experiment 4 was to examine whether a feedback consistency effect emerges when inconsistent words are chosen carefully with respect to the probability of occurrence of individual letters.

Method

Participants. Twenty-three psychology students from the University of Bourgogne participated in the experiment for course credit. None of them was involved in the previous experiments.

Materials and procedure. A set of 24 inconsistent words and 24 consistent words four to six letters long was extracted from the LEXOP database. As in Experiment 3, sound-to-print consistency was defined by reference to the body. Inconsistent words thus included atypically spelled rimes. In addition they were selected so that the letters within the bodies were rarely represented in the same position in the alternative spellings. Letter probabilities for inconsistent rime/body correspondences were estimated as follows. For each correspondence we computed the number of times a particular letter occurs at a specific position within all possible orthographic renderings and divided by the number of times the same rime unit occurs in the whole lexical corpus. For example, given the rime /ɑ̃/, the probability of the letter N in the second position (as in -ANG) was equal to .87 (39/45), and the probability of the letter M in the same position (as in -AMP) was equal to .07 (3/45). These estimates are also sensitive to differences in orthographic length between the alternative orthographic codes and to the absence of letters at specific positions. Inconsistent words included rime/body correspondences for which letter probability was less than .25 for at least two positions. Note that such constraints strongly reduce the number of potential stimuli. Among the 2449 words of the LEXOP corpus, there were only 147 words that met the criteria. Not surprisingly, these words were the most feedback inconsistent.

Both feedback inconsistent and consistent words were perfectly consistent on body/rime correspondences and were of low frequency. Stimulus characteristics appear in Table 4, together with the mean subjective frequency as estimated by an independent group of 34 students. Appendix 3 provides the full list of stimulus words.

The stimulus words were matched for length with 48 pseudowords. The pseudowords were all pronounceable and they varied in number of orthographic neighbors and in rime/body consistency. The experiment started with a list of 20 practice trials. After completion of the ex-

TABLE 4

Characteristics of the Words Used in Experiment 4 (mean values), Mean Lexical Decision Latencies (in ms), and Percentage of Errors

Variables	Inconsistent words	Consistent words	*p* values
Length (number of letters)	5.04	5.04	ns
Log frequency[a]	0.77	0.74	ns
Subjective frequency	3.48	3.59	ns
Log bigram frequency[b]	2.88	2.90	ns
Number of orthographic neighbors	1.3	1.8	ns
Number of higher frequency neighbors	0.33	0.83	.03
Body/rime consistency[c]	1 (1)	1 (1)	ns (ns)
Rime/body consistency[c]	.11 (.03)	1 (1)	.001 (.001)
Mean latencies	685	680	
Error rates	7.2	6.9	

Note. p values of the *t*-tests reported only when lower than .20.
[a] From Imbs (1970).
[b] From Content and Radeau (1988).
[c] By type (by token in parentheses).

perimental list, knowledge of the experimental words was assessed as in previous experiments. The other aspects of the procedure were identical to Experiment 1.

Results and Discussion

Words declared as unknown by the participants were removed from the analyses. This led to the rejection of 1.8% of consistent words and 3.6% of inconsistent words. RTs smaller than 200 ms or larger than 1800 ms were also excluded from the analyses. The deadline criteria led to the discarding of only one long reaction time on a consistent word. Mean lexical decision latencies and percentages of errors as a function of word consistency are reported in Table 4.

Analyses of variance indicated that neither lexical decision latencies nor error rates differed between consistent and inconsistent words ($p = .58$ by subjects and $p = .78$ by items, for the analyses on latencies; $p = .77$ by subjects and $p = .86$ by items, for the analyses on errors).

COMBINED ANALYSES

Taken together, Experiments 1 through 4 failed to demonstrate a significant effect of sound-to-print consistency on lexical decision performance. However, as the reader may have noticed, small trends in the expected direction were observed in most experiments. To increase the power of the statistical analyses, combined analyses were conducted on the lexical decision performance from Experiments 1a, 3, and 4. The data from Experiment 2 were not included because they were collected with a different procedure (masked lexical decision) and because the stimuli were identical to those of Experiment 1a.

Dependent variables. Both mean latencies and mean percentages of errors by items were used as dependent variables. Two different analyses were performed. The first was based on raw data, as a preliminary examination indicated that there was no significant difference between the three experiments in terms of overall speed or error rate. The second used a procedure proposed by Massaro and Cohen (1994) in which the dependent variable corresponded to the performance for each item minus the experiment grand mean. As both analyses yielded similar conclusions, only the latter will be described.

Words used in the analyses. When words

TABLE 5

Partial Correlations Between the Seven Predictors and Lexical Decision Performance to Inconsistent Words in Experiments 1, 3, and 4

Independent variables	Partial correlation with latencies	Partial correlation with error rates
Number of letters	−.01	−.17
Number of orthographic neighbors	.03	−.05
Subjective word frequency (*z* score)	−.63**	−.45**
Log word frequency	−.06	−.06
Log bigram frequency	.08	.24*
Number of higher frequency neighbors	−.10	−.05
Rime/body consistency	−.06	−.03

* $p = .05$.
** $p < .0001$.

appeared in more than one experiment, the lexical decision performance used in the analyses was determined randomly. The five items excluded in Experiment 3 were also discarded. The remaining set of words included 75 consistent words and 72 inconsistent words.

Analyses and Results

Preliminary analyses were carried out to assess whether word sets were comparable for objective and subjective word frequency, number of letters, bigram frequency, number of neighbors, and number of higher frequency neighbors. In spite of the efforts deployed in stimulus selection, there was a small but significant difference in number of orthographic neighbors between consistent and inconsistent words (3.0 and 2.0 in average, respectively; $t(145) = 2.95$, $p < .01$). In addition, consistent words tended to have more higher frequency neighbors than inconsistent words (1.5 and 1.0, respectively; $t(145) = 2.06$, $p < .05$). Because both neighborhood variables have been shown to affect lexical decision performance (Andrews, 1989, 1992; Grainger, 1990), the analyses contrasting performance to consistent and inconsistent words were performed using both the number of neighbors and the number of higher frequency neighbors as covariates. In spite of the large number of observations used, there was no reliable effect of consistency either for latencies (701 and 691 ms for inconsistent and consistent words, respectively; $p = .65$) or for errors (9.0 and 6.7, $p = .16$).

Several findings indicate that the influence of print-to-sound consistency in reading is a matter of degree (Jared, McRae, & Seidenberg, 1990; Kay & Bishop, 1987; Laxon, Masterson, & Coltheart, 1991; Peereman, 1995). Similarly, it might be argued that the detrimental effect of feedback consistency would be more apparent for very highly inconsistent words. Therefore, the first aim of the combined analysis was to investigate whether performance varied as a function of the degree of inconsistency. In addition to rime/body consistency (token counts), we added the following variables as predictors in the analyses: number of letters, logarithm of objective word frequency, log bigram frequency, number of orthographic neighbors, and number of higher frequency neighbors. We also included subjective word frequency as predictor after *z* transformation of each estimation. The only significant simple correlations between the dependent variables and the seven predictors were for log frequency (−.35 and −.27 for latencies and errors, respectively) and subjective frequency (−.68 and −.48).[1] Partial correlations, shown in Ta-

[1] The absence of significant correlations with the other predictors is not surprising in the present context given their limited range of variation.

ble 5, indicated a significant relation between subjective frequency and both latencies and errors. Errors were also dependent on bigram frequency. In none of the analyses did rime/body consistency correlate with lexical decision performance.

Because objective word frequency norms might give an inaccurate index of actual word frequency for low-frequency words (Gordon, 1985), several inconsistent words used in the experiments could be highly familiar to the students who participated in the experiments. If so, feedback consistency effects might have less chance to emerge as some authors have suggested that phonological constraints on word identification apply more strongly to low-frequency words (Jared & Seidenberg, 1991; Seidenberg, Waters, Barnes, & Tanenhaus, 1984; but see Van Orden, Pennington, & Stone, 1990). We therefore assessed whether consistency affected performance for words for which mean subjective frequency was smaller than 3.5. There were 25 inconsistent words and 19 consistent words involved in the analyses. The mean latencies were 761 and 753 ms for inconsistent and consistent words, respectively. The 8-ms difference was not significant ($p = .75$). Inconsistent words gave rise to more errors than consistent ones (16.3 and 13.3%, respectively), but the difference did not reach statistical significance ($p = .56$).

EXPERIMENT 5A: A REPLICATION OF ZIEGLER ET AL.'S EXPERIMENT

The main outcome of the previous series of experiments is the lack of strong positive evidence supporting the influence of FB consistency in lexical decision. Such a conclusion stands in sharp contrast to the findings of Ziegler, Montant, and Jacobs (1997a), showing FB consistency effects with French readers. The aim of Experiment 5a was to replicate Ziegler et al.'s (1997a) Experiment 2 in which feedback consistency and feedforward consistency were orthogonally manipulated. In their study, both consistency effects were significant across subjects, but not across items. However, reliable differences by subjects and items were observed with one-tailed *t*-tests when comparing feedforward consistent and inconsistent words that were feedback consistent and feedback consistent and inconsistent words that were feedforward consistent.

Method

Participants. Twenty-one psychology students from the University of Bourgogne took part in the experiment for course credit.

Stimuli. The word stimuli were identical to those used by Ziegler et al. (1997a). They consisted of 40 low-frequency words four or five letters long. Half of the words were feedback consistent and the remaining ones were feedback inconsistent. Among both the feedback consistent and feedback inconsistent words, half were feedforward consistent, and the other half were feedforward inconsistent. The four categories of words were matched for word frequency, word length, bigram frequency, number of neighbors, and number of higher frequency neighbors. More details about stimulus characteristics can be found in Ziegler et al. (1997a).

The 40 target words used by Ziegler et al. were mixed with 40 pseudowords for the purpose of the lexical decision. All pseudowords were pronounceable and matched in length with the target words.

Procedure. The presentation of the 80 experimental trials was determined randomly. A set of 20 additional trials served as practice. The other aspects of the procedure were identical to previous experiments.

Results and Discussion

As in Ziegler et al.'s analyses, no deadline criteria were applied to latency analyses. Mean correct lexical decision latencies and percentages of errors are reported in Table 6. Analyses of variance including the factors FF consistency and FB consistency were conducted on correct decision latencies and on errors. In the by-subject analysis on latencies, there were significant effects of both FF consistency and FB consistency ($F_1(1,20) = 10.89$, $MS_e = 2777.6$, $p < .01$; $F_1(1,20) = 4.65$, $MS_e = 4787.8$, $p < .05$, respectively), as well as a significant interaction ($F_1(1,20) = 9.97$; $MS_e = 2560.1$, $p < .01$). In contrast, there was no significant effect

TABLE 6

Mean Lexical Decision Latencies and Percent Errors (in Parentheses) in Experiment 5a

		FB consistency	
		Consistent	Inconsistent
FF consistency	Consistent	629 (8.6)	697 (23.8)
	Inconsistent	702 (27.6)	700 (11.9)

in the by-item analysis ($p = .42$ for FF consistency; $p = .30$ for FB consistency; $p = .14$ for the interaction). Analyses on errors revealed no significant effect of FF and FB consistency, either in the by-subject analysis ($p = .12$ and $p = .92$, respectively), or in the by-item analysis ($p = .63$ and $p = .98$, respectively).

The results of the ANOVAs replicate faithfully Ziegler et al.'s observations. The 38-ms advantage (34 ms in Ziegler et al.) for FF consistent words over FF inconsistent words and the 32-ms advantage (36 ms in Ziegler et al.) for FB consistency over FB inconsistent words reached statistical significance only in the by-subject analysis. On errors, as in Ziegler et al., no consistency effects were observed. The only different statistical outcome is that a significant interaction between FF consistency and FB consistency on errors emerged in the present experiment but was only marginally significant in Ziegler et al.

As in Ziegler et al.'s study, we carried out one-tailed *t*-tests to examine the effect of FF consistency for FB consistent words and the effect of FB consistency for FF consistent words. In the by-subject analyses, FB consistent words were responded significantly faster and with fewer errors when they were FF consistent than when they were FF inconsistent ($t(20) = 4.75$, $p < .01$ for latencies; $t(20) = 8.00$, $p < .01$ for errors). The corresponding by-item analyses yielded nearly significant differences ($t(18) = 1.47$; $p = .08$ for latencies; $t(18) = 1.66$, $p = .06$ for errors). Turning to the analyses on FB consistency for FF consistent words, it appeared that FB consistent words gave rise to shorter latencies and less errors than FB inconsistent words. The differences were significant by subjects ($t(20) = 3.10$, $p < .01$ for latencies; $t(20) = 5.12$, $p < .01$ for errors) and nearly significant by items ($t(18) = 1.67$, $p = .057$ for latencies; $t(18) = 1.51$, $p = .074$ for errors).

In short, latency analyses replicated Ziegler et al.'s observations. Although analyses on errors were not significant in Ziegler et al.'s study, the differences were marginally significant in the present experiment.

EXPERIMENT 5B: WHEN SUBJECTIVE FREQUENCY DOES THE JOB FOR CONSISTENCY

Over the past 30 years, psycholinguistic research has identified several structural variables that influence word recognition processes. Among them, word frequency has produced consistent and reliable effects in a large variety of tasks and has been shown to interact with other variables such as imageability (Strain et al., 1995), neighborhood size (Andrews, 1989; Peereman & Content, 1995), and print-to-sound consistency (Seidenberg et al., 1984), most of the effects being confined to low-frequency words. Hence, studies often use low-frequency words to assess the effects of various factors on word processing. One potential problem, already underlined by previous authors (Gernsbacher, 1984; Gordon, 1985), is that word frequency tables may be less reliable for low-frequency than for high-frequency words, because the size and nature of the text corpus is more critical to differentiate among rare words. An alternative way to assess word frequency is to collect ratings from a pool of human participants, as we did to control stimulus selection in the previous experiments. Gernsbacher's well-known study (1984) suggests that such subjective estimates might help to explain discrepancies in previous investigations of word recognition.

In the present case, a close look at Ziegler et al.'s stimuli led us to believe that FB consistent and inconsistent words were not equally common. Therefore, Experiment 5b was carried out to assess whether the word sets differed on subjective frequency ratings.

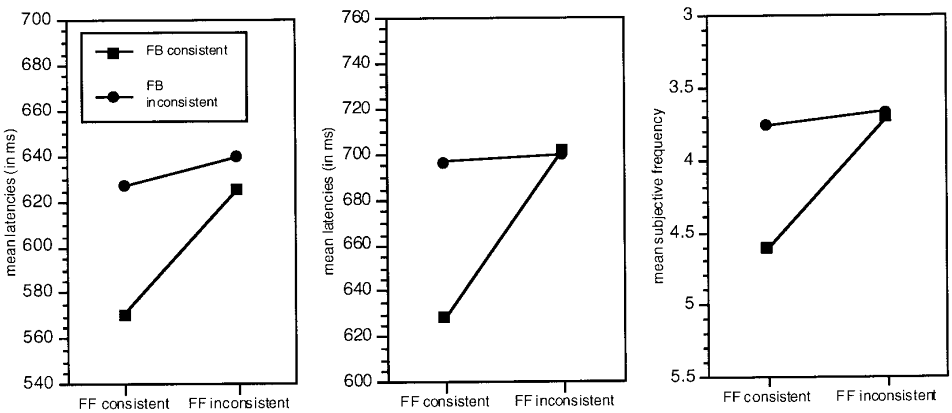

FIG. 1. Mean lexical decision latencies in Ziegler et al.'s Experiment 2 (left panel), in Experiment 5a (middle panel), and mean subjective frequency of the corresponding stimulus words (right panel). Note that the *y* axis in the rightmost graph is reversed so that lower points correspond to high frequency ratings.

Method

Participants. Eighty-three students in psychology at the University of Bourgogne were tested in a single group.

Materials and procedure. Two-page forms were prepared including all stimulus words used in Ziegler et al.'s first ($N = 46$) and second ($N = 40$) experiment. Because several words occurred in both experiments, only 75 words were used. Word order was determined alphabetically, and page order varied across the participants. Six squares were printed in front of each word. The first square was labeled "unknown" and the last "very frequent." Students were asked to rate each word for its frequency in print by putting a cross in the square corresponding to their choice. Participants' judgments were converted to numerical values ranging from 1 (unknown) to 6 (very frequent).

Results and Discussion

The results were analyzed separately for the items used in Ziegler et al.'s Experiment 1 and Experiment 2.

Stimuli used in Experiment 1. Subjective frequency ratings for feedback consistent and inconsistent words differed significantly by subjects ($F(1,82) = 164.58$, $MS_e = .055$, $p < .001$) and by items ($F(1,44) = 5.51$, $MS_e = .452$, $p <$ .05). Feedback consistent words were judged more frequent than feedback inconsistent words (means of 4.65 and 4.18, respectively).

Stimuli used in Experiment 2. The mean subjective frequency ratings appear in Fig. 1 together with the mean lexical decision latencies from Ziegler et al.'s Experiment 2 and from Experiment 5a. Analyses of variance similar to those performed in Experiment 5a were carried out on subjective frequency estimates. In the by-subject analyses, there were significant effects of feedforward consistency ($F(1,82) =$ 221.31, $MS_e = .093$, $p < .001$) and feedback consistency ($F(1,82) = 166.67$, $MS_e = .098$, $p < .001$), as well as a reliable interaction between feedforward and feedback consistency ($F(1,82) = 129.36$, $MS_e = .106$, $p < .001$). In the by-item analyses, neither feedback nor feedforward consistency reached significance. Finally, as in Experiment 5a, we performed one-tailed *t*-tests on item data to contrast feedforward consistent and inconsistent words (for feedback consistent items) and feedback consistent and inconsistent words (for feedforward consistent items). The effect of feedforward consistency was significant ($t(18) = 1.81$, $p < .05$) and the effect of feedback consistency was nearly significant ($t(18) = 1.69$, $p = .054$).

Reanalyzing Ziegler et al.'s data while controlling for subjective frequency. Because there was a confound between consistency and subjective frequency in Ziegler et al.'s experiments, it seemed appropriate to reanalyze the data on consistency while controlling for subjective frequency. The analyses were restricted to the data reported for Ziegler et al.'s Experiment 1 in which a larger set of items were used. First, we reanalyzed the item RTs introducing subjective frequency as a covariate. Not surprisingly, the ANCOVA indicated that feedback consistency was far from significant ($F(1,42) = .095$, and $p = .76$). Second, we reanalyzed the RTs after removing some items in order to match FB consistent and inconsistent words carefully on subjective frequency. Five FB consistent words with the highest subjective frequency scores and 5 FB inconsistent words with the lowest subjective frequency scores were removed. The corresponding subjective frequency scores for the remaining items were 4.43 and 4.44 for FB consistent and inconsistent words, respectively. An ANOVA performed on the remaining matched sets of items showed no significant effect of FB consistency ($F(1,34) = .295$, $p = .59$).

In sum, the results of Experiment 5b indicate that the FF and FB word sets used in Ziegler et al.'s study differed in subjective frequency. Such observations strongly suggest that the difference between consistent and inconsistent stimuli obtained in Ziegler et al.'s experiments and in Experiment 5a follow from differences in word frequency. However, it remains possible to argue that the confound is the other way round; that is, that participants are sensitive to FB consistency and rely on this dimension in evaluating word frequency. In that case, the differences observed in Experiment 5b would simply reflect the influence of consistency on word frequency ratings. This issue is addressed in Experiment 5c.

EXPERIMENT 5C: DOES FB CONSISTENCY INFLUENCE FREQUENCY JUDGMENTS?

Following Howes' (1954) seminal work, large correlations have often been reported between subjective word frequency and objective word frequency, as indexed in Thorndike and Lorge (1944) or Kučera and Francis (1967), for example (e.g., Carroll, 1971; Galbraith & Underwood, 1973; Gordon, 1985; Segui, Mehler, Frauenfelder & Morton, 1982; Shapiro, 1969; Tryk, 1968; Whalen & Zsiga, 1994). Such a relation has proved to be reliable using different methodologies for evaluating subjective frequency, such as magnitude estimation, multiple rank ordering, or a forced-choice procedure. However, little is known about the variables that affect subjective frequency judgments besides frequency of occurrence. Galbraith and Underwood (1973) reported that abstract words tended to be judged as more frequent than concrete words matched for objective frequency. This result held in the forced-choice procedure in which participants were presented with pairs of words and were asked to decide which one was the most frequent. The difference disappeared when magnitude estimations were required, unless the instructions emphasized that words had to be evaluated with regard to the number of possible contexts in which they can occur. Segui et al. (1982) had French and English participants judge open- and closed-class words for their frequency in written materials and observed very similar correlations between subjective estimates and word frequency for both types of words, suggesting that semantic characteristics do not influence frequency judgments.

Although the influence of FB consistency in frequency judgments cannot be excluded *a priori*, it is unclear how such a hypothesis would account for the observation made in Experiment 5b that subjective frequency ratings distinguished among FF consistent words but not among FF inconsistent ones. Furthermore, no difference in subjective frequency was observed between FB consistent and inconsistent words in previous experiments (Experiments 1, 3, and 4).

Nevertheless, one possible way to test the hypothesis is to examine whether subjective frequency correlates with consistency for a set of words in which there is no correlation between consistency and objective frequency. This approach was applied to a set of 122 monosyllabic French words for which subjective frequency was assessed by Flieller and Tournois (1994) using a seven-point scale. FB consistency values on rime/body correspondences was given by LEXOP (Peereman & Content, 1998). A first analysis confirmed that there was a large correlation between log

TABLE 7

Mean Frequency Ratings for the Ten Classes of Words Varying in Rime/Body Consistency in Experiments 5c and 5d, and Mean Subjective Consistency in Experiment 5d

	Consistency class									
Variables	.01–.10	.11–.20	.21–.30	.31–.40	.41–.50	.51–.60	.61–.70	.71–.80	.81–.90	.91–1.0
Experiment 5c										
FB Cons.	.03	.16	.25	.36	.46	.56	.66	.76	.84	1.0
Log freq.	1.60	1.55	1.35	1.22	1.32	1.65	1.17	1.33	1.46	1.56
Subj. freq.	4.47	4.03	3.74	3.86	3.88	4.14	3.91	3.63	4.00	4.13
Experiment 5d										
FB cons.	.01	.14	.25	.37	.47	.56	.66	.76	.84	1.0
Log freq.	.69	.82	.78	.65	.79	.78	.72	.79	.79	.69
Subj. freq.	3.82	3.66	3.57	3.51	3.71	3.40	3.90	3.25	3.60	3.63
Subj. cons.	2.48	3.58	3.86	4.35	4.22	4.52	4.41	4.62	4.39	5.35

objective frequency (from Imbs, 1971) and frequency judgments ($r = .78$, $p < .001$) but no correlation at all between log objective frequency and rime/body consistency (computed by token; $r = .01$). The critical result was that subjective frequency and rime/body consistency did not correlate significantly ($r = .04$). Hence, it does not appear that frequency estimations are affected by FB consistency.

A potential limitation in using the subjective frequency values reported by Flieller and Tournois (1994) is that their participants were not instructed to rate the words for frequency in print specifically. Therefore, the purpose of Experiment 5c was to examine further whether rime/body consistency affects subjective judgments of word frequency in printed materials.

Method

Participants. Thirty-three students in Educational Science from the University of Bourgogne took part in the experiment as a group.

Materials and procedure. The stimulus words were selected from the Lexop database. To ensure a rectangular distribution of the words along the rime/body consistency variable, we arbitrarily split the consistency continuum in ten classes (from .01 to .10, from .11 to .20, from .21 to .30, . . .) and we chose 16 words in each one. Words varied in frequency within each class, and the sets of 16 words were matched as far as possible for objective frequency (from Imbs, 1971). Mean frequency and mean consistency for each of the ten sets of 16 items are provided in Table 7. A correlation analysis performed on the 160 words indicated no relation between objective frequency and rime/body consistency ($p > .20$ by type and by token).

The 160 words were presented in different orders to the participants who were instructed to estimate the frequency of each word in written materials, along a six-point scale, as in previous experiments.

Results and Discussion

If FB consistency influences frequency ratings, words from lower consistency classes should be judged less frequent than words from higher consistency classes in spite of the fact that they were of comparable objective frequency. Table 7 shows the mean frequency values obtained for the ten classes of items. Correlation analyses indicated a large correlation between log frequency and subjective frequency ($r = -.90$, $p < .001$), but no significant correlation between FB consistency and subjective frequency ($r = -.04$, $p = .59$). Hence, the results of Experiment 5c strengthen the conclusions reached in Experiment 5b and demonstrate that the differences in subjective frequency observed in Experiment 5b are not due to FB consistency. The observation that subjective frequency and consistency do not correlate parallels the finding that subjective judgments of the number of spelling variants of phonemes occurring in words are unrelated to objective word frequency (Kreiner, 1996).

Given that the lexical decision experiments used only low-frequency items, additional analyses were conducted on the 74 low-frequency words which had a log frequency lower than 1.20. The number of items in each of the ten consistency classes was 8, 8, 8, 8, 7, 7, 8, 6, 7, and 7, from class 1 to 10, respectively. The pattern of correlations was similar to that observed for the whole set, with a significant correlation between subjective frequency and log frequency ($r = .52$, $p < .001$) and no correlation between FB consistency and either frequency estimate ($r = -.11$, $p = .33$ and $r = -.13$, $p = .26$ for log frequency and subjective frequency, respectively). In fact, the correlations indicate small trends in the direction opposite to the hypothesis. In sum, FB consistency does not seem to affect frequency ratings even within a pool of low-frequency words.

EXPERIMENT 5D: CONSISTENCY JUDGMENTS AND FREQUENCY JUDGMENTS

The purpose of Experiment 5d was to examine further whether participants are influenced by consistency when assessing printed word frequency. In Experiment 5c, subjective estimates were analyzed as a function of objective FB consistency values. One potential concern with this technique is that FB consistency, as assessed from quantitative analyses of lexical databases, might not constitute an optimal index of readers' knowledge of sound-to-print mappings. In the present experiment, participants performed both frequency and consistency ratings, on a selection of words in which, as in the previous experiment, the two critical dimensions were statistically independent. This provides a stronger test of the hypothesis that participants are capable of assessing word frequency and orthographic consistency as distinct dimensions. If frequency and consistency ratings correlate positively, it would indicate that consistency influences frequency judgments or/and that frequency affects consistency judgments. Conversely, no correlation is expected if frequency judgments are performed without regard to consistency.

Method

Participants. Fifty-three psychology students at the University of Bourgogne were tested as a single group. All were native speakers of French.

Materials and procedure. Ninety low-frequency words (less than 20 per million) served as stimuli. As for Experiment 5c, there was an identical number of words within each of the ten arbitrary classes of the consistency continuum. Mean frequency and mean FB consistency for each of the ten sets of nine words appear in Table 7. There was no correlation between word frequency and FB consistency ($r = .01$, $p = .93$).

The 90 words were randomly arranged in six different lists. A six-point scale was printed beside each word. The instructions for the familiarity judgment were as in the previous experiments. For the consistency judgment, participants were told to evaluate the sound-to-print consistency of each word along a six-point scale. Participants had to rate the word as 1 if they felt that the word spelling was very consistent, that is, when only a single spelling was possible and this spelling corresponded to the word. Words had to be rated as 6 (very inconsistent) when the word spelling was atypical and diverged from the more typical orthography that could be associated to the same sounds. Consistency values of 2 and 3 had to be used when various orthographic renderings existed, but that the word had the more typical orthographic transcription. Finally, values 4 and 5 were dedicated to words with atypical spellings. Each participant performed the frequency rating before the consistency rating. Different lists were used for the two tasks for each participant.

Results and Discussion

Subjective consistency ratings were transformed[2] to provide a scale with the same polarity as other measures. Mean frequency and con-

[2] Transformed ratings corresponded to 7 minus participants' ratings, so that 1 indicates the lowest and 6 the highest consistency value.

sistency ratings for the ten classes of words are reported in Table 7. Independent analyses revealed significant correlations between log frequency and subjective frequency ($r = .46$, $p < .001$) as well as between objective FB consistency on rime/body correspondences and subjective consistency ($r = .63$, $p < .001$). As for Experiment 5c, there was no significant correlation between subjective frequency and objective FB consistency ($r = -.04$, $p = .69$). Also, as previously observed by Kreiner (1997), there was no correlation between consistency ratings and log frequency ($r = .05$, $p = .66$). The critical correlation between frequency and consistency ratings was not reliable ($r = .09$, $p = .41$). In sum, the data converge with Experiment 5c and confirm that frequency ratings are not contaminated by FB consistency.

GENERAL DISCUSSION

The purpose of the present study was to collect evidence about the influence of FB consistency in the lexical decision task. In Experiment 1, we explored the role of FB consistency in three tasks. Whereas a clear effect was obtained in the writing task, thus validating the stimulus selection, there were no significant RT differences either in lexical decision or in naming, although a small effect was observed on naming errors. In order to enhance the contribution of reverberating activation from phonology to orthography, the same stimuli were used in Experiment 2 under degraded visual conditions, with no more success. The third and fourth experiments used more stringent criteria for stimulus selection, to no further avail. Finally, a combined analysis using the data from all lexical decision experiments failed to demonstrate a significant contribution of FB consistency, despite numerical differences in the expected direction.[3]

In Experiment 5a, we replicated Ziegler et al.'s second experiment, in which significant FB consistency effects had been observed with French materials. Although the pattern of results in our study very closely reproduced theirs, we discovered a confound between consistency and word frequency, as estimated by subjective frequency ratings. To ensure that subjective frequency judgments were not themselves biased by consistency, we collected frequency ratings for a set of words in which consistency was gradually varied while word frequency was held constant across consistency classes. In both Experiments 5c and 5d, no correlation appeared between frequency ratings and consistency, thus leading us to reject the possibility that consistency partially determines frequency judgments.

The main conclusion that emerges from the present study is that the FB consistency effect on lexical decision, if it exists at all in French, is too meager to materialize under the experimental telescope. As a matter of fact, the small trends observed in each experiment (5, 7, 5 ms, for latencies and 0.9, 2.5, and 0.3% for errors, respectively, for Experiments 1, 3, and 4) would require an unusually large sample of participants to be tested. Based on the standard deviations observed in the present experiments, we calculated that 100 to 200 participants would be required to detect a 10-ms effect with a statistical power above .90 ($\alpha = .05$). In contrast, the present studies were powerful enough to detect a 5% effect on errors ($\beta = .06$, .01, and .01, respectively, for Experiments 1, 3, and 4). Thus, at the present time, we cannot be sure that the effect is fictitious, but our results at the very least should encourage psycholinguists not to embrace the notion of bidirectional influences of consistency too hastily.

The failure to observe robust effects of FB consistency and the discovery of a possible frequency confound in Ziegler et al.'s study leads

[3] One issue that arises in presence of nonsignificant effects concerns the statistical power of the experiments. We computed power values using the GPOWER software (Erdfelder, Faul, & Buchner, 1996). Estimations of the effect size were derived from the means and standard deviations of the difference between FB consistent and inconsistent items in previously published data (Stone et al., 1997, Experiment 2; Ziegler et al., 1997, Experiments 1 and 2). Data from FF consistent conditions were used when possible. Standard deviations were computed from *t* values. For RT and error analyses, the power for a matched *t*-test on a group of 20 participants was .95 or above, except when the effect size on errors was estimated from Ziegler et al.'s Experiment 2, in which the effect was not significant. Thus we conclude that our studies are suitable to detect an effect of the size previously reported by other authors.

us to wonder whether the results reported for English by Stone et al. reflect crosslinguistic differences between English and French or whether there are also reasons to suspect that Stone et al.'s findings are spurious.

If one assumes that feedback consistency does influence lexical decision performance in English, the absence of a clear effect in French might reveal subtle differences in processing across languages. At the outset, we believed that because the French orthography is highly consistent in the print-to-sound direction, the FB consistency phenomenon would be easier to isolate. Obviously, we were wrong. As a matter of fact, the FB effect might be harder to capture precisely because of the high FF consistency of the French orthography. Reliance on phonological coding appears to increase with the degree of systematicity of the orthography (Frost & Katz, 1992). If readers of French (or any other regular orthography) rely on phonological information in the lexical decision task more than readers of English, the reverberation of phonological activation to orthographic units might not affect decision processes, thus failing to affect lexical decision performance. According to this hypothesis, one would predict the consistency effect to arise in regular orthographies only if the lexical decision situation is designed to hamper reference to phonological information or in a different task that *demands* access to orthographic information.

One such task is the letter detection paradigm. Interestingly, there is evidence that letter detection performance may be influenced by the flow of activation from phonology to orthography, even in highly regular writing systems. Ziegler and Jacobs (1995; see also Ziegler, Van Orden & Jacobs, 1997c for a replication in English) demonstrated that letter detection was more difficult when the letter string including the target letter was homophonic with a word that did not contain that letter (I in GAIM, homophone of GAME), and conversely, that more false detections occurred when the letter string was homophonic with a word that included the target letter (I in GANE, homophone of GAIN). This phenomenon was initially observed in the German language, which is known for its regular orthography. These findings suggest that activation from phonological candidates evoked by the input feeds back to the orthographic codes and appear thus to corroborate the hypothesis that the evidence of a FB consistency effect depends on the nature of the task and the regularity of the orthography.

However, there is an important principled difference between these phenomena and the FB consistency effect. Whereas the influence of homophony in the letter search task or the Reicher task (Hooper & Paap, 1997) can be interpreted as showing that lexical phonological codes reverberate to orthographic word forms, they do not imply interactions between orthographic codes and phonological codes *at the sublexical level*. Hence, the letter detection results provide no direct support for a strong interactive view that assumes feedback influence during all stages of processing, including prelexical levels. They fit equally well with a restricted interactivity account in which interactions are limited to lexical processing levels.

Conversely, the most likely interpretation of the FB consistency effect is at the sublexical level. According to this analysis, the FB consistency effect may provide a major index for chosing between a strong and a restricted interactivity account and Stone et al.'s findings remain as the only evidence supportive of the strong interactivity notion.

However, although Stone et al. carefully matched stimuli for many relevant dimensions, including word frequency, they did not use subjective frequency. This may be particularly critical when word frequency estimates are based on a corpus of relatively limited size. Interestingly, while Stone et al. reported close average values of word frequency for their consistent and inconsistent words, based on the Kučera–Francis counts, a reanalysis using the Cobuild word frequency data from the Celex database (Baayen, Piepenbrock, & Gulikers, 1995) showed a significant trend for consistent words to be more frequent than inconsistent words ($t(1,84) = 2.07$, $p < .05$, based on Cobuild Log frequency). Note that the Cobuild counts are based on a sample of 16.6 million words, whereas the Kučera–Francis table used a corpus of only 1,014,232 words. Following this observation, we ran a multiple regression analysis on the 86 items from both experiments in Stone et al.'s study, using the logarithm of Cobuild frequency, neighborhood density (from the MRC database; Coltheart, 1981), and a dummy variable coding for FB consistency as predictors. The results showed that all three predictors were significant (Frequency: $t(82) = 6.18$, $p < .0001$; Neighborhood density: $t(82) = 2.26$, $p < .05$; Consistency category: $t(82) = 2.10$, $p < .05$). Thus, whereas word frequency appears to account for a substantial part of the difference between FB consistent and inconsistent stimuli, it does not completely wash out the consistency effect. It remains to be demonstrated that, contrary to what happened with the French stimuli used by Ziegler et al., the effect resists to a control for subjective frequency.

In conclusion, the experimental observations reported in this paper cast doubts on the existence of reciprocal constraints between orthography and phonology at prelexical stages of processing. Although the absence of a clear FB consistency effect in French might be explained by crosslinguistic differences and task specific decision strategies, additional studies are needed to ascertain the validity and generality of the phenomenon.

APPENDIX 1

Target Words Used in Experiments 1 and 2

Inconsistent words. benne, bord, buis, daim, dard, douane, drap, feinte, flux, gendre, gland, greffe, grêle, grès, kyste, lynx, noix, pince, plomb, rein, score, seigle, tempe, tiers, torse, trombe, tronc, trône

Consistent words. bave, boxe, bribe, charte, douche, dune, dupe, fade, feutre, fibre, fourbe, globe, gourde, larve, lime, meute, pacte, palme, piste, pulpe, rive, rixe, tarte, tigre, torche, tube, tuile, volt

APPENDIX 2

Target Words Used in Experiment 3

Inconsistent words. bail, benne, blues, bourg, buis, cèpe, clerc, comte, cran, dard, dense, feinte, flair, flash, gaze, glaive, gland, greffe, grès, grêle, hall, heurt, houx, jarre, jeun, joug, leurre, luth, mythe, nain, noce, noeud, noix, pause, plomb, porc, quinte, score, seiche, snack, steppe, stèle, taux, tempe, tors

Consistent words. lest, nonne, catch, pompe, bise, fisc, rhume, songe, golf, cuve, coude, fluide, prune, bonze, boxe, scribe, meute, broche, luge, guise, tige, grive, taxe, bègue, urne, laps, lionne, malt, ouest, buse, tube, ruse, linge, bave, fugue, moine, buse, poigne, dinde, trogne, ronce, fronde, louve, loge, taupe, volt

APPENDIX 3

Target Words Used in Experiment 4

Inconsistent words. blues, chèque, frêle, colle, geôle, glaive, greffe, houppe, jarre, menthe, mythe, noce, paon, plomb, prompt, puce, seiche, steppe, tank, thym, voeu, zèle, barre, noeud

Consistent words. zèbre, poisse, guise, terne, celte, fronde, poigne, trogne, cèdre, lionne, ouest, loge, golf, fugue, bronze, laps, fougue, fluide, volt, taxe, onze, boxe, borne, pompe

REFERENCES

Alegria, J., & Mousty, P. (1994). On the development of lexical and non-lexical spelling procedures in French-speaking normal and disabled children. In G. D. A. Brown & N. C. Ellis (Eds.), *Handbook of spelling: Theory, process and intervention* (pp. 211–226). Chichester: Wiley.

Allen, P. A., Wallace, B., & Weber, T. A. (1995). Influence of case type, word frequency, and exposure duration on visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance,* **21,** 914–934.

Andrews, S. (1989). Frequency and neighborhood effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory and Cognition,* **15,** 802–814.

Andrews, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory and Cognition,* **18,** 234–254.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The*

CELEX Lexical Database (CD-ROM). Philadelphia: Linguistic Data Consortium, University of Pennsylvania.

Borowsky, R., & Masson, M. E. J. (1996). Semantic ambiguity effects in word identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **22,** 63–85.

Carroll, J. B. (1971). Measurement properties of subjective magnitude estimates of word frequency. *Journal of Verbal Learning and Verbal Behavior,* **10,** 722–729.

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

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

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review,* **100,** 589–608.

Content, A., & Peereman, R. *Quantitative analyses of orthography to phonology mappings in English and French.* Manuscript in preparation.

Content, A., & Radeau, M. (1988). Données statistiques sur la structure orthographique du français. *Cahiers de Psychologie Cognitive,* **8,** 399–404.

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers,* **28,** 1–11.

Flieller, A., & Tournois, J. (1994). Imagery value, subjective and objective frequency, date of entry into the language, and degree of polysemy in a sample of 998 French words. *International Journal of Psychology,* **29,** 471–509.

Frost, R., & Katz, L. (1992). *Orthography, phonology, morphology, and meaning.* Amsterdam: Elsevier, North-Holland.

Galbraith, R. C., & Underwood, B. J. (1973). Perceived frequency of concrete and abstract words. *Memory and Cognition,* **1,** 56–60.

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General,* **113,** 256–281.

Gordon, B. (1985). Subjective frequency and the lexical decision latency function: Implications for mechanisms of lexical access. *Journal of Memory and Language,* **24,** 631–645.

Grainger, J. (1990). Word frequency and neighborhood frequency effects in lexical decision and naming. *Journal of Memory and Language,* **29,** 228–244.

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review,* **103,** 518–565.

Grossberg, S., & Stone, G. O. (1986). Neural dynamics of word recognition and recall: Attentional priming, learning, and resonance. *Psychological Review,* **93,** 46–74.

Hawkins, H. L., Reicher, G. M., Rogers, M., & Peterson, L. (1976). Flexible coding in word recognition. *Journal of Experimental Psychology: Human Perception and Performance,* **2,** 380–385.

Hino, Y., & Lupker, S. J. (1996). Effects of polysemy in lexical decision and naming: An alternative to lexical accounts. *Journal of Experimental Psychology: Human Perception and Performance,* **22,** 1331–1356.

Hooper, D. A., & Paap, K. R. (1997). The use of assembled phonology during performance of a letter recognition task and its dependence on the presence and proportion of word stimuli. *Journal of Memory and Language,* **37,** 167–187.

Howes, D. (1954). On the interpretation of word frequency as a variable affecting speed of recognition. *Journal of Experimental Psychology,* **48,** 106–112.

Imbs, P. (1971). *Etudes statistiques sur le vocabulaire français. Dictionnaire des fréquences. Vocabulaire littéraire des XIXe et XXe siècles.* Centre de Recherche pour un trésor de la langue franąise (CNRS), Nancy, Paris: Librairie Marcel Didier.

Jacobs, A. M., & Grainger, J. (1992). Testing a semistochastic variant of the Interactive Activation model in different word recognition experiments. *Journal of Experimental Psychology: Human Perception and Performance,* **18,** 1174–1188.

Jacobs, A. M., & Grainger, J. (1994). Models of visual word recognition—Sampling the state of the art. *Journal of Experimental Psychology: Human Perception and Performance,* **20,** 1311–1334.

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

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

Jared, D., & Seidenberg, M. S. (1991). Does word identification proceed from spelling to sound to meaning? *Journal of Experimental Psychology: General,* **120,** 358–394.

Kawamoto, A. H., & Zemblidge, J. (1992). Pronunciation of homographs. *Journal of Memory and Language,* **31,** 349–374.

Kay, J. (1987). Phonological codes in reading: Assignment of sub-word phonology. In D. A. Allport, D. Mackay, W. Prinz, & E. Scheerer (Eds.), *Language perception and production* (pp. 181–196). London: Academic Press.

Kay, J., & Bishop, D. (1987). Anatomical differences between nose, palm and foot, or, the body in question: Further discussion of the processes of sub-lexical spelling-sound translation. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 449–469). London: Erlbaum.

Kreiner, D. S. (1992). Reaction time measures of spelling: Testing a two-strategy model of skilled spelling. *Jour-*

nal of Experimental Psychology: Learning, Memory, and Cognition, **18,** 765–776.

Kreiner, D. S. (1996). Effects of word familiarity and phoneme-grapheme polygraphy on oral spelling time and accuracy. *Psychological Record,* **46,** 49–70.

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

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

Laxon, V., Masterson, J., & Coltheart, V. (1991). Some bodies are easier to read: The effect of consistency and regularity on children's reading. *Quarterly Journal of Experimental Psychology,* **43A,** 793–824.

Marslen-Wilson, W., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology,* **10,** 29–63.

Massaro, D. W., & Cohen, M. M. (1991). Integration versus interactive activation: The joint influence of stimulus and context in perception. *Cognitive Psychology,* **23,** 558–614.

Massaro, D. W., & Cohen, M. M. (1994). Visual, orthographic, phonological, and lexical influences in reading. *Journal of Experimental Psychology: Human Perception and Performance,* **20,** 1107–1128.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part I. An account of basic findings. *Psychological Review,* **88,** 375–407.

Mewhort, D. J. K., & Johns, E. E. (1988). Some tests of the interactive-activation model for word identification. *Psycholinguistic Research,* **50,** 135–147.

Paap, K. R., Newsome, S. L., McDonald, J. E., & Schvaneveldt, R. W. (1982). An activation-verification model for letter and word recognition: The word superiority effect. *Psychological Review,* **89,** 573–594.

Paap, K. R., & Noel, R. W. (1991). Dual-route models of print to sound: Still a good horse race. *Psychological Research,* **53,** 13–24.

Peereman, R. (1995). Naming regular and exception words: Further examination of the effect of phonological dissension among lexical neighbours. *European Journal of Cognitive Psychology,* **7,** 307–330.

Peereman, R., & Content, A. (1995). The neighborhood size effect in naming: Lexical activation or sublexical correspondences? *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **21,** 409–421.

Peereman, R., & Content, A. (1997). Orthographic and phonological neighborhoods in naming: Not all neighbors are equally influential in orthographic space. *Journal of Memory and Language,* **37,** 382–410.

Peereman, R., & Content, A. (1998). LEXOP. A LEXical database with Orthography-Phonology statistics for French monosyllabic words. *Behavior Research Methods, Instruments, & Computers,* in press.

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology,* **81,** 275–280.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review,* **89,** 60–94.

Schneider, V. I., & Healy, A. F. (1993). Detecting phonemes and letters in text: Interactions between different types and levels of processes. *Memory and Cognition,* **21,** 739–751.

Segui, J., Mehler, J., Frauenfelder, U., & Morton, J. (1982). The word frequency effect and lexical access. *Neuropsychologia,* **20,** 615–627.

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

Seidenberg, M. S., & Tanenhaus, M. K. (1979). Orthographic effects on rhyme monitoring. *Journal of Experimental Psychology: Human Learning and Memory,* **5,** 546–554.

Shapiro, B. J. (1969). The subjective estimation of relative word frequency. *Journal of Verbal Learning and Verbal Behavior,* **8,** 248–251.

Spoehr, K. T. (1978). Phonological encoding in visual word recognition. *Journal of Verbal Learning and Verbal Behavior,* **17,** 127–141.

Stone, G. O., & Van Orden, G. C. (1994). Building a resonance framework for word recognition using design and systems principles. *Journal of Experimental Psychology: Human Perception and 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.

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

Taraban, R., & McClelland, J. L. (1987). Conspiracy effects in word pronunciation. *Journal of Memory and Language,* **26,** 608–631.

Thorndike, E. L., & Lorge, I. (1944). *The teacher's word book of 30,000 words.* New York: Bureau of Publications, Teachers College, Columbia University.

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.

Tryk, H. E. (1968). Subjective scaling of word frequency. *American Journal of Psychology,* **81,** 170–177.

Van Orden, G. C. (1987). A ROWS is a ROSE: Spelling, sound, and reading. *Memory and Cognition,* **10,** 434–442.

Van Orden, G. C., & Goldinger, S. D. (1994). Interdependence of form and function in cognitive systems explains perception of printed words. *Journal of Experimental Psychology: Human Perception and Performance,* **20,** 1269–1291.

Van Orden, G. C., Johnston, J. C., & Hale, B. L. (1988). Word identification in reading proceeds from spelling to sound to meaning. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **14,** 371–385.

Van Orden, G. C., Pennington, B. F., & Stone, G. O. (1990). Word identification in reading and the promise of subsymbolic psycholinguistics. *Psychological Review,* **97,** 488–522.

Whalen, D. H., & Zsiga, E. C. (1994). Subjective familiarity of English word / name homophones. *Behavior Research Methods, Instruments, & Computers,* **26,** 402–408.

Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology,* **1,** 59–85.

Ziegler, J. C., & Jacobs, A. M. (1995). Phonological information provides early sources of constraint in the processing of letter strings. *Journal of Memory and Language,* **34,** 567–593.

Ziegler, J. C., Jacobs, A. M., & Stone, G. O. (1996). Statistical analysis of the bidirectional inconsistency of spelling and sound in French. *Behavior Research Methods, Instruments, & Computers,* **28,** 504–515.

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

Ziegler, J. C., Stone, G. O., & Jacobs, A. M. (1997b). What's the pronunciation for OUGH and the spelling for /u/ ? A database for computing feedforward and feedback inconsistency in English. *Behavior Research Methods, Instruments, & Computers,* **29,** 600–618.

Ziegler, J. C., Van Orden, G. C., & Jacobs, A. M. (1997c). Phonology can help or hurt the perception of print. *Journal of Experimental Psychology: Human Perception and Performance,* **23,** 845–860.

(Received September 15, 1997)
(Revision received March 12, 1998)